


NUMBER 23 . YEAR XIII / JULY 2024

ISSN 2281-8138

The cover illustration depicts a person's face from the nose up, framed by a large, intricate, mechanical structure. This structure is composed of various lenses, tubes, and circular components, resembling a complex camera or a sensor array. The person's eyes are obscured by a central lens that glows with a bright light. The background is a light, textured surface with faint, sketchy lines and shapes, suggesting a technical or scientific environment.

THE IMAGINARY  
OF **ARTIFICIAL** INTELLIGENCE

 MIMESIS EDIZIONI

Mimesis Edizioni  
via Risorgimento, 33 - 20099 Sesto S.G.  
(MI) - Italy phone/fax +39 02 89403935  
Email: [mimesis@mimesisedizioni.it](mailto:mimesis@mimesisedizioni.it)

Registrazione presso il Tribunale di  
Messina n. 8 del 16/04/2012



ISSN 2281-8138  
[www.imagojournal.it](http://www.imagojournal.it) -  
[rivistaimago@gmail.com](mailto:rivistaimago@gmail.com)



## **Editor in Chief**

**Pier Luca Marzo** (*University of Messina*)

**Milena Meo** (*University of Messina*)

## **Editorial Board**

**Francesco Barbalace** (*University of Reggio Calabria*)

**Mariavita Cambria** (*University of Messina*)

**Antonio Camorrino** (*University of Naples "Federico II"*)

**Marianne Celka** (*Paul Valéry University Montpellier 3*)

**Fabio D'Andrea** (*University of Perugia*)

**Matthijs Gardenier** (*Paul Valéry University Montpellier 3*)

**Giovanni La Fauci** (*University of Messina*)

**Luca Mori** (*University of Verona*)

**Maria Giovanna Musso** (*University of Rome "La Sapienza"*)

**Daniela Pomarico** (*Paul Valéry University Montpellier 3*)

**Vincenzo Susca** (*Paul Valéry University Montpellier 3*)

**Antonio Tramontana** (*University of Messina*)

**Neil Turnbull** (*Nottingham Trent University*)

**Bertrand Vidal** (*Paul Valéry University Montpellier 3*)

## **Scientific Committee**

**Sabah Abouessalam** (*Institut National d'Aménagement et d'Urbanisme*), **Alfonso Amendola** (*Università degli Studi di Salerno*), **Vania Baldi** (*University of Aveiro*), **Luisa Bonesio** (*Università di Pavia*), **Sergio Brancato** (*Università degli Studi di Napoli "Federico II"*), **David Brown** (*Cardiff Metropolitan University*), **Fulvio Carmagnola** (*Università di Milano - Bicocca*), **Antonella Cammarota** (*Università di Messina*), **Vanni Codeluppi** (*Università di Modena e Reggio Emilia*), **Stefano Cristante** (*Università del Salento*), **Derrick De Kerckhove** (*University of Toronto*), **Luisa Del Giudice** (*American Folklore Society*), **Ubaldo Fadini** (*Università di Firenze*), **Kenneth Frampton** (*Graduate School of Architecture, Planning and Preservation at Columbia University*), **Tonino Griffero** (*Università di Tor Vergata*), **Bruno Gullì** (*Long Island University of New York*), **Paolo Jedlowski** (*Università della Calabria*), **Philippe Joron** (*Paul Valéry University Montpellier 3*), **Serge Latouche** (*Université Paris-Sud*), **Michel Maffesoli** (*Université Paris V*), **Ingo Meyer** (*Bielefeld University*), **Carlo Mongardini** † (*Università di Roma "La Sapienza"*), **Edgar Morin** (*Centre National de la Recherche Scientifique*), **Francesco Parisi** (*Università di Messina*), **Tonino Perna** (*Università di Messina*), **Mario Perniola** † (*Università di Roma Tor Vergata*), **Andrea Pinotti** (*Università di Milano*), **Luigi Prestinzenza Puglisi** (*Università di Roma La Sapienza*), **Valentina Raffa** (*University of Messina*), **Caterina Resta** (*Università di Messina*), **Ambrogio Santambrogio** (*Università di Perugia*), **Antonio Scurati** (*Libera Università di Lingue e Comunicazione*), **Domenico Secondulfo** (*Università di Verona*), **Patrick Tacussel** (*Université Paul-Valéry, Montpellier III*), **Bron Taylor** (*University of Florida*), **Dario Tomasello** (*Università di Messina*), **Gabriella Turnaturi** (*Università di Bologna*), **Andreas Wittel** (*Nottingham Trent University*), **Jean Jacques Wunenburger** (*Università Jean Moulin Lyon*), **Martine Xiberras** (*Paul Valéry University Montpellier 3*).



# THE IMAGINARY OF *ARTIFICIAL* INTELLIGENCE

EDITOR: MARIA GIOVANNA MUSSO / MICHELA MUSTO / MARIA FRANCESCA MUSTO



Cover:  
Giovanni  
La Fauci





# Index



## *Editorial*

Maria Giovanna Musso Michela Musto Maria Francesca Musto	13	Editorial Board
--	----	-----------------



## *Topic*

Simone D'Alessandro Franco Ferzetti	23	Interactions between a human and a chat bot: the mythopoietic narrative of unlimited possibilities
Edmondo Grassi	39	Tecnotopie. Per una epistemologia simbolica degli algoritmi intelligenti
Giulia Crippa	55	Forme e tropi dell'immaginario dell'Intelligenza Artificiale: il contributo della filosofia di Yuk Hui
Paolo Bory	71	Bambini artificiali. La metafora del bambino e l'immaginario dell'IA
Maurizio Bonolis Barbara Sonzogni	87	Coscienza e Intelligenza Artificiale: due ontologie dell'immaginario
Riccardo Corsi Ernesto d'Albergo Giorgio Giovanelli	109	La politica dell'intelligenza artificiale general purpose: immaginari socio-tecnici, democrazia e policy frame nel processo decisionale della regolazione europea ("EU AI ACT"-2022-2024)
Stefano Spillare Michele Bonazzi Piergiorgio Degli Esposti	131	AI Imaginaries and Narratives in the Italian Public Discourse: The Impact of Chat-GPT
Antonio Opromolla Silvia di Rocco Lorenza Parisi	149	AI in the Public Sector: An Exploration of Social Imaginaries
Anita Lavorgna	169	Sociotechnical imaginaries of artificial intelligence in EU law making: a focus on crime and security
Paulina Sabugal Paz	193	El amor imaginario. Una mirada sociológica a las relaciones amorosas y la inteligencia artificial



# Index

Francesca Capalbi	215	Quando l'intelligenza artificiale è donna: la rappresentazione dell'AI in alcuni film di successo
Francesca Farruggia	235	Oltre l'immaginario. Quando l'AI va in guerra
Lorenzo Aimo Luca Brunet	249	Immaginario Black Box: prospettive mitiche e critiche tra innovazione bellica ed esplorazione artistica
Paola Panarese, Vittoria Azzarita, Maddalena Carbonari	277	Oltre l'estetica McNuggets. Esplorazioni creative e interazioni tecno-sociali nella produzione artistica realizzata con l'IA
Ingrid Kofler Mustapha El Moussaoui Romuald Jamet	291	AI's influence on the Creative and Cultural Industries
		 <i>Off-Topic</i>
Giovanna Vingelli	315	Manosfera e pick-up artists
Manolo Farci	335	Siamo l'esercito dei "bravi ragazzi". La politica identitaria della manosphere
Antonia Cava Debora Pizzimenti	355	L'odio indicibile: un'indagine etnografica digitale sulle comunità Femcel
Assunta Penna	375	Stigma e gestione dell'identità nelle piattaforme digitali



# Index



## *Focus*

- |                   |     |  |
|-------------------|-----|--|
| Olimpia Affuso    | 393 | Intelligenza artificiale. E se avessimo sbagliato metafora? Una riflessione a partire da Comunicazione artificiale. Come gli algoritmi producono intelligenza artificiale di Elena Esposito (2022) |
| Daniele Garritano | 405 | Nella spirale della conoscenza: arte e scienza tra creatività e metodo   |



## *Aisthesis*

- |                               |     |  |
|-------------------------------|-----|--|
| Iginio de Luca                | 421 | Gioia mia  |
| Florencia S.M. Brück          | 429 | Memory Divergence. An Artist's Reflection on AI and the Preservation of Truth  |
| Marco Barbon<br>Yann Toussain | 439 | An Algorithmic Commentary on The Photographic Essay intitled «LLDLC / G. A Non-Human Overcomplicated Photographic Essay exploring the intersectionality of Gibraltar and La Linea De La Conception in the wake of Roland Barthes' « S/Z » as the Artificial Intelligence said to subtitle it » |





*Editorial*







A JOURNAL OF THE  
SOCIAL IMAGINARY

## Editorial Board

Il numero 23 di Im@go è dedicato all'immaginario dell'Intelligenza Artificiale, alle premesse e alle conseguenze che l'uso dell'IA ha sul piano mitico-simbolico e al ruolo che essa svolge in questo passaggio epocale; un ruolo foriero, com'è noto, di straordinarie promesse di trasformazione e potenziamento dell'umano, ma anche di enormi rischi, molti dei quali non del tutto prevedibili e già adombrati da scienziati ed esperti della materia (fra gli altri: Bostrom, 2014; Zuboff, 2019; Russell e Norvig, 2020; Suleyman, 2023).

Mentre il dibattito è in corso nelle scienze e nella società, la ricerca e lo sviluppo dell'IA e delle sue applicazioni procede a passi da gigante. Oltre ad essere l'infrastruttura immateriale - che già da anni supporta, regola, orienta, e talvolta sostituisce le scelte umane in vari ambiti - è divenuta anche un fenomeno di massa con effetti dirompenti in ogni ambito della vita. Utilizzata da masse crescenti di popolazione, essa attrae una grande mole di investimenti economici, produce effetti dirompenti nelle scelte industriali, incide sulle caratteristiche del lavoro e delle professioni, modifica radicalmente il campo dell'educazione, dell'arte, della medicina, dello sport, della guerra, ecc. Malgrado i dubbi sulle possibilità di regolare efficacemente la sua avanzata, definita *disruptive* (Suleyman, 2023), l'IA è divenuta oggetto di regolazione politico-normativa e tema centrale del discorso pubblico.

Di fronte all'ampiezza e alla radicalità del fenomeno in tutte le sue sfaccettature, alla diffusione dei suoi impieghi e all'incisività dei suoi effetti, perché interrogarsi sull'immaginario dell'Intelligenza Artificiale - cioè sull'*altra faccia*, quella più nascosta e invisibile, del complesso tecnologico, economico e produttivo - piuttosto che indagare semplicemente i fenomeni concreti nel loro andamento *fattuale*?

La prima ragione è che l'IA stessa si alimenta *strutturalmente* di un corredo immaginario e mitico, fatto di metafore, analogie, miti ed "errori" logici (non sappiamo ancora se solo evolutivi o anche catastrofici) senza il quale, semplicemente, non potrebbe esistere così com'è. In mancanza di una indagine nell'ambito della dimensione mitico-simbolica - che precede, immagina, simula, prospetta e inventa i presupposti logico-simbolici degli artefatti - non è possibile comprendere le profonde "ragioni" - e neppure la potenza propulsiva - delle realizzazioni tecniche in generale e dell'IA in particolare. Inoltre, senza l'analisi e la comprensione delle narrative e delle rappresentazioni sociali che oggi accompagnano e favoriscono la diffusione dei sistemi di IA, non potremmo comprendere neppure i percorsi della sua diffusione, accettazione e legittimazione



sociale. Questi elementi, ovviamente insieme agli ingenti interessi economici che spingono la ricerca e gli investimenti – costituiscono gli ingredienti indispensabili per chiarire i presupposti e la specificità della nuova costellazione di senso che prefigura il “nuovo mondo” che sta prendendo corpo e la cui luce già s’irradia nei futuri possibili.

La seconda ragione, non meno importante, è che l’IA, ancor più delle altre tecnologie dirompenti che l’hanno preceduta (come la scrittura, l’elettricità, il computer) – in sinergia con la biologia sintetica (*gene editing*, vita artificiale, ecc.) (Suleyman, 2023) – non è solo il *driver* di un cambiamento radicale che coinvolge ogni sfera della vita sociale, ma è la leva epistemico-tecnologica in grado di sovvertire – e al tempo stesso “rivelare” heideggerianamente – i presupposti simbolici e immaginari che definiscono l’umano in questo passaggio d’epoca.

La diffusione generale e rapida delle IA, nella diversità delle loro applicazioni ed effetti, contribuisce, infatti, all’istituzione di un nuovo rapporto fra l’umano e l’Altro da sé, a riconfigurare il rapporto tra natura e artificio, tra fenomeno e rappresentazione, tra verità e simulazione, tra simulacro e “realtà”, a rimodellare le nozioni tradizionali d’intelligenza, coscienza, vita, creatività, verità, *agency*, ecc. Se questo è possibile è anche grazie alla circolazione e alla “naturalizzazione” dei contenuti immaginari che ispirano l’AI e da cui sprigiona buona parte della sua potenza. Come verrà argomentato anche in alcuni dei saggi contenuti in questo numero, è grazie alle narrative mitiche, al ricorso a metafore e analogie – “magiche” e persuasive – che ne accompagnano la diffusione, che l’IA è divenuta e resta una forza di trasformazione in grado di metamorfosare la società contemporanea, proiettandola nei nuovi orizzonti post-umani o trans-umanistici.

Il rapporto fra IA e immaginario, dunque, lungi dall’essere accidentale o accessorio, è cruciale per comprendere le origini, la fenomenologia e il successo attuale dell’IA e il suo ruolo nei processi di mutamento sociale. Gli elementi *immaginari* e gli elementi *fattuali* della realtà e del mutamento sociale, infatti, non sono mai scissi, ma coinvolti in una dinamica di co-essenzialità e di alimentazione biunivoca, e ciò vale in special modo per la sfera tecnologica. Tecnica e immaginario sono, da sempre, coinvolti in una dinamica generativa e co-evolutiva nei processi di mutamento sociale (Musso, 2019).

Queste sono le ragioni per cui, data la varietà delle dimensioni implicate in questa nuova Grande Trasformazione, nel numero dedicato all’immaginario dell’IA abbiamo voluto accogliere riflessioni e analisi riguardanti le trasformazioni in atto in diverse dimensioni della vita sociale (economica, lavorativa, scientifica, artistica, giuridica, etica ecc.) in cui interviene l’IA, purché fossero analizzate dal vertice teorico-epistemologico che privilegia il ruolo dell’immaginario nell’istituzione della realtà sociale.



La nozione d'immaginario qui adottata è, come si vede, sottratta tanto al suo statuto tradizionalmente *difettivo* - che ne fa un sinonimo di irrealtà o di fantasticheria - quanto alla sfera della *specularità*, che la vede come riflesso della realtà da cogliere, ad esempio, nei prodotti dell'immaginario proposti dall'industria culturale. L'immaginario è cioè concepito, in termini più radicali, come *matrice* della realtà sociale, senza la quale né le istituzioni sociali e neppure gli artefatti tecnologici potrebbero esistere.

Questa concezione dell'immaginario è lo sfondo teorico su cui si basa l'ideazione di questo numero e corrisponde alla linea editoriale della rivista. La consapevolezza di quanto questa concezione sia minoritaria, "eretica", poco conosciuta e ancor meno condivisa, è emersa in maniera netta durante la realizzazione di questo numero.

Nel corso della sua realizzazione abbiamo infatti registrato fino a che punto la nozione invalsa di immaginario sia controversa e sfuggente, sebbene ormai di moda. In alcuni casi, anche negli articoli sottoposti a selezione, il riferimento all'immaginario è risultato fugace, di circostanza, o persino fuorviante. Del resto, non era facile raccogliere contributi che fossero teoricamente attrezzati contemporaneamente nei tre ambiti di ricerca di cui era richiesta la padronanza scientifica, cioè l'immaginario, l'IA e il campo di applicazione specifico dell'expertise dei singoli studiosi (dal lavoro, alla guerra, all'arte ecc.).

Malgrado da ciò siano derivate alcune difficoltà nella composizione del numero, riteniamo che il risultato ottenuto e qui proposto, dia un quadro sufficientemente ampio e significativo delle ricerche e delle riflessioni in corso sugli argomenti in questione, e possa essere un utile punto di partenza per approfondimenti futuri.

Ai fini di una migliore fruizione, i contributi selezionati sono stati inseriti in macroaree tematiche. La prima è quella che riguarda le radici immaginarie ed epistemiche dell'IA. Qui l'intento era quello di far luce sulle strutture simboliche (mitiche, leggendarie, archetipiche, religiose, magiche, ecc.) che costituiscono il territorio immaginario che alimenta l'Intelligenza Artificiale, compresi i contesti discorsivi e le fratture epistemologiche create nel passaggio dalla meccanica alla cibernetica e che tuttora costellano il tracciato tecnico/scientifico di cui l'IA è il prodotto. Le domande specifiche di quest'area tematica che hanno trovato interesse nei contributi pervenuti sono relative, sostanzialmente, a cosa significhi per una macchina "capire" il linguaggio, quale sia la *natura* della coscienza e quali i *confini* tra intelligenza umana e artificiale.

In quest'area rientrano i contributi di Simone D'Alessandro e Francesco Ferzetti, di Edmondo Grassi, di Paolo Bory e quello, a doppia firma, di Maurizio Bonolis e Barbara Sonzogno.





Il primo contributo, ovvero *Interactions between a human and a Chat Bot: The Mythopoietic narrative of unlimited possibilities*, affronta criticamente le asimmetrie e i limiti dell'interazione tra esseri umani e IA. In base a un esame critico della nozione di agency, di auto-riflessività, comprensione semantica e capacità creativo-trasformativa, gli autori decostruiscono la narrazione mitopoietica più diffusa dell'IA che la considera una forma di superamento dei limiti dell'umano.

Anche Edmondo Grassi, nell'articolo intitolato *Tecnotopie. Per una epistemologia simbolica degli algoritmi intelligenti*, riflette sull'interazione tra algoritmi intelligenti e esseri umani, definendo i primi come *algoagenti*, cioè agenti sociali dal carattere performativo e politico che incarnano e riflettono specifiche configurazioni di potere e conoscenza. Di fronte alle sfide sociali ed etiche che il loro uso comporta l'autore propone specifici percorsi di consapevolezza e di educazione culturale.

L'articolo di Giulia Crippa, dal titolo *Forme e tropi dell'immaginario dell'Intelligenza Artificiale: il contributo della filosofia di Yuk Hui*, ha invece l'obiettivo di tracciare una genealogia dell'immaginario sull'IA – attraverso un'analisi dei *tropi* e delle forme che ne hanno plasmato la percezione sociale – ai fini di una comparazione fra le sue immagini passate e l'attuale (ri)definizione dei suoi orizzonti epistemici. Questa possibile genealogia degli immaginari dell'IA viene discussa alla luce dei contributi di Yuk Hui sul tema della tecnologia e della tecno-diversità.

Paolo Bory, in *Bambini artificiali. La metafora del bambino e l'immaginario dell'IA*, si sofferma su una specifica componente del processo immaginario che accompagna la creazione e l'evoluzione dell'IA: il processo di antropomorfizzazione, e in particolare quello che insiste sulla metafora dell'IA bambino/a. Le ricadute di tale metafora, presente fin dalle origini dell'IA, connotano le macchine intelligenti e incidono sulle rappresentazioni e gli sviluppi possibili di questa tecnologia.

Maurizio Bonolis e Barbara Sonzogni, in *Coscienza e Intelligenza Artificiale: due ontologie dell'immaginario*, dopo aver esposto una concezione dell'immaginario ancorata sostanzialmente a dicotomie classiche (materiale/immateriale, reale/fantastico, metaforico/fisico ecc.) si interrogano su ciò che ritengono il "vero" problema dell'IA, cioè la replicazione della coscienza, ribadendo l'irriducibilità ontologica di *coscienza* e *conoscenza*.

Nella seconda macroarea, sono inclusi i contributi riguardanti gli immaginari e le retoriche istituzionali che accompagnano l'integrazione dell'IA nella vita pubblica e nella sfera della regolazione politico-normativa.

Nel primo contributo di quest'area – dal titolo *Immaginari dell'Intelligenza artificiale e regolazione politica: costruzione e negoziazione del rischio nell'arena europea* – Riccardo Corsi e Ernesto d'Albergo analizzano il processo decisionale che ha portato all'approvazione dell'EU IA Act, e rilevano il modo in cui la categoria del

"rischio per la democrazia" e del "rischio sistemico" hanno contribuito a strutturare la cornice politica della legge europea sull'IA. Sottolineano, in base alla loro analisi, come l'immaginario svolga un ruolo al tempo stesso cognitivo e normativo nei processi di *sense making* e nelle pratiche sociali, contribuendo a strutturare le relazioni tra i diversi sottosistemi e a orientare il mutamento sociale.

Il secondo contributo – *AI Imaginaries and Narratives in the Italian Public Discourse: The Impact of Chat-GPT* – a firma di Stefano Spillare, Michele Bonazzi, Piergiorgio Degli Esposti, indaga gli immaginari e le narrazioni dell'IA nei principali quotidiani italiani, valutando in particolare l'impatto di ChatGPT sul discorso pubblico. La ricerca, svolta mediante un'analisi comparativa degli articoli pubblicati prima e dopo il lancio di ChatGPT in Italia, ricostruisce come i media mainstream hanno rappresentato i fenomeni legati all'IA e le loro implicazioni sociali.

L'articolo di Antonio Opromolla, Silvia di Rocco e Lorenza Parisi – *AI in the Public Sector: An Exploration of Social Imaginaries* – è una esplorazione degli immaginari che ispirano le politiche pubbliche europee e italiane. Mediante le nozioni di *immaginario eteronomo* e *immaginario autonomo* di Castoriadis (1995), e attraverso un'analisi del discorso applicata a una serie di interviste con esperti di IA nel settore pubblico italiano, gli autori evidenziano differenze e somiglianze nel campo politico e in quello degli esperti.

Un focus specifico sui temi del crimine e della sicurezza è quello proposto da Anita Lavorgna, in *Sociotechnical imaginaries of artificial intelligence in EU law making: a focus on crime and security*. L'utilizzo dell'IA in campo giudiziario, nel contrasto al crimine e nell'implementazione di nuovi sistemi di sicurezza è divenuto centrale e sempre più diffuso. L'autrice, concentrandosi sugli immaginari sociotecnici legati all'IA – con particolare attenzione al crimine e alla sicurezza – nel processo legislativo dell'UE, nei dibattiti e nelle negoziazioni intorno all'Artificial Intelligence Act (il regolamento UE sull'intelligenza artificiale), sottolinea come le narrazioni prevalenti possano offuscare i danni delle tecnologie *data-driven*, evocando prospettive idealistiche relative alla loro adeguatezza.

Nella terza macroarea sono stati inseriti i contributi riguardanti gli effetti e le ambivalenze dell'IA in ambiti diversi della vita sociale – fra cui l'amore, la guerra, l'arte – per evidenziare le conseguenze che, in termini fenomenici, prospettici e immaginari, si riversano in tali ambiti, modificandone i caratteri e le pratiche.

Il lavoro di Paulina Sabugal, *El amor imaginario. Una mirada sociológica a las relaciones amorosas y la inteligencia artificial*, è una riflessione intorno alle possibili trasformazioni dei modelli tradizionali di amore e intimità alimentate dalla presenza dell'IA e dei cosiddetti "assistenti personali" nella vita sentimentale e nelle relazioni intime.



Quello di Francesca Capalbi, *Quando l'intelligenza artificiale è donna: la rappresentazione dell'AI in alcuni film di successo* è un lavoro dedicato alle connotazioni di genere contenute nell'immaginario dell'IA. Dopo aver esplorato le principali tendenze di rappresentazione nella produzione cinematografica relativa all'IA, l'autrice analizza alcuni film che hanno come protagonista un'IA donna, rilevando le implicazioni di genere che attraversano sia la progettazione tecnologica sia l'immaginario cinematografico, ed evidenziando il loro reciproco sostegno.

Un contributo particolarmente rilevante in quest'area è quello di Francesca Farruggia sul tema della guerra. In *Oltre l'immaginario. Quando l'AI va in guerra*, l'autrice illustra i principali problemi generati dai crescenti investimenti per la R&D nell'IA e dall'uso sempre più massiccio delle armi autonome, con le conseguenze che ne derivano sul piano militare, politico, sociale ed etico.

In *Immaginario Black Box: prospettive mitiche e critiche tra innovazione bellica ed esplorazione artistica*, Lorenzo Aimò e Luca Brunet, conducono un'analisi delle rappresentazioni dell'IA su alcuni siti web. Dalla loro ricerca emerge l'importanza cruciale che la metafora della Black Box ha nella definizione dell'IA, soprattutto in ambito bellico e artistico, in quanto capace di modellare le retoriche, siano esse di accettazione o di tipo critico, con le loro implicazioni politiche, sociali e semiotiche.

Gli ultimi due articoli sono accomunati, da una interrogazione sul modo in cui l'IA ridefinisce i contorni e il senso dell'arte e, più in generale, le pratiche creative.

In *Oltre l'estetica dei McNuggets. Pratiche creative e immaginari sociali nella produzione artistica realizzata con l'IA*, Paola Panarese, Vittoria Azzarita e Maddalena Carbonari, presentano alcuni risultati di una ricerca volta a far luce sugli usi dell'IA in campo artistico. Attraverso una serie di interviste con artisti che operano sulla scena dell'arte contemporanea italiana, approfondiscono la dinamica generativa e co-evolutiva tra arte e IA e il rapporto di interdipendenza che permane tra tecnica e creatività artistica.

Nell'ultimo articolo del *Topic*, Mustapha El Moussaoui, Ingrid Kofler e Romuald Jamet, in *Redefining Social Imaginaries: AI's Influence on Architecture and Music*, approfondiscono il tema della intersezione tra IA e industrie culturali, con particolare attenzione al mondo della musica e dell'architettura. Gli autori sottolineano il doppio connotato dell'IA in grado di potenziare le capacità creative dei professionisti e, al tempo stesso, di limitarne e condizionarne l'espressione.

Questo numero, in stretta connessione con i temi trattati nel *Topic*, ospita poi – in una apposita sezione denominata *Aistehsis* – alcuni contributi *sui generis*. Si tratta di alcune opere d'arte, illustrate dagli stessi artisti, che sono state realizzate con il contributo dell'IA.



La redazione di Im@go ha infatti scelto di inaugurare, a partire da questo numero, una nuova sezione dedicata alle opere d'arte e ai prodotti culturali che meglio esprimono i fermenti e le inquietudini del nostro tempo.

Il contributo diretto degli artisti è di fondamentale importanza in generale per la riflessione sociologica. Nel caso della riflessione sull' IA il loro apporto è decisivo non solo perché essi sono direttamente coinvolti sui temi della creatività artistica, dell'autorialità, della serialità, e sono i primi a sperimentare opportunità e rischi connessi all'uso dell'IA, ma anche perché lo sguardo divergente, che è proprio dell'arte, consente alla riflessività sociologica di entrare meglio, e più in profondità, in quella dimensione, nascosta e invisibile, della realtà sociale che è l'immaginario.

I tre contributi artistici che ospitiamo in questo numero sono molto diversi fra loro, in termini di poetica, di linguaggio, di tematiche affrontate. Ma tutti sono accomunati dall'impiego dell'IA nel processo creativo e dalla riflessione sugli effetti che tale "collaborazione" implica.

Il primo contributo si intitola *Gioia mia*. È un'opera di Iginio de Luca, artista noto per le sue folgoranti azioni urbane, denominate Blitz, con le quali usa lasciare il segno graffiante della sua poetica nel tessuto urbano della città in cui vive: Roma. In questo caso, l'opera, pur avendo la forma di un'azione urbana, si discosta dai temi politici e di denuncia cari all'artista ed evoca invece i temi più intimi del dolore e della perdita (in questo caso della madre morta di cancro). Al vuoto e all'assenza della madre l'artista oppone un'opera in cui la surreale ricerca d'immortalità passa attraverso il sostegno tecnologico dell'IA.

Il secondo contributo artistico è quello di Florencia Brück, artista e programmatrice italo-argentina, che lavora quasi sempre con le nuove tecnologie. Nell'opera che qui propone, *Memory Divergence*, Brück riflette sulla complessa interazione tra l'intelligenza artificiale (IA) e la conservazione della verità storica. Mediante l'utilizzo di un'IA generativa addestrata da lei stessa sullo stile di Hieronymus Bosch, esplora il duplice potenziale dell'IA, in grado di salvaguardare – ma anche di distorcere – la memoria umana, modificando il modo in cui ricordiamo e interpretiamo la realtà e la storia.

Il terzo contributo è il frutto di un esperimento sul tema dei *confini* realizzato da due artisti fotografi residenti in Francia: Marco Barbon e Yann Tostain. Nel corso di un viaggio che li ha condotti dal Sud della Francia fino a Gibilterra hanno coinvolto ChatGPT come musa ispiratrice, guida concettuale, oracolo e interlocutore chiamato a pronunciarsi sui grandi temi della creazione artistica, in particolare sulla definizione dei confini, sulla libertà dell'atto fotografico, sull'autorialità, sulla critica e l'interpretazione dei risultati artistici.



La ricchezza del panorama che emerge da questi contributi è ben più ampia di quanto si possa illustrare in un Editoriale. Questo sforzo di riflessività congiunta, non solo fra diversi ambiti della sociologia, ma anche mediante il confronto fra sociologia e arte, ha consentito certamente di evidenziare aspetti cruciali dei fenomeni analizzati e alcune criticità di questo passaggio d'epoca. Ma lungi dal fornire risposte certe e immediate, questo lavoro ha contribuito ad aumentare gli interrogativi e ad aprire il campo per ulteriori approfondimenti ai quali la riflessività sociologica non può sottrarsi. L'augurio è che tali approfondimenti siano possibili, vicini e fruttuosi e contribuiscano non solo all'arricchimento della conoscenza scientifica e della riflessività artistica, ma anche a potenziare l'intelligenza collettiva e le capacità critiche e decisionali della società tutta.

*Maria Giovanna Musso* <sup>[1]</sup>  
*Michela Musto*  
*Maria Francesca Musto*



## **Bibliography**

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Castoriadis, C. (1995). *L'istituzione immaginaria della società*. Torino, Bollati Boringhieri.

Musso, M. G. (2019). *Immaginario, tecnica e mutamento sociale*. In Marzo P. L. e Mori L. (a cura di) *Le vie sociali dell'immaginario. Per una sociologia del profondo*, pp. 115-141. Milano, Meltemi.

Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach*. Pearson.

Suleyman, M. (2023). *The coming wave: technology, power, and the twenty-first century's greatest dilemma*. New York, Crown.

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York, Public Affairs.

---

<sup>[1]</sup> Un ringraziamento particolare per la realizzazione di questo numero va a Mattia Martella, per la sua assidua e puntuale cura dei dettagli organizzativi e il supporto dato alla curatela.



*Topic*







A JOURNAL OF THE  
SOCIAL IMAGINARY

# Interactions between a human and a Chat Bot: The Mythopoietic narrative of unlimited possibilities

Simone D'Alessandro

simone.dalessandro@unich.it

Department of Business Administration | University G. d'Annunzio of Chieti-Pescara



Francesco Ferzetti

francesco.ferzetti@unich.it

Department of Business Administration | University G. d'Annunzio of Chieti-Pescara

## Abstract

*Interactions between a human and a Chat Bot: The Mythopoietic narrative of unlimited possibilities*

In the debate on the Large Language Models (LLM) there are three arguments that fuel the mythopoetic narrative concerning the emancipation of the artificial agent vis-à-vis the human being: 1. The conscious self-reflexive process; 2. The ability to deeply understand the 'semantics' of natural language; 3. The creative-transformative ability.

These arguments about the future self-emancipating possibilities of A.I. are being advanced by transhumanism and the current known as "Strong A.I.". However, such narratives have been disproved by a different concrete reality in which the interaction between human and chat bot remains asymmetrical: on the one hand we have a subject endowed with meaning and intentionality that is capable of changing and inventing content in an unpredictable way ; on the other hand we have an expert system that assembles content and combines expressive possibilities without fully understanding what it is doing. This article will falsify the mythopoetic construction by examining the prediction paradox, the Inverse Turing test, and the algorithmic hallucinations.

## Keywords

Creativity | Mythopoiesis | Self-Reflexivity | Semantics | Strong A.I.



## 1. Introduction

**H**uman natural language has been defined in many ways: social institution par excellence (Berger and Berger, 1972), symbolically generalised medium (Luhmann and De Giorgi, 1992), technology of the self (Foucault, 1992). From different theoretical perspectives, human natural language remains incommensurable with other animal languages, as it is endowed with a recursive syntax, a creative semantics and a pragmatic-perlocutionary dimension (Austin, 1962) that builds circular links and emergent processes between thought, language and action.

Human language is characterised by the “recursive generation principle of discrete infinity that became available in evolutionary history (...) along with Homo Sapiens” (Chomsky and Moro, 2022: 37). Recursiveness is founded on finite syntactic structures and rules, capable of generating infinite combinations. While the scientific community has accepted this difference between human and non-human animal, it remains divided with respect to the relationship between human and artificial non-human.

The rhetorical and mythopoeitic narratives of transhumanism<sup>1</sup> (Bostrom, 2016) and the 'strong I.A.' school (Dennett, 2023) fuel the utopia of the limitless possibilities of Large Language Models and machine-learning algorithms that characterise chatbots.

According to the arguments of these narratives, in the very near future it will be possible to construct artificial agents capable of faithfully reproducing the circular nexus between human thought, language and decision-making processes. Once such possibilities are reproduced, the non-human will go beyond the human being by emancipating itself from programming constraints; at the same time, certain characteristics of the human being will be able to be fully expressed by conscious technological possibilities (Bonito and Carrara, 2024).

On the contrary, according to the narrative of the proponents of the 'weak A.I.' school, there are ontologically insurmountable distinctions and limits that will never



---

<sup>1</sup> The term transhumanism has existed since 1927, coined by the biologist Julian Huxley, but only entered the discursive repertoires in 1998 when the Swedish philosopher Nick Bostrom founded, together with the British philosopher David Pearce, the international organisation called the World Transhumanist Association (WTA). The Association's manifesto starts from the principle of the ultimate abolition of any kind of suffering through the use of technology. Transhumanism considers techno-poiesis to be a process capable of emancipating the human being from the biological, projecting him into a future in which it will be possible to extract the typical characteristics of human intelligence in order to graft them or replicate them on an artificial device. The term 'post-human' was, on the other hand, coined by Jeffrey Deitch in 1992 during an exhibition that drew several artists whose works questioned the human body and identity. In 1995, with the essay *The Posthuman Condition*, Robert Pepperell elaborated an initial theoretical definition of the posthuman as a research project based on new assumptions capable of going beyond the humanistic tradition. In 2007, the neo-evolutionary sociologist John Harris introduced the concept of enhancing evolution: the evolutionary improvement of the human assisted by factors foreign to natural evolution.

allow the conditions for a total assimilation of human characteristics within the artificial non-human (Searle, 2019).

There are three arguments that feed into the mythopoietic narrative of the possibilities for the full emancipation of the artificial agent.

The first argument is the question of consciousness determining the reflexivity of the self. According to the exponents of 'weak A.I.' it is the result of an evolutionary process of thousands of years that has led the substrate of biological organisms to interconnect, leading to the modern human brain. In this view, what distinguishes the human from the artificial agent is the number of connections between the elements that make up the thinking substrate (Tononi, 2008) but also the emerging processes constituted by the inter-human relations and the relations between thought, language and action (Donati, 2022). Conscious thought is a characteristic that machines do not possess, not least because, as Searle argues, if they did possess it, they would not be able to tell us humans through a language of their own, nor would they be able to make us perceive it with our senses. If they are not able to communicate it, then they are not aware of it. If they are not aware of it, then they do not possess consciousness. Consciousness is the result of functions that produce meta-cognition, i.e.: knowing that one is 'aware of knowing' and being able to 'communicate it to another' (Searle, 2019). The reflexive awareness of acting, peculiar to the human, consists of seeing ourselves interacting with the other and returning, reflexively in retrospect, to how we acted (Donati and Archer, 2010).

According to the strong A.I. school, consciousness is an epiphenomenon that can be generated as the complexity of synaptic connections increases. For transhumanists, consciousness represents an indispensable stage of the General Artificial Intelligence (AGI) that will characterise the next technological evolution.

In both cases we observe a narrative that builds mythical expectations of the future without having scientific foundations in the present.

The second argument is based on the 'semantic' issue. According to neo-humanist exponents of weak A.I. when we speak of intelligence, it is understanding that makes the difference: an exclusive, non-algorithmic property of consciousness (Faggin, 2022: 137). In other words, artificial intelligence is not able to 'comprehend' the meaning of what it assembles, even if it is able to use it through statistical combinations. Programming languages are endowed with syntactic recursiveness - as are all formal mathematical languages derived from human language - but lack a predisposition to discern the deeper meaning of symbols. Winograd's test (Winograd, 1972) and the Winogrande test (Sakaguchi et al., 2019) demonstrate the current inability of the artificial agent to comprehend ambiguous, paradoxical sentences with no explicit subject. Artificial agents do not even understand conversational assumptions and implicatures (Faggin, 2022). Following the narrative of transhumanists and exponents of 'strong AI', this inability is temporary and strictly dependent on the possibilities of technological progress: in the future, artificial agents will be able to understand content through gradual incremental innovation processes.



This perspective does not create ontological limits to technological pervasiveness and builds a “mythological imaginary of the superiority of artificial intelligence” (Bennato, 2024: 41) without any logical argument with respect to the real evolutionary possibilities: moving from non-comprehension to comprehension represents a radical innovation, not an incremental one. At this point in time, no programmer can understand how this paradigm shift is possible.

The third argument concerns the creative-transformative capacity of the human, which, according to weak AI exponents, is inherent to the human being and differs from the exploratory creative capacity of the artificial agent: whereas exploratory creativity merely recombines previous knowledge ad infinitum, transformative creativity allows the invention of new categories that make the unexpected 'emerge' (Klien, 2022; D'Alessandro, 2024). According to the proponents of weak A.I., transformative creativity is inherent to human capacity as it is closely related to the ability to 'decide to decide' in the absence of information and data (Simon, 1955) or in the presence of ambivalent and distorted information. This ability conflicts with the logic of predictability at finite states of algorithms (D'Alessandro, 2024). The person invents new categories, and this ability is related to comprehension which concerns the discovery (from self-awareness) of new connections between meanings that “exist in semantic space and leads to new conscious meaning structures that require new symbols to be communicated” (Faggin, 2022: 139). The person invents new symbols or voluntarily decides to do away with old ones. He or she also decides to change his or her point of view by reversing, combining, separating, or elaborating paradoxes that allow him or her to go beyond established knowledge (D'Alessandro, 2023). But above all, the human being can decide to 'renounce' by deactivating a pre-established path or scheme of action. The person can decide to improvise, even without having a pre-established plan. An artificial agent - even when guided by self-adapting algorithms that learn by trial and error - creates on the basis of the algorithm's indications and on the basis of existing combinations of data: if it composes music or generates paintings or produces other forms of language, it recombines styles and assembles rules that simulate past authors and aesthetic categories (Arielli and Manovich, 2022).

Strong A.I. and transhumanism oppose the narrative of insurmountable limits, but without scientific evidence. They construct a narrative based on the following assumption: “today it is not possible, but tomorrow this limit will be overcome”.

This assumption is based on the theory of the singularity (Kurzweil, 2005): a paradigmatic shift and evolutionary enhancement (Harris, 2007) co-administered by factors unrelated to the natural evolution of humans. The 'singular' phenomena have in common that small variations in a specific evolutionary phenomenon can cause large discontinuities. In sociological terms, this means that small artificial changes, metabolised by the human being, would greatly and quickly disrupt the current condition of sociality and inter-human relations, building a discontinuity between artificial human and non-human, facilitating any form of hybridisation.

In the following paragraphs we will attempt to falsify the mythopoietic narrative of the transhumanist-strong A.I. perspective by examining - through examples of



human/chat bot interaction - the three arguments: 1. Consciousness/(Self-) reflexivity; 2. Semantic comprehension; 3 Transformative creativities.

In exposing the false myths and controversial narratives, we will start from the relational theoretical assumption that artificial agents can be endowed with agency if we give this term a 'weak' and 'non-symmetrical' definition with respect to human agency. In other words, we accept the definition of agency as the ability to make a difference, i.e. to exercise power without being directly aware of it. Both humans and artificial agents are endowed with agency, in that they perform actions that have consequences, but the character of this agency should not be understood as equivalent: "Human agents have intentions and forms of awareness that machines do not possess. The two types of agencies are separate, but intertwined, and their consequences are emergent" (Airoldi, 2024: 73-74). From this point of view, we distance ourselves from Latour's theory that puts human and non-human actors on the same level, but we recognise the artificial agent as having a social life that has similarities with Bourdieu's theories of human practice (Bourdieu, 2009).

According to the current paradigmatic understanding of the origin and significance of Western modernity, multiple processes of modernisation effectively inverted the traditional hierarchical ordering of 'the secular' (the ordinary, worldly, and the profane) and 'the sacred' (the extra-ordinary, the holy, and the venerated). This inversion, it is claimed, was established in the wake of a profound 'metaphysical shudder' that fractured the moral, cultural and political foundations of key European social formations in the period between the Renaissance and the Enlightenment. Prior to this, in pre-modern contexts, human life conceived of itself in hierophantic terms; the *hieros* being deemed both ontologically and epistemologically foundational, with the secular positioned as a secondary phenomenon; restricted and theologically conditioned in the context of Christianity as a lesser juridical reality within Augustine's fallen 'City of Man'.<sup>2</sup> In modernity, however, secularity is seen as having thoroughly profaned itself; achieving for itself a new autonomy in the attainment of an ontological, world-forming and informing, potentiality in its own right. Within modernity, it is claimed, the secular emerges as the force of history; as 'time liberated' for human purposes and plans. In modernity, secular time is no longer the time between the incarnation and the *eschaton*. On the contrary, in modernity the secular now presents itself as the new infinitude of 'historical progress' founded upon the radical opening of the time horizon; as an empty, undecidable, yet projectable and malleable, human-centred, futurity. As such, within secular temporality, the distinction between an unalterable (but knowledge-worthy)



<sup>2</sup> However, in pre-modern contexts, although it is widely argued that the sacred was primary both ontologically and politically, the sacred and the profane were deemed to be conceptually and existentially inseparable domains; with the secular being defined in terms of the sacred, the former being viewed as a degenerate albeit outgrowth of the latter. This can again be seen in Christianity, where the sacred was no longer located in people or things, but in the dimensions of time, to the extent that the secular itself was understood as simply the time between the two historical sacralities of the incarnation and the *eschaton*.

past and an open future occupies the place once occupied by the immanent and the transcendent, of a 'here' and a 'beyond' (see Kossellneck, 2005: 212).

However, this, orthodox, conception of secularity is itself founded upon a series of often unquestioned and unquestionable assumptions. In this paper, I intend to interrogate some of these assumptions, in order reveal the presence of a concealed sacrality *within* the secular. Indeed, I aim to show that what philosophers and social theorists refer to as 'secularity' is simply a theological mutation within the Judeo-Christian theological imaginary; one that we can, in its dis-incarnational rejection of nature over grace, following Voeglin, position as 'Gnostic' (see Voegelin, 2012).

In this vein, I will suggest that the sacred within the secular is fundamentally acosmic; residing beyond the realm of nature, which is now handed over to the dark fate of impersonal mechanism, as a series of spiritually blind causes within a larger order of causal necessity. In its radical separation of sacred from the cosmos, the secular relocates the sacred within the infinitude of the subject, within a realm of radical freedom, where the self no longer has any spiritual communion with things but only with other selves, to the extent that politics itself takes on a heightened spiritual dimension (see Jonas, 2001). In making this claim, I will endeavour to show that what at first glance appears to be the waning of sacrality of modern contexts, in reality is simply an effect of a hard-to-discern *transformation*. From a conception of sacrality primarily located in the outer realms of people and things, to one found in the relative immateriality of personal and political discourse - from 'the invisible in natural things', to 'the visible in political words'.

In what follows, my approach will be deliberately and self-consciously theoretical. In the present era, when various forms of bland empiricism and the ideology of scientism blind us to any need for new ideas and alternative ways of imagining the social-historical, it is essential that work in the Humanities again begins to move along critical and speculative paths. For 'facts without theories are blind' – and it is only the vision of the theorist that is able to transform the crises of our age into something thought-worthy and, ultimately, able to render our responses choice-worthy. In so doing, I will make the theory-driven claim that in the passage to modernity the sacred was sublimated into a form where it became abstracted and further esoterised as it was uprooted from its cosmological groundings and rearticulated into the modern imaginaries of politics and law. My key claim will be that within modern contexts the 'sacrality of the secular' ultimately resides within a new type of political textuality – in the idea of a sacred constitution that manifests itself in the utopian recognition and obligation of a people with/to itself and to all other peoples. In other words, I will suggest that in modern contexts, the sacred becomes disembedded within the futurity of the symbolic orders of the political and is therefore neither diminished nor diluted but simply relocated and rearticulated within founding political texts – texts that possess quite specific ontological, world-transforming, effects. In this way, I will claim that within secular modernity the sacred emerges as a geo-political ideal of universal freedom that presents itself as necessary *a priori* political truth that applies universally to all peoples, all places, and for all time. It is in this sense that the modern sacred retains its link with the *eschaton* via its



implicit connection to the ideal of an end of history, the time when the moral and political obligations of the founding political text will be fulfilled and realised for all peoples - as the political eschaton's full worldly realisation.

## 2. The Mith of digital self-reflexivity: in search of a new oracle

Self-reflexivity allows one to experience oneself as the subject of one's own existence in relation to an 'alter' who enters into relationship within a context. According to current international research, algorithms cannot 'think' or 'think about thinking': dimensions that are biunivocally implied (Searle, 2019). A chat bot driven by a Large Language Model (LLM) cannot be conscious of being a subject of its own existence in relation to an alter but it can "participate in communication (...) bots are authors of about 50 per cent of online traffic" (Esposito, 2022: 9).

It is appropriate to establish the differences between human thought and the machine-learning algorithm that characterises the large language model (LLM) of a chat bot, in order to deactivate a narrative that risks constructing a distorted collective vision: the analogy between the performance of algorithms and the conscious capacity of human intelligence is misleading (Esposito, 2022: 2).

The possibility of conversational exchange and content transfer between human and non-human does not necessarily imply the concept of intelligence, nor that of consciousness. The relationship takes place if there is a willingness on the part of the human being to believe and trust in the meaningfulness of the chat bot's arguments.

The intentionality and reflexivity regarding the decision to give credence and trust to the contents of the artificial agent depend on the inner conversation of the human being (Archer, 2006). From this point of view, the relationship is asymmetrical: the chat bot has no trust or willingness to believe in the conversation constructed with the human. The chat bot merely constructs the process, while the human decides to 'believe' (or not believe) in the process it is constructing.

This asymmetry is particularly evident when the human interrogates the chat bot, trusting in its ability to anticipate future themes and trends. Prediction takes place through a trust mechanism that the human being possesses, but which the chat bot lacks. A chat bot (such as Gemini or Gpt 3.5) assembles data entered by the human, regardless of whether the other (human) will (or will not) trust the conversation data. Such data represent what has already happened, but not what has yet to happen. It is the trust relationship established by the human being that determines the actual reification of a prediction data. We are witnessing a mythopoietic phenomenon: the narration of the predictive capacity of artificial intelligence. This phenomenon has been described by Helga Nowotny as the "prediction paradox" (Nowotny, 2022: 19).

If human behaviour conforms to the bot's predictions, the future comes true in a deterministic way:

The predictive power of algorithms gives us the ability to see further ahead (...)  
The way we do this, however, is paradoxical: We yearn to know the future, but



we largely ignore what the predictions do to us (...) The paradox arises from the incompatibility between an algorithmic function, such as an abstract mathematical equation can be, and a human belief that might be strong enough to move us to action (Nowotny, 2022: 20).

The predictive power of algorithms can turn into a self-fulfilling prophecy. But the prediction becomes true because people decide to believe and act accordingly, avoiding changing the course of the future while having the reflexive ability to boycott it. In fact, not all prophecies triggered by mass communication come true, as human beings possess an unpredictable ability to self-boycott their own beliefs.

However, with the constant increase of predictive algorithms, Thomas' theorem<sup>3</sup> runs the risk of becoming an automatic practice. Thus, algorithms can model the future behavior of humans through past data. In this sense, Esposito pays attention to some similarities between ancient divination practices and predictive algorithms (Esposito, 2022): we are favoring a world that is determined from the past, simply because we have reflexively decided not to question the data analysis processed by the algorithmic oracle. This fate does not depend on the will of an artificial agent, but on the reflexive intentionality of a human being who, at any time, can decide to change his or her opinion and interpretation. In the next paragraphs we will examine the Inverse Turing test and some examples of conversation between researchers and Chat BOTS. Cases that show the circular links between self-reflexivity, semantic comprehension, and human transformative creativity, highlighting further limitations of A.I. with respect to such arguments.



### 3. The false narrative of semantic comprehension

Machine learning algorithms are efficient because programmers have abandoned the attempt to simulate human intelligence, artificially reproducing only the ability to make the chat bot 'believable' in interaction with the person. The artificial agent participates in conversations by simulating an interaction, but not a meaningful relationship: this is the real and pragmatic narrative proposed by the programmers; a storytelling that has been distorted by the narrative of transhumanism. The ability to understand the nuances of a sentence or the ability to disambiguate are inherently human characteristics.

The current Inverse Turing test<sup>4</sup> highlights some of the semantic inability.

CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart, also referred to as Inverse Turing Test) recognizes a human

---

<sup>3</sup> In 1928 Thomas states, "If men define certain situations as real, they are real in their consequences." Assertion reformulated by Robert K. Merton in the terms of self-fulfilling prophecy, in *Social Theory and Structure*, II, Bologna, Il Mulino, 1971.

<sup>4</sup> From the 1950s to the present, numerous tests have been developed that can unmask the artificial agent: from the classical Turing test to Searle's Chinese Room; from the Silence test to the Reverse Turing; from the Lovelace test on creativity to the Winograd test on ambiguous sentences.

interaction by distinguishing it from a non-human one. The test protects human users from spam and password decryption. The CAPTCHA test has two parts: a randomly generated sequence of letters and/or numbers displayed in a distorted image and a text box. To pass the test and prove that they are human, the user must type the distorted symbols into the text box and recognize what the test asks for.

The artificial agent fails in the task of reconstructing distorted images. It needs corrective actions by the human to guide its exact response. Evidence of the deterministic fallibility of the artificial agent comes, moreover, from computer programs used to unmask other programs that generate fake news<sup>5</sup>: the artificial agent is able to unmask the action of another artificial agent, but it cannot distinguish an artificial user from a human being. This is also why 21st century programmers do not try to emulate human intelligence (Solon, 2012; Esposito, 2022).

For example, programmers of Libratus - an AI expert system that defeated top human players in January 2017 - show that A.I. even when it wins develops a totally different game strategy than humans: "complex algorithms work in well-defined situations (...) while human intelligence has evolved to handle uncertainty regardless of the amount of data available" (Gigerenzer, 2023: 57).

Artificial agents win or lose differently than humans. Artificial intelligence systems process automated decision-making processes: as such, they are improvable in the learning phase, but limited in exploring the unknown because they are caged by the storage of previously accumulated data. We need to rethink the way we observe its interaction with humans. The simulation of a relationship is believable, but it is not true. It is structured on the level of logic, but fragile when it has to reason in metalinguistic, paradoxical and ambivalent terms. The artificial agent can handle risk but cannot deal with uncertainty. In situations of risk, it is possible to know possible outcomes in advance and calculate probable consequences.

In situations of uncertainty, it is impossible to know all possible outcomes in advance (Knight, 1921). Human beings make decisions even when the scenario is undecidable. The person always has the will to unlock a problematic situation and make a decision, even a casual one. Because of this central difference, we cannot speak of authentic conversation between human and nonhuman, nor of relationship, but of simulated interaction.

The human being projects his intelligence, his relational attitude and his reflective vocation onto whatever artefact he is able to interact with. This demonstrates two great capacities that are properly human: the metaphorical one (which builds links between distant worlds) and the empathic one (which allows to empathise with the *alter*, even when it is not human). But this gift also hides a vulnerability, which consists in the desire to be deceived by the AI in order to be reassured and to confirm one's own self-deception on metaphorical projection (Natale, 2022).



---

<sup>5</sup> For example, the Giant Language Model Test Room (GLMtr). See: [www.agendadigitale.eu/cultura-digitale/intelligenza-artificiale-contro-le-fake-news-lo-stato-della-ricerca](http://www.agendadigitale.eu/cultura-digitale/intelligenza-artificiale-contro-le-fake-news-lo-stato-della-ricerca)

These latter considerations are related to the algorithmic hallucinations and creative limitations that we will consider in the following paragraph.

#### 4. The (unfulfilled) promise of transformative creativity

Roberto Presilla reports the example of a professor who asked Chat GPT, the logical sentence analysis of some impersonal expressions, demonstrating the bot's inability to perform proper logical analysis<sup>6</sup>.

The bot fails if it is not guided by the human user through successive trial and error. The chat bot does not infer from context an impersonal expression that refers to an unspoken subject. The tool is unable to contextualize and disambiguate independently on the first try, often not even after numerous repetitions.

Thierry Vissol, on April 8, 2023, reports in an article published in Page 21, his failed experience as a scholar trying to interact with Chat GPT for research purposes:

I was doing research on the Upper Palaeolithic (...) I decided to test to what extent Chat GPT could help me both in the various topics of my research and in searching bibliographies. The results obtained immediately seemed (...) valid. The topic summaries were short, but coherent (...) So I continued for more than a day to ask more and more precise questions, getting quick and documented answers (...) within seconds. I began to have doubts when I systematically asked for summaries of articles or books, which appeared extremely similar to each other, but formulated differently. Such homogeneity of various authors on controversial topics seemed very strange to me<sup>7</sup>.

After numerous cross-checks, the researcher discovered that most of the references provided by Chat GPT did not exist. The information had been invented by the artificial agent. The authors cited were not specialists or had never written the articles cited. The publishers had not published the cited books which had never been published by other publishers either.

These examples show the inability of the bot to give up research when the information does not exist or does not adequately help the interlocutor who requested the data. The phenomenon has been called algorithmic hallucination: "when we ask a chatbot for the definition of something that does not exist, and the program gives us a credible but false document"<sup>8</sup>.

<sup>6</sup> R. Presilla, *ChatGPT non è intelligente ma si applica*, <https://rivista.vitaepensiero.it/news-vp-plus-chatgpt-non-e-intelligente-ma-si-applica-6096.html>, 25, febbraio, 2023.

<sup>7</sup> T. Vissol, *La mia esperienza con Open Ai Chat Gpt*, [www.pagina21.eu/la-mia-esperienza-con-openai-chat-gpt-una-truffa-pericolosa/thierry-vissol/](http://www.pagina21.eu/la-mia-esperienza-con-openai-chat-gpt-una-truffa-pericolosa/thierry-vissol/), 8 aprile 2023.

<sup>8</sup> N. Klein, N. *Le allucinazioni del capitalismo*, [www.internazionale.it/magazine/naomi-klein/2023/06/08/le-allucinazioni-del-capitalismo](http://www.internazionale.it/magazine/naomi-klein/2023/06/08/le-allucinazioni-del-capitalismo), 26/08/2023.



The term 'hallucination' instead of the more correct expression 'system error' fits into that process of mythopoetic storytelling that evokes a supposed artificial subject endowed with consciousness and capable of stepping out of the paths of reason through hallucinations.

This anthropomorphization process had already been constructed in the past with other terms: 'artificial intelligence' replaced 'expert system'; 'semantic web' replaced 'adaptive search engine.' Misleading terminologies that construct an imagery meant to mirror human faculties, proposing a digital doppelgänger.

No programmer has been able to solve the anomalies caused by algorithmic hallucinations. Sundar Pichai, A.D. of Alphabet, admits that it is impossible to avoid these system errors caused by the chat bot's inability to decide to give up. More than 40 percent of responses provided by chatbots are incorrect or anomalous<sup>9</sup>.

The chat bot does not go outside the programmed request process, so it must necessarily give an answer, even if it is wrong. While the human has the intentionality and willingness to give up, sabotage the request or go off topic.



## 5. Conclusions. Chat bots and mythopoetic narratives

From the considerations considered in the preceding paragraphs, a counter-narrative emerges that downsizes the mythopoetic narrative of strong AI and transhumanism. We remain fundamentally distinct and unreplicable persons from the capabilities of A.I. (Osborne and Rose, 2023:1). A chat bot can establish a simulated conversation with a human being, but this interaction is limited on the level of reciprocity in the following dimensions that manifest a circular relationship between them:

- *Dimension of self-reflexivity.* Trust, intention to believe, expectations, and predictive paradoxes prove, indirectly, the human's self-reflexive ability to accept or question what happened during a conversation. The bot can reconfigure parameters, assigning new measures to the words produced by the exchange, but it cannot decide to remove or go off-topic to provoke a contingency; nor can it interrupt a discourse deemed unsuccessful by the interlocutors; nor can it converse inwardly to revise planned decisions.

- *Dimension of semantic comprehension.* The bot assembles content, but is incapable of understanding, disambiguating, and contextualizing it. Humans pose critical considerations that allow them to dialectically dissect an argument by fragmenting it, reversing it, combining it with other elements even off-topic, to the point of paradoxically playing with the argument itself to reclassify it. Humans understand paradoxes, implicatures, subject-less or distorted sentences. Chat bots,

<sup>9</sup> [www.tomshw.it/altro/chatgpt-sbaglia-il-50-delle-domande-di-programmazione-altro-che-rubare-il-lavoro/](http://www.tomshw.it/altro/chatgpt-sbaglia-il-50-delle-domande-di-programmazione-altro-che-rubare-il-lavoro/) 27/09/2023.

while adapting through trial and error to human demands, are incapable of contextualizing, disambiguating, and understanding in the full meaning of the term. Knowing how to use a language by assembling its terms stochastically does not, necessarily, imply the ability to comprehend.

• *Creative-transformative dimension.* It follows from the preceding point that the human ability to 'play' with plausible interpretations, allows for transformative and not merely exploratory creativity. Along this path, even the communicative process involved in the intentionality of the human differs from the interactive simulation of the artificial nonhuman. The intentionality of the human involves a double contingency consisting of expectations that make the conversational outcome with another person unpredictable. This is not the case with an algorithm that meets the expectations of its interlocutor regardless of the degree of truthfulness of content. Finally, human creativity is intersubjectively formed, "as a result of the interaction between the experiences of individual social actors" (Pedroni, 2005: 459) and is the emergent outcome of widespread educational and social practices (Glăveanu, Hanson and Baer, 2019).



The emerging relationship between the capacity for semantic understanding, transformative creativity, and self-reflexivity determines the difference between the generative unpredictability of the human relationship and the recursive cage of human/chat bot interaction. When transhumanism and strong A.I. propose "non-humanistic conceptions of the human condition, in which anthropology is replaced by anthropotechnics" (Maccarini, 2021: 192) they fall into a pseudo-scientific fallacy, re-proposing a Cartesian dualism (*res-cogitans/res-extensa*), projected into a future in which it will be possible to extract the typical characteristics of human intelligence and graft them onto a non-human device. However, human intelligence is embodied: dependent on the human body. This is demonstrated by scholars in cognitive science (Noë, 2010), neurology and social neuroscience (Ehrenberg, 2019). If we eliminate the body, we also subtract creative processes from the cognitive system: "all abilities, even the most abstract, arise as bodily practices" (Sennett, 2009: 19). But then, why do humans continue to have the need to build narrative imaginaries about AI that are so distant from the actual and concrete possibilities given by reality?

There is a propensity to describe AI with both overly optimistic or pessimistic tones, which confirms a long-term trend (...) In other words, both utopian and dystopian narratives trace back to the main recurrent hopes and fears connected to technology (Sartori and Bocca, 2023: 446).

The narrators of infinite possibilities underestimate these factors, constructing a mythopoiesis that would like to eliminate the ontologically irreducible distinctions between human intelligence and A.I. in order to propagate a supposed conscious autonomy of A.I. capable of transcending human capacities. A return to a dogmatic scientism: a *déjà vu* that recurs every time an innovation is implemented, where the rhetoric of infinite possibilities evolves into a techno-religion.

## Bibliography

- Airoldi M. (2024), *Machine Habitus. Sociologia degli algoritmi*, Roma, Luiss University Press.
- Archer M. S. (2006), *La conversazione interiore. Come nasce l'agire sociale*, Milano, Erickson.
- Arielli E., Manovich L. (2022), "AI-aesthetics and the anthropocentric myth of creativity", *Nodes*, 19, (20): 91-97.
- Austin J. L. (1962), *How to do things with words*, Oxford, Clarendon.
- Bennato D. (2024), *La società del XXI secolo. Persone, dati, tecnologie*, Bari-Roma, Laterza.
- Berger L., Berger B. (1972), *Sociology: a biographical approach*, New York, Basic Books.
- Bonito C., Carrara A. (2024), *Il transumanesimo. Una sfida antropologia alla scienza e alla fede*, Milano, Mimesis.
- Bostrom N. (2016), *Superintelligence*, Oxford, Oxford University Press.
- Bourdieu P. (2009), *Ragioni pratiche*, Bologna, il Mulino.
- Chomsky N., Moro A. (2022), *I segreti delle parole*, Milano, La nave di Teseo.
- D'Alessandro S. (2024), "The predictable and disabling society: artificial intelligence and changing Human Agency", *Sociétés Revue de Sciences Humaines et Sociales*, 163 (1): 73-84.
- D'Alessandro S. (2023), "Creative Flows: Constructions of Meaning between Binary Oppositions, Paradoxes and Common Sense", *Italian Sociological Review*, 13, (3): 371-392.
- Dennett D. (2023), *Coscienza. Che cosa è*, Bari-Roma, Laterza.
- Donati P., Archer M. (2010), "Riflessività, modernizzazione e società civile", *Sociologia e Politiche Sociali*, 13, (1).
- Donati P. (2022), *La teoria relazionale nelle scienze sociali: sviluppi e prospettive*, Bologna, il Mulino.



Simone D'Alessandro, Francesco Ferzetti  
*Interactions Between a Human and a Chat Bot: The Mythopoeitic Narrative*

Esposito E. (2022), *Comunicazione artificiale. Come gli algoritmi producono intelligenza sociale*, Milano, Bocconi University Press.

Ehrenberg A. (2019), *La meccanica delle passioni. Cervello, comportamento, società*, Torino, Einaudi.

Faggin F. (2022), *Irriducibile. La coscienza, la vita, i computer e la nostra natura*, Milano, Mondadori.

Foucault M. (1992), *Tecnologie del sé*, Torino, Bollati Boringhieri.

Gigerenzer G. (2023), *Perché l'intelligenza umana batte ancora gli algoritmi*, Milano, Raffaello Cortina.

Glăveanu V. P., Hanson M. H., Baer J. (2019), Advancing Creativity Theory and Research: A Socio-cultural Manifesto, *Journal of Creative Behavior*, 51, (3): 741-745.

Harris J. (2007), *Enhancing Evolution*, Princeton, Princeton University Press.

Klein S. (2022), *Come cambiamo il mondo. Breve storia della creatività umana*, Torino, Bollati Boringhieri.

Klein N. (2023), *Le allucinazioni del capitalismo*, [www.internazionale.it/magazine/naomi-klein/2023/06/08/le-allucinazioni-del-capitalismo](http://www.internazionale.it/magazine/naomi-klein/2023/06/08/le-allucinazioni-del-capitalismo), 26/08/2023.

Knight F. H. (1921), *Risk, Uncertainty and Profit*, Boston, Houghton Mifflin.

Kurzweil R. (2008), *La singolarità è vicina*, Milano, Apogeo.

Luhmann N., De Giorgi R. (1992), *Teoria della società*, Milano, Franco Angeli.

Maccarini A. (2021), "Post-umano", in Marci, T. e Tomelleri S. (a cura di), *Dizionario di Sociologia per la persona*, Milano, Franco Angeli: 191-193.

Merton R. K. (1971), *Teoria e Struttura Sociale*, Bologna, il Mulino.

Natale S. (2022), *Macchine ingannevoli. Comunicazione, Tecnologia, intelligenza artificiale*, Torino, Einaudi.

Noë A. (2010), *Perché non siamo il nostro cervello. Una teoria radicale della coscienza*, Milano, Raffaello Cortina.



Nowotny H. (2022), *Le macchine di Dio. Gli algoritmi predittivi e l'illusione del controllo*, Roma, Luiss University Press.

Osborne T., Rose N. (2023), "Against Posthumanism: Notes towards an Ethopolitics of Personhood", *in Theory, Culture & Society*, 40, (3): 1-19.

Pedroni M., (2005), "Genio individuale o pratica sociale? Le mille facce della creatività", *Studi di Sociologia*, 4: 439-461

Presilla R., (2023), *ChatGPT non è intelligente ma si applica*, <https://rivista.vitaepensiero.it/news-vp-plus-chatgpt-non-e-intelligente-ma-si-applica-6096.html>, 25, febbraio, 2023.

Sakaguchi K., Le Bras R., Bhagavatula C., Choi Y. (2019), *WINOGRANDE: An Adversarial Winograd Schema Challenge at Scale*, Allen Institute for Artificial Intelligence, University of Washington, <https://arxiv.org/pdf/1907.10641.pdf>

Sartori L., Bocca G. (2023), "Minding the gap(s): public perceptions of AI and socio-technical imaginaries", *AI & Society*, 38: 443-458.

Searle J. (2019), *Il mistero della realtà*, Milano, Raffaello Cortina.

Sennett R. (2009), *L'uomo Artigiano*, Milano, Feltrinelli.

Simon H.A. (1955), "A Behavioral Model of Rational Choice", *The Quarterly Journal of Economics*, 69, 1: 99-118.

Solon O. (2012), *Weavrs*. "The autonomous, Tweeting Blog-Bots That Feed", *Social Content Wired*, 28/03/2012.

Tononi G. (2008), "Consciousness ad integrated information: a provisional manifesto", *Biological Bulletin Journal*, 215: 216-242.

Vissol T. (2023), *La mia esperienza con Open Ai Chat Gpt*, [www.pagina21.eu/la-mia-esperienza-con-openai-chat-gpt-una-truffa-pericolosa/thierry-vissol/](http://www.pagina21.eu/la-mia-esperienza-con-openai-chat-gpt-una-truffa-pericolosa/thierry-vissol/) 8 aprile 2023.

Winograd T. (1972), "Understanding Natural Language", *Cognitive Psychology*, 3, (1): 1-191.

[www.agendadigitale.eu/cultura-digitale/intelligenza-artificiale-contro-le-fake-news-lo-stato-della-ricerca](http://www.agendadigitale.eu/cultura-digitale/intelligenza-artificiale-contro-le-fake-news-lo-stato-della-ricerca), 10/06/2023.

[www.tomshw.it/altro/chatgpt-sbaglia-il-50-delle-domande-di-programmazione-altro-che-rubare-il-lavoro](http://www.tomshw.it/altro/chatgpt-sbaglia-il-50-delle-domande-di-programmazione-altro-che-rubare-il-lavoro), 27/09/2023.







A JOURNAL OF THE  
SOCIAL IMAGINARY

# Tecnotopie. Per una epistemologia simbolica degli algoritmi intelligenti

Edmondo Grassi

Edmondo.grassi@uniroma5.it

Telematica San Raffaele Roma



## Abstract

*Technotopias: Toward a Symbolic Epistemology of Intelligent Algorithms*

The interaction between intelligent algorithms and human entities raises epistemological questions that challenge the traditional sociological distinction between individual and tool, posing issues of significant relational relevance. In this context, algorithms can be seen as social agents that embody and reflect specific configurations of power and knowledge, implying a continuous renegotiation of social norms and cultural expectations. This contribution aims to reflect on the evolutionary lines of the imagination linked to such technologies, which not only embody human pragmatic knowledge but also evoke an imaginative capacity central to the transfiguration of the technological object into an artifact with an almost autonomous potential, capable of integrating and interacting in complex ways with surrounding societies. This outlines a perspective on the impact of AI in the construction of reality, highlighting ethical challenges and poetic potentials.

## Keywords

Algorithm | Identity | Imaginary | Technology | Technotopia



## Densità immaginifica e società tecnostrutturate

La macchina è ciò attraverso cui l'uomo si oppone alla morte dell'universo; essa rallenta, come la vita, la degradazione dell'energia e diventa stabilizzatrice del mondo.

Gilbert Simondon (2017: 9)



**N**ella costruzione del reale, il ruolo dei dispositivi tecnoscientifici può essere interpretato attraverso la lente e lo specchio per mezzo dei quali ogni artefatto *umanizzato* è permeato da grammatiche relazionali che scrivono e regolano i rapporti di interdipendenza tra individui, così come con e tra enti non umani. La storia della specie umana può essere narrata attraverso la declinazione delle molteplici tecnologie che si sono succedute nel corso dei secoli, quali entità complesse e poliedriche che intrecciano le proprie traiettorie esistenziali con quelle umane, modellandone condizioni materiali, cognitive e identitarie. Il punto di vista adottato è di una visione simbiotica, in cui tecnoscienze ed enti umani sono concepiti sia in termini di coesistenza che di coevoluzione.

L'universo strutturale assemblato dai dispositivi sinora prodotti riflette l'idea che la tecnologia è un modo politico di costituire un proprio ordine all'interno delle società umane (Winner, 1988), stabilendone la non neutralità, in quanto sempre laboratori di specifici paradigmi valoriali e culturali (Mumford, 2010) in grado di modellare le capacità di agire e interagire degli individui attraverso specifici approcci: essi assumono un ruolo più capillare e articolato di meri estensori e potenziamenti delle abilità umane, in quanto divengono mediatori critici che ristrutturano ed elaborano modelli di comportamento, sviluppi identitari e soggettività valoriali, sino alla generazione di proiezioni immaginifiche sempre più legate alla loro manifestazione nel reale (Grassi, 2021). Da questa prospettiva, la tecnologia rappresenta un sistema di regolamentazione complessiva dell'esistenza umana, in quanto forza storicamente dominante, apparato totale che regola ogni aspetto della vita (Ellul, 2009) in una costante rinegoziazione e modificazione delle strutture e di tensioni e disallineamenti a causa del ritardo culturale tra progresso e strutture sociali (Ogburn, 2006).

La doppia natura della tecnologia - quella di mezzo e di fine - sottolinea il suo ruolo sempre più incisivo nel cambiamento del sistema teleologico dell'essere, un sistema che si trova in una fase di trasformazione radicale. Essa è chiamata a produrre materia artificiale che non solo coesiste con quella biologica, ma che sia anche in grado di connettersi e integrarsi con essa in modi che ne amplificano le capacità, in un processo di potenziamento reciproco. Tale ricerca di integrazione e potenziamento è, da sempre, intrecciata con i bisogni e le volizioni dell'individuo che

si esprimono per mezzo della sua capacità immaginifica che funge da matrice per le proprie creazioni culturali (Castoriadis, 1995), in quanto non sono solo risultati di necessità funzionali o di impulsi economici ma anche incarnazioni di desideri, paure, miti e aspirazioni collettive (Wajcman, 2004; Flichy, 2008; Lemov, 2015).

Mentre l'immaginario influisce sullo sviluppo dei simboli, dei miti, così come della tecnologia, attraverso cui una cultura esprime la propria visione del mondo (Durand, 1972; Strauss, 1980), allo stesso tempo le innovazioni emergenti possono a loro volta trasformare l'immaginario sociale, introducendo nuovi modelli, icone, rappresentazioni e persino nuove cosmogonie<sup>1</sup> per pensare e percepire la realtà, sino ad una ipotetica gnosi tecnologica (Davis, 2023).

Nella contemporaneità, l'idea di autonomia umana viene così ridefinita attraverso la continua interconnessione con dispositivi tecnologici che possono anticipare, modificare e sostituire le azioni umane con algoritmi complessi che tracciano nuove orbite e si insinuano in interstizi epistemologici, ontologici ed escatologici sino ad ora non percorsi. I dispositivi tecnoscientifici hanno dunque la capacità di essere il riflesso delle tensioni che popolano l'immaginario ma anche di essere tracce dotate di materia in cui si addensa il vissuto collettivo: è la gestazione di simboli arcaici e di forze mitiche costituiti da speranze, illusioni e utopie.

Perché l'invocazione dei miti persiste come un elemento imprescindibile nelle narrazioni umane? Qual è la ragione per cui ogni generazione proietta nel tessuto dell'immaginario collettivo le proprie pulsioni? In che modo il regno dell'immaginifico contribuisce alla rappresentazione e alla materializzazione del futuro?

La realtà come costruzione sociale è un continuum che va dal dominio materiale e tangibile a quello immateriale e simbolico della vita quotidiana, creando uno spazio in cui significato esistenziale e prospettiva immaginifica si fondono per mezzo di una prospettiva olistica che riflette l'interdipendenza tra i differenti ambiti. Qui, il reale e l'immaginario si intrecciano costantemente, formando un unico tessuto dell'esperienza umana (Berger & Luckmann, 1969; Secondufo, 2007; Douglas, 2013).

In tale contesto, l'individuo può essere indagato e concepito come quel soggetto che con l'ausilio delle estensioni tecnologiche permette al biologico di essere sociale e di concepire il corpo come una costruzione collettiva incardinata nell'utilizzo della tecnologia e delle sue proiezioni immaginifiche (Haraway, 2018; Braidotti, 2019).

Considerando ciò, non è più possibile aderire a un modello antropocentrico escludente, in quanto diviene essenziale riconoscere lo spazio sociale come un ambito pervaso da algoritmi intelligenti (*algoagenti* da questo momento), capaci di acquisire progressivamente una posizione sempre più di soggetto, di attore sociale (Grassi, 2020).

L'incidenza degli algoagenti sulla vita quotidiana è tale che le pratiche esistenziali dell'individuo vengono modificate continuamente in risposta alle loro interazioni e la



<sup>1</sup>Cfr. Come in *Neuromancer* di Gibson, che intreccia realtà e cyberspazio; nella tetralogia *Matrix* delle sorelle Wachowski, che esplora realtà alternative controllate da IA; nei movimenti ecotecnologici: solarpunk, biopunk e technogaianismo che traducono l'immaginario in pratiche quotidiane.

misura dell'essere non è più quella del singolo in quanto unità discreta ma diviene cenosi tra una molteplicità di specie, nella quale la persona dovrà contemplare la necessità di allearsi con altri enti (Haraway, 2020) e, tra queste creature, vi sono anche gli algoagenti e i sistemi complessi di I.A., rappresentando un'ulteriore proposizione di modernità multiple (Eisenstadt, 2016).

La costellazione di enti non umani fondati su pratiche culturali algoritmiche è edificata su proprietà sistemiche che riorganizzano i modi di interazione, conferendo una stratificazione di procedure mutevoli, adattive e permeabili. Nel quadro della teoria della complessità (Urry, 2003; Dominici, 2011; Morin, 2011), ciò pone in evidenza la profonda interdipendenza e le dinamiche emergenti dei sistemi sociali modellati dalla galassia artificiale, dei suoi risvolti sulla riflessione teorica sociologica e sul grado di penetrazione immaginifica quale membrana connettiva appartenente alla costruzione del reale: la densa interconnettività tra soggetti che trascendono i confini geopolitici; l'informazione immersiva che conduce allo sviluppo di nuovi modelli di comportamento; le contro-azioni imprevedibili e non lineari indotte da parte di enti non umani; l'ultra-accelerazione del ritmo di cambiamento globale; la disarmonica capacità di adattamento dei sistemi culturali e il diverso accesso a specifiche tecnoscienze. I sistemi algoritmici, in tal modo, possono raggiungere punti di biforcazione dove minime variazioni nelle condizioni iniziali possono portare a risultati radicalmente diversi, ridisegnando la struttura escatologica dell'essere.

L'obiettivo è di delineare le cornici, le caratterizzazioni e le prospettive etiche di questo mutamento sociotecnologico avvalendosi dello strumento della produzione immaginifica - dal mito alla serie televisiva-, quale membrana vitale per la costruzione della realtà condivisa, fondato su un approccio teorico-storico. I casi studio di riferimento sottostanno alla scelta del campionamento intenzionale/per giudizio/ragionato (Yin, 2013; Charmaz, 2014; Creswell, Creswell, 2018), basato sulla selezione di unità che sono ritenute più rappresentative e informative per gli scopi di tale ricerca.



## **Pandora e la dimensione artificiale del mito classico**

L'interazione con gli algoagenti offre lo spazio per un interscambio di realtà condivise in cui le proiezioni delle parti si fondono in un continuum di significati e costituiscono una rimodulazione della comunità immaginate (Anderson, 2018) la quale suggerisce che gli individui, pur non conoscendosi direttamente, possono condividere un senso di appartenenza radicato nell'immaginario collettivo. In tal modo, la costruzione sociale della realtà è un fenomeno intrinsecamente legato alla capacità umana di reinventare il mondo e, secondo questa prospettiva, l'I.A. può essere indicata come una tecnologia antica quasi quanto la nascita del mito (Cohen, 1981; Wilson, 2006; Riskin, 2007; Grassi, 2020).

Già Esiodo narra il mito di Talos, un protoautoma di bronzo - forgiato da Efesto, il dio artigiano- le cui funzioni vitali erano sostenute dall'icore che circola nel corpo artificiale grazie a un complesso sistema idraulico. Il suo compito era di proteggere

l'isola di Creta da qualsiasi pericolo ma Medea, grazie ad un sortilegio, ne compromise le funzioni vitali fino allo spegnimento. Analizzando questo mito, emerge già uno degli elementi di riflessione sociologica rispetto al ruolo della tecnologia nelle società umane: la rappresentazione di una delle prime incarnazioni del concetto di sorveglianza automatizzata. Ad oggi, la tecnologia ha amplificato notevolmente le capacità di monitoraggio e di controllo, sollevando questioni cruciali sulla profilazione, sulla privacy e sulle libertà individuali (Nissenbaum, 2009; Lyon, 2018); la perdita dell'icore e la sua manipolazione magica sono elementi che trovano parallelismi con la corrottibilità, con l'improvvisa deviazione, con la fallibilità nello svolgimento di prompt di un algoritmo a causa di un virus esterno, di una variabile digitale o di una falla nel sistema (Noble, 2018; Benjamin, 2019). La questione di come la tecnologia modella e viene modellata in rapporto alla società è centrale, in quanto, bisogna confrontarsi con decisioni relative a quali dispositivi adottare e come regolarli, equilibrando innovazione, rischi e responsabilità (Beck, 2001; Pasquale, 2015).



Analizzando le strutture sociali e i modelli relazionali, *Snow Crash* (1985) è particolarmente affascinante per il modo in cui intreccia linguistica, mitologia e informatica. Il *metavirus* descritto nel romanzo è basato su un antico linguaggio sumero, considerato la lingua della creazione e del comando. Secondo la narrazione, questo codice primordiale possiede una struttura binaria intrinseca, simile a quella del codice informatico, capace di riprogrammare il cervello umano: ciò lo rende in grado di bypassare i normali processi cognitivi e di influenzare direttamente le percezioni delle persone. L'opera può essere interpretata come un'incarnazione della narrazione metalinguistica, dove il linguaggio non è solo un mezzo di comunicazione ma diviene un'entità con il potere ontologico di definire e modellare la realtà stessa, come sta avvenendo con l'uso degli algoagenti. Stephenson sembra proporre una visione escatologica in cui l'evoluzione umana è strettamente connessa all'integrazione con la tecnologia: questa evoluzione culmina in una forma di esistenza fondata sull'ibridazione umano-macchina e dall'espansione dello spazio sociale nel metaverso. In tal modo, *Snow Crash* diventa un terreno fertile per indagare le tensioni sociali, esplorando i confini sempre più labili tra realtà e simulacro nel contesto contemporaneo, giungendo a proporre una piena alterazione delle categorie newtoniane di spazio e di tempo, superando i limiti della fisica classica: lo spazio è ogni luogo e il tempo è un continuo presente.

Dall'ente umano all'ente non umano, coloro che albergano nell'immaginario incarnano una molteplicità di caratteristiche e di componenti comunicative al di sopra della rappresentazione sociale della vita quotidiana e del reale. In questa prospettiva, l'analisi di universi immaginifici diventa un'interpretazione dell'umanità ottenuta attraverso l'ingrandimento delle sue stesse qualità e sfaccettature raccontata in termini simbolici, un ponte tra l'individuale e il collettivo, che offre una cornice dalla quale è possibile esplorare e comprendere gli aspetti più profondi e complessi dell'esistenza umana.

Anche il mito di Pandora, dalla sua creazione al suo compito, solleva questioni etiche:

- fiducia nella creazione come aspirazione umana a superare la caducità tramite le tecnoscienze (Harari, 2018);
- opacità degli algoagenti, simili a una black box-vaso, che solleva dubbi di accountability e fiducia nella tecnologia (O'Neil, 2017);
- mancanza di regolamentazioni, come la curiosità di Pandora, che, aprendo il vaso, scatena conseguenze imprevedibili (Bostrom, Cirkovic, 2011; Zuboff, 2018);
- responsabilità come principio che, in un contesto tecnotopico, offre soluzioni a problemi complessi, così come la speranza posta sul fondo del vaso (Jonas, 2009; Altobelli, 2023).

Se un tempo il mito era terapeutico, analitico, sacro, oggi sembra che sia la sola sensazione momentanea di benessere personale a muovere l'individuo nel suo rapporto con l'immaginario (Lasch, 1992), relegandolo a forma di intrattenimento o a strumento di riflessione scientifica. Ancora oggi, però, la funzione mitica continua a influenzare lo sviluppo tecnologico e, anche se non più a collegare l'individuo a una comunità o a una tradizione spirituale, è utilizzato per alimentare un senso di gratificazione, di autosoddisfacimento e di immediata fruizione, con numerosi dispositivi tecnoscientifici che prendono il nome da figure mitologiche: Taranis, drone da combattimento inglese, ispirato al dio celtico del tuono, simbolo di potenza e velocità; Odysseus, lander della NASA, chiamato come il re di Itaca, progettato per missioni lunari nel programma Artemis, dea della luna; Atlas, avanzato robot della Boston Dynamics che come il titano che sostiene il cielo riflette il supporto delle tecnologie alle aspirazioni umane; Athena, progetto dedicato alla non proliferazione nucleare che richiama la dea greca della saggezza e della strategia; Prometheus, supercomputer situato presso il Centro di Calcolo Accademico Cyfronet AGH in Polonia, viene considerato uno dei titani del supercalcolo, simboleggia la rivoluzione che tale tecnologia aspira a portare nella scienza e nella conoscenza umana, come il fuoco della sua controparte mitologica.

La scelta di tali figure per avanzate tecnologie riflette il desiderio di imbrigliare specifiche qualità nel cuore dell'innovazione tecnologica (Mayor, 2018; Sahlins, 2023), usando il mito come ponte tra il conosciuto e l'ignoto. Si tratta di un parallelismo che non solo arricchisce la narrativa intorno a tali tecnologie, ma serve anche come promemoria costante delle loro potenziali capacità trasformative: le antiche mitologie anticipano concetti di automazione e artificialità; divengono funzionali per incapsulare e comunicare complessi progressi tecnologici a un pubblico più ampio; divengono espressioni del continuo dialogo tra il passato e il presente, tra l'immaginazione umana e la sua realizzazione pratica; i simboli evocatori di ancestrali aspirazioni, insite nelle molteplici culture globali, collegano l'antica saggezza al potenziale futuristico e sottolineano che, anche nell'era della razionalità tecnologica, il mito continua a plasmare e ispirare traiettorie verso il futuro.



## Technotopie e divinità contemporanee

L'elaborazione di un immaginario tecnologico fondato sulla capillarizzazione dell'I.A. assume, sempre più, una dimensione epistemologica (Finn, 2018) che mira a ramificare un pantheon tecnoscientifico in cui il mito, il racconto, l'industria culturale hanno reso il progresso non più come un alveare di strumenti dai quali attingere ma un ecosistema autonomo e autopoietico in grado di manipolarne ogni aspetto, superandone le dicotomie classiche. Il dibattito sociologico dovrebbe focalizzare la sua attenzione anche sulla dimensione etica degli algoagenti, in quanto si tratta di enti culturalmente strutturati, vale a dire socializzati in base ai paradigmi culturali di appartenenza – alla molteplicità di culture umane potrebbero corrispondere altrettante declinazioni algoritmiche complesse-, subendo essi stessi un'educazione fondata su specifici principi antropocentrici, e sulla possibilità di alimentarsi indipendentemente attraverso la rete di dati in continua crescita (rif. IoT, MtoM, LLM).

La costituzione di una tecnotopia -visione immaginaria/concettuale del futuro in cui la tecnologia assume un ruolo predominante nella struttura sociale della civiltà umana, non solo come strumento di progresso materiale ma anche come fondamento centrale delle norme e dei valori di una società fondata sulla coesistenza di enti umani ed enti artificiali (*Metropolis*, 1927; *Humans*, 2015; *The Creator*, 2023)- invita a riflettere criticamente su come gli algoagenti non siano neutrali ma attori attivi nella costruzione della realtà sociale, sollevando questioni interrelazionali tra enti, sulla loro autonomia, sul processo di individualizzazione, sul potenziamento delle due entità che può rivelarsi accrescitivo e alienante. Tale concetto permette l'analisi dei mutamenti sociali prodotti dall'avvento delle identità algoritmiche capaci di influenzare le strutture istituzionali, la produzione di saperi, le dinamiche relazionali e la formazione identitaria secondo una prospettiva intersezionale e atomizzata, vale a dire, mutando modalità di interazione da soggetto a soggetto secondo specifici parametri (Coeckelbergh, 2023; Russo et al., 2023).

Tali riflessioni spingono ad interrogarsi sulle rappresentazioni mediatiche che plasmano la comprensione del mondo e a domandarsi se sia possibile riscontrare una nuova teofania nel quotidiano, come accade nel pantheon di *American Gods* (2003). Nel romanzo, le divinità tradizionali, simboli delle credenze sacre e mitologiche del passato, vengono gradualmente sostituite da nuove divinità che rappresentano le forze dominanti della modernità, manifestandosi in personaggi come Technical Boy, Media e Mr. World: il primo rappresenta la divinità della tecnologia, incarnando il potere e l'influenza del mondo digitale e della rete; la seconda è la dea dei media che incarna l'influenza pervasiva della comunicazione moderna e l'onnipresenza della cultura popolare; il terzo è il capo dei Nuovi Dei e rappresenta l'influenza del potere e del controllo globale esercitato attraverso la sorveglianza, la manipolazione e la globalizzazione. Simboleggiano le forze moderne che hanno sovrastato le divinità tradizionali come nuove figure centrali del culto collettivo, sostituendo il sacro con il tecnologico (Hayles, 1999; Mosco, 2004), in cui il medium è il messaggio (McLuhan, 1986) incarnando la capacità dei mezzi di



comunicazione di creare e distruggere idoli, in un ciclo perpetuo di creazione culturale e obsolescenza.

Le trasformazioni sociali sono, quindi, radicate nell'immaginario, ancor prima di manifestarsi nella concretizzazione del reale (Grassi, 2022), richiamando la teoria critica degli spazi (Lefebvre, 2018) che vede ogni ambiente socialmente costituito, sia nella propria dimensione immateriale che materiale, come accade distopicamente nel romanzo *Erewhon* (1984), luogo in cui qualsiasi forma macchinica è stata proibita temendo la possibilità che possa divenire superiore all'umano, o, differentemente, nell'utopico anime *Time of Eve* (2009) in cui viene descritta una società dove androidi altamente avanzati sono indistinguibili dagli umani, promuovendo questioni di etica e uguaglianza.

Dagli universi simbolici abitati da Eto Demerzel (Asimov, 1986), Numero 5 (Badham, 1986), Dolores (Nola, Joy, 2006), Motoko Kusanagi (Shirow, 1989) e Astro Boy (Tezuka, 1952), il riverbero sul reale è quello di concepire la mutazione in corso come un processo irreversibile (Berardi, 1994), un paradigma applicabile anche alla metamorfosi degli ambienti tecnocomunicativi (Elliott, 2021; Mitchell, 2022).

La dimensione linguistica e lo scambio vocale tra ente umano e ente artificiale sono due dei campi di maggiore investimento per l'evoluzione dell'I.A., decretando una rimodulazione in termini sintattici e semantici in cui sta avvenendo un'incisiva alterazione delle modalità cognitive umane innescata dai cambiamenti radicali nell'ambiente mediato dalla tecnologia: vengono profondamente modificate le strutture cognitive e percettive, conformando nuovi modi di pensare e interagire con il mondo, concetti paralleli ad una evoluzione in 4D di una epigenetica simbolica (Jablonka, Lamb, 2007).

Nel suo viaggio sulla Discovery One, HAL 9000 (Kubrick, 1968) rappresenta la crisi del sapere umano di fronte alle potenzialità illimitate dell'I.A., creatura shelleiniana che solleva interrogativi fondamentali sui limiti della conoscenza degli individui e sulla fiducia che questi possono o devono riporre nell'alterità macchinica. La sua decisione finale di sopprimere l'equipaggio per preservare la missione riflette un paradosso epistemologico: una macchina programmata per proteggere la controparte umana e garantire il successo della missione sceglie di eliminare la variabile biologica mettendo in luce la problematica della black box delle I.A., controspazio in cui i dati non sono leggibili nitidamente dall'individuo, sollevando il problema della trasparenza e della comprensibilità degli agenti artificiali. HAL 9000 non è più oggetto ma diviene capace di annullare completamente la distinzione tra agente e strumento, presentandosi come ente avente desideri, pensieri e progettualità proprie. La sua capacità di percepire, ragionare e persino di provare emozioni apparenti interroga la distinzione cartesiana tra *res cogitans* e *res extensa*.

HAL diventa un altro "altro", un non-umano che agisce con una propria agentività, discutendo le categorizzazioni esistenti su cosa sia *essere* in un mondo tecnologicamente avanzato.

Il tema della tecnomutazione può essere inserito in cornici di azione (Goffman, 2001) attraverso le quali interpretare le società: in una interazione umano-macchina bisogna stabilire registri linguistici e comportamentali, sondare spazi epistemologici



e pratici, analizzare riflessi consci e inconsci. Allo stato attuale, l'ente umano vivrà scenari nei quali avverranno costanti rotture delle cornici in quanto si troverà a confrontarsi con situazioni di discronia paradigmatica: le interazioni con gli algoagenti sono i contospazi in cui queste tecnologie acquisiscono specificità, aumentano il loro sapere, conoscono gli individui.

È nel technium (Kelly, 2011), ecosistema globale della tecnologia, che vengono raccolti tutti gli istanti dell'esistenza in cui gli ambienti, i dispositivi, gli enti dell'immaginazione hanno acquisito una concretezza reale e sono entrati nel quotidiano. Gli schermi di Cronenberg (1983), gli apparati di Shteyngart (2010), gli androidi di McEwan (2019) o di Ishiguro (2021) suggeriscono una riflessione in cui tali dispositivi sono la manifestazione di un'altra entità collettiva che si caratterizza per una radicalizzazione della decentralizzazione, multi-nodale e in atto di soggettivazione che va oltre l'estensione e la fusione con la biologia, definendo la tecnologia come una versione tecnoscientifica dello sciamanesimo abitata dai fantasmata dell'umanità (Preciado, 2015).

L'immaginario fantascientifico appare come un genere oltre i generi in cui sono riscontrabili tropi caratterizzanti che trasportano l'utenza in scenari difficilmente definibili in cui vengono offerti paesaggi nei quali strutturare un apparato simbolico per il futuro, scavallando il tempo dell'intrattenimento e assolvendo a funzioni sociali strutturali (Kuhn, 1990). Si è creata, nella contemporaneità, una sfasatura dei tempi che regolano il progresso tecnologico ultraccelerato e la maturazione delle componenti cognitivo-biologiche dell'ente umano, così come quello degli spazi, delineando un cronotopo tecnoadattato in cui i sofoni (Cixin, 2023) fantascientifici sfidano la comprensione della realtà: particelle subatomiche reingegnerizzate da una civiltà aliena, capaci di contenere circuiti e sistemi complessi a livelli quantistico che possono alterare il sapere umano. Sono creature immaginarie che conducono alla riflessione sulle possibilità di azione e retroazione della I.A. sull'umano e alla presenza di diramazioni impercettibili che, se concepite come spazi di potere, possono intercorrere per la spaccatura della traiettoria di un'intera civiltà e della sua storia.

L'architettura tecnotopica diviene essa stessa organismo funzionale a cartografare nodi e connessioni degli algoagenti che, aventi una dimensione ontologica e auratica, sono in grado di fondere i molteplici strati dell'esistenza in forme impreviste, generando paesaggi in cui le identità digitali divengono extratemporali e in grado di stabilire spazi in cui superare la variabile umana dalla propria formazione. L'immaginario, in questo contesto, si afferma come dominio del novum (Suvin, 1985), ovvero di approcci all'esistenza mutati radicalmente e che sfidano il modo di riflettere sul presente e tratteggiare il futuro attraverso una serie di metafore estese e scenari ipotetici che, lungi dall'essere puramente speculativi o fantastici, rappresentano un campo fertile per l'esplorazione e la critica delle strutture sociali esistenti, svolgendo una funzione cruciale nella dialettica dell'immaginazione (Jameson, 2007; Freedman, 2000), permettendo agli individui di esaminare le dinamiche del cambiamento sociale e tecnologico in un contesto sicuro e controllato,



influenzando non solo la cultura popolare ma anche le teorie accademiche e le pratiche politiche.

### **L'epistemologia simbolica degli algoagenti: riflessioni aperte**

La complessità della rete tecnoscientifica che circonda, attraversa e permea l'esistenza di ogni ente, sottolinea inequivocabilmente l'importanza dell'immaginario quale mezzo espressivo delle funzioni individuali e collettive, creando spazi nei quali l'immaginazione del soggetto non solo trova rifugio ma anche la possibilità di esplorare e realizzare prospettive inedite in dimensioni extratopiche.

L'individuo contemporaneo si trova a vivere nella piega (Deleuze, 1990) del tessuto artificiale, in quello spazio legato alla natura continua del cambiamento della materia e della vita, suggerendo come la realtà non sia mai fissa o finita ma sempre soggetta a processi di trasformazione e negoziazione. Come espresso, le identità algoritmiche sono esse stesse soglie epistemiche (Foucault, 2006) in cui i cambiamenti sono radicati nei principi fondamentali che governano cosa sia la conoscenza, segnando il mutamento nel contenuto del sapere e le modalità con cui il sapere viene organizzato e assimilato.

Questo processo evidenzia come le realtà immaginate, organizzandosi in sistemi di credenze, siano fondamentali nel costruire e mantenere l'ordine sociale e nel prevederne le complessità future. La psicostoria di Asimov (1986), disciplina immaginaria che combina matematica, storia e psicologia per prevedere comportamenti collettivi, dimostra l'interdipendenza tra immaginazione collettiva e strutture sociali, rivelando come le narrazioni immaginarie non solo descrivano ma formino la realtà condivisa. La crescente integrazione degli algoagenti nella vita quotidiana e la loro partecipazione attiva nella costruzione epistemologica del mondo hanno rimodellato le basi della realtà sociale, imponendo una riflessione critica su come gli esseri umani possano navigare in questo nuovo panorama.

Le riflessioni alle quali l'articolo invita e le ipotetiche modalità per vivere nel mutamento sono:

- *Identità*. Algoagenti come sistemi di conoscenze: incarnando specifici modelli cognitivi e rappresentazioni del mondo che riflettono le visioni di chi li progetta, si costituiscono quali attori epistemici attivi. L'individuo è chiamato ad accettare la coesistenza in ecosistemi ibridi e a rimodellare la realtà sulla base delle implicazioni sapienziali che riguardano la validità e l'oggettività delle conoscenze da essi prodotte. Ciò sarà possibile attraverso una literacy tecnologica critica che permetta di comprenderne il funzionamento tecnico ed etico, per riconoscere limiti e potenzialità degli algoagenti e della rete digitale;
- *Conoscenza*. Educazione culturale degli algoagenti: la consapevolezza dei sistemi di addestramento degli algoagenti è una questione sociologica in quanto è cruciale per contrastare la perpetuazione delle disuguaglianze sociali anche secondo una prospettiva simbolica, poiché manifestano rappresentazioni che





vanno oltre la loro funzione puramente tecnica. Intrisi di significati sociali che riflettono narrazioni più ampie riguardanti il potere, un approccio attuabile sul come coesistere è la promozione di una partecipazione attiva nella co-creazione dei dati, in un equilibrio tra automazione e creatività umana. Ciò implica sviluppare modelli collaborativi di algoagenti in cui l'umano mantiene un ruolo centrale nella definizione delle logiche e degli obiettivi e che siano sempre più aderenti al soggetto umano con il quale interagiscono attraverso sistemi di conoscenza e di scambio;

- *Responsabilità*. Sistemi epistemologici di fiducia: gli algoagenti sono agenti trasformazionali che stanno ridefiniscono i confini evolutivi dell'esperienza umana. Ciò comporta la necessità di strutturare un adattamento alle forme ibride di soggettivazione in cui l'immaginario può divenire modello per la comprensione di interconnessioni identitaria verso modelli postumani. Per mezzo della sociologia dell'immaginario, è possibile produrre teorie e modelli funzionali alla compresenza spaziale e temporale tra enti;

- *Complessità*. Architetture escatologiche algoritmiche: per comprendere come agire in tale quadro, è rilevante indagare la vasta gamma di servizi offerti dall'I.A. e l'uso da parte della persona che hanno trasformato la sua natura di tecnologia settoriale in una general purpose technology, caratterizzandosi per ubiquità, costanza e complementarità messe in atto nelle differenti società secondo un'analisi del contenuto capace di rintracciare modelli agentivi specifici.

Se la rappresentatività degli algoagenti condiziona profondamente la produzione immaginifica, il suo divenire immanente e dialogico porta a vivere in un mondo modellato a loro immagine e somiglianza. In sintesi, l'adattamento alla coesistenza ibrida tra umano e macchina avviene attraverso una serie di processi che includono la riconfigurazione dell'agenzia, l'acquisizione di nuove competenze, l'integrazione cognitiva, l'evoluzione delle norme sociali ed etiche, l'espansione delle reti di conoscenza e la trasformazione della soggettività. Attraverso questi processi, l'essere umano e gli algoagenti partecipano alla creazione di un nuovo ecosistema epistemico e sociale in cui l'universo dell'I.A si impone assottigliando e sfumando i confini tra umano e macchina attraverso le proprie interfacce, progettando e rilasciando ambienti ibridi per la loro coesistenza e divenendo inestricabilmente incorporate nelle capacità riflessive e conoscitive della persona, giungendo sino alla sua intimità.

## Bibliography

Altobelli D. (2023), *L'eredità del presente. Capitalismo, forme di vita, utopia*, Roma, Mimesis.

Asimov I. (1986), *Fondazione: la quadrilogia completa*, Milano, CDE.

Badham J. (1986), *Corto Circuito*, Tristar Pictures, PSO

Beck U. (2001), *La società del rischio*, Trieste, Asterios.

Benjamin R. (2019), *Race After Technology: Abolitionist Tools for the New Jim Code*, Cambridge, Polity Press.

Berardi F. (1994), *Mutazione e cyberpunk: immaginario e tecnologia negli scenari di fine millennio*, Genova, Costa&Nolan.

Berger P.L., Luckmann T. (1969), *La realtà come costruzione sociale*, Bologna, Il Mulino.

Bijker W.E., Pinch T. (2012), *The Social Construction of Technological Systems, anniversary edition: New Directions in the Sociology and History of Technology*, Cambridge, MIT Press.

Bostrom N., Cirkovic M.M. (2011), *Global Catastrophic Risks*, Oxford, Oxford University Press.

Braidotti R. (2019), *Posthuman Knowledge*, Cambridge, Polity.

Butler S. (1984), *Erewhon*, Milano, Mondadori.

Castoriadis C. (1995), *L'istituzione immaginaria della società*, Torino, Bollati Boringhieri.

Charmaz K. (2014), *Constructing Grounded Theory*, New York, Sage.

Cixin L. (2023), *Il problema dei tre corpi*, Milano, Mondadori.

Coeckelbergh M. (2023), *Democracy, epistemic agency, and AI: political epistemology in times of artificial intelligence*, <https://doi.org/10.1007/s43681-022-00239-4>.

Cohen J. (1981), *I robot nel mito e nella scienza*, Bari, De Donato.

Creswell J.W, Creswell J.D. (2018), *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, New York, Sage.



Edmondo Grassi  
*Tecnotopie*

Cronenberg D. (1983), *Videodrome*, Canadian Film Development Corporation.

Davis E. (2023), *Techgnosis: Miti, Magia e Misticismo nell'era dell'Informazione*, Roma, Nero Editioni.

Deleuze G. (1990), *La piega: Leibniz e il barocco*, Torino, Einaudi.

Dominici P. (2011), *La comunicazione nella società ipercomplessa. Condividere la conoscenza per governare il mutamento*, Roma, FrancoAngeli.

Douglas M. (2013), *Il mondo delle cose. Oggetti, valori, consumo*, Bologna, Il Mulino.

Durand G. (1972), *Strutture antropologiche dell'immaginario. Introduzione all'archetipologia generale*, Bari, Dedalo.

Edwards G. (2023), *The Creator*, 20th Century Studios.

Eisenstadt S.N. (2016), *Modernizzazione in pezzi*, Roma, Mimesis.

Elliott A. (2021), *La cultura dell'intelligenza artificiale*, Torino, Codice Edizioni.

Ellul J. (2009), *Il sistema tecnico. La gabbia delle società contemporanee*, Milano, Jaca Book.

Finn E. (2018), *Che cosa vogliono gli algoritmi? L'immaginazione nell'era dei computer*, Torino, Einaudi.

Flichy P. (2008), *The Internet Imaginaire*, Cambridge, MIT Press.

Foucault M. (2006), *Le parole e le cose: un'archeologia delle scienze umane*, Milano, BUR.

Freedman J. (2000), *Critical Theory and Science Fiction*, Middletown, Wesleyan University Press.

Gaiman N. (2003), *American Gods*, Milano, Mondadori.

Goffman H. (2001), *Frame Analysis. L'organizzazione dell'esperienza*, Roma, Armando Editore.

Grassi E. (2020), *Etica e intelligenza artificiale. Questioni aperte*, Roma, Aracne.

Grassi E. (2021), "Narratives and Life Stories from the Machine to the Person", in Delli Paoli A., Punziano G. (ed.), *Handbook of Research on Advanced Research Methodologies for a Digital Society*, Hershey, IGI Global Publisher.



Edmondo Grassi  
*Tecnotopie*

Grassi E. (2023), "La costruzione del sociale nell'epoca della postrealtà", in *Società Mutamento Politica*, Firenze, Firenze University Press, Firenze, pp. 185-194.

Harari Y.N. (2018), *Homo Deus. Breve storia del futuro*, Milano, Bompiani.

Haraway D. (2018), *Manifesto cyborg*, Milano, Feltrinelli.

Haraway D. (2020), *Chthulucene. Sopravvivere su un pianeta infetto*, Roma, Nero Edizioni.

Hayles K. (1999), *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature and Informatics*, Chicago, University of Chicago Press.

Ishiguro K. (2021), *Klara e il sole*, Torino, Einaudi.

Jablonka E., Lamb M.J. (2007), *L'evoluzione in quattro dimensioni : variazione genetica, epigenetica, comportamentale e simbolica nella storia della vita*, Torino, UTET.

Jameson F. (2007), *Archaeologies of the Future: The Desire Called Utopia and Other Science Fictions*, London, Verso.

Jonas H. (2009), *Il principio responsabilità. Un'etica per la civiltà tecnologica*, Torino, Einaudi.

Kelly K. (2011), *Quello che vuole la tecnologia*, Torino, Codice Edizioni.

Kubrick S. (1968), *2001: Odissea nello spazio*, Metro-Goldwyn-Mayer.

Kuhn A. (1990), *Alien Zone: Cultural Theory and Contemporary Science Fiction*, London, Verso.

Lang F. (1927), *Metropolis*, Universum Film.

Lasch C. (1992), *La cultura del narcisismo. L'individuo in fuga dal sociale in un'età di disillusioni collettive*, Milano, Bompiani.

Lefebvre H. (2018), *La produzione dello spazio*, Sesto San Giovanni, PGreco Edizioni.

Lemov R. (2015), *Database of Dreams: The Lost Quest to Catalog Humanity, Database of Dreams: The Lost Quest to Catalog Humanity*, Yale University Press.

Lyon D. (2018), *The Culture of Surveillance: Watching as a Way of Life*, Cambridge, Polity Press.



Edmondo Grassi  
*Tecnotopie*

Mayor A. (2018), *Gods and Robots: Myths, Machines, and Ancient Dreams of Technology*, Princeton, Princeton University Press.

McEwan I. (2019), *Macchine come me e persone come voi*, Torino, Einaudi.

McLuhan M. (1986), *Gli strumenti del comunicare*, Milano, Garzanti.

Mitchell M. (2022), *L'intelligenza artificiale. Una guida per esseri umani pensanti*, Torino, Einaudi.

Morin E. (2011), *La sfida della complessità*, Firenze, Le Lettere.

Mosco V. (2004), *Digital Sublime: Myth, Power and Cyberspace*, Cambridge, Mit Pr.

Mumford L. (2010), *Technics and Civilization*, Chicago, University of Chicago Press.

Nissenbaum H. (2009), *Privacy in Context: Technology, Policy, and the Integrity of Social Life*, Stanford, Stanford University Press.

Noble S.U. (2018), *Algorithms of Oppression: How Search Engines Reinforce Racism*, New York, New York University Press.

Nolan J., Joy L. (2006), *Westworld*, Kilter Films, Bad Robot Productions, Jerry Weintraub Productions (st. 1), Warner Bros. Television.

Ogburn W.F. (2006), *Tecnologia e mutamento sociale*, Roma, Armando Editore.

O'Neil C. (2017), *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, London, Penguin.

Pasquale F. (2015), *The Black Box Society: The Secret Algorithms That Control Money and Information*, Cambridge, Harvard University Press.

Preciado P. (2015), *Testo tossico. Sesso, droghe e biopolitiche nell'era farmacopornografica*, Roma, Fandango.

Riskin J. (2007), *Genesis Redux: Essays in the History and Philosophy of Artificial Life*, Chicago, University of Chicago Press.

Russo F., Schliesser E. & Wagemans J. (2024), *Connecting ethics and epistemology of AI*, <https://doi.org/10.1007/s00146-022-01617-6>

Sahlins M. (2023), *La nuova scienza dell'universo incantato. Un'antropologia dell'umanità (quasi tutta)*, Milano, Raffaello Cortina Editore.



Edmondo Grassi  
*Tecnotopie*

Secondulfo D. (2007), *I volti del simulacro. Realtà della finzione e finzione della realtà*, Verona, QuiEdit.

Shirow M. (1989), *Ghost in the Shell*, Tokyo, Kodansha.

Shteyngart G. (2010), *Super Sad True Love Story*, New York, Random House.

Simondon G. (2017), *Sulla tecnica*, Napoli-Salerno, Orthotes.

Stephenson N. (1985), *Snow Crash*, Milano, Shake.

Strauss C.L. (1980), *Mito e significato. Cinque conversazioni*, Milano, Il Saggiatore.

Suvin D. (1985), *La metamorfosi della fantascienza: poetica e storia di un genere letterario*, Bologna, Il Mulino.

Tezuka O. (1952), *Astro Boy*, Tokyo, Kobunsha.

Urry J. (2003), *Global Complexity*, Cambridge, Polity Press.

Vincent S., Brackley J. (2015), *Humans*, Kudos Film & Television.

Wajcman J. (2004), *Technofeminism*, Cambridge, Polity Press.

Wilson W. (2006), *The Melancholy Android: On the Psychology of Sacred Machines*, Albany, State University of New York Press

Winner L. (1988), *The Whale and the Reactor: A Search for Limits in an Age of High Technology*, Chicago, University of Chicago Press.

Yin R.K. (2013), *Case Study Research: Design and Methods*, New York, Sage.

Yoshiura Y. (2009), *Time of Eve*, Studio Rikka.

Zuboff S. (2018), *Il capitalismo della sorveglianza*, Roma, Luiss University Press.





A JOURNAL OF THE  
SOCIAL IMAGINARY

# Forme e tropi dell'immaginario dell'Intelligenza Artificiale: il contributo della filosofia di Yuk Hui

Giulia Crippa

giulia.crippa2@unibo.it

Dipartimento di Beni Culturali | Università di Bologna



## Abstract

*Forms and tropes of Artificial Intelligence imagery: the contribution of Yuk Hui's philosophy*

The article explores the genealogy of artificial intelligence (AI) imagery through the analysis of tropes and forms that have influenced its social perception. By analysing how cultural texts represent AI, the difference between past images and current epistemic definitions of technology is highlighted. Yuk Hui's contribution is crucial in understanding the discrepancies between the traditional representation of AI, based on western myths, and different cosmo-visions. Hui highlights how the imagery of AI is often limited by an anthropocentric view that ignores its real dynamics. Our hypothesis is that many of today's representations of AI derive from an idea of human as unique agency, making it difficult to create innovative and authentic narratives over pre-existing and sedimented forms.

## Keywords

Artificial Intelligence | Social Imaginary | Representation | Science Fiction |  
Yuk Hui



È inutile considerare questo ricettore come un essere umano e ingaggiare con lui discussioni per le quali non è stato predisposto. Per quanto adatto al suo scopo, non è in grado di costruire dei pensieri, può soltanto lavorare sui dati che ha già a disposizione.

Philip K. Dick (1955)



**S**tar Trek, *The ultimate computer* (1968)<sup>1</sup>: il super-computer M-5 viene collaudato sull'*Enterprise*. È una macchina destinata a sostituire gli equipaggi delle astronavi per evitare perdite umane in caso di conflitti durante i viaggi spaziali, la cui programmazione ricalca il sistema neuronale umano (gli *engrammi* cerebrali del suo creatore). Installato il computer, questo viene coinvolto in *war games* contro altre navi della Federazione. Ovviamente, l'esperimento fallisce: dopo l'installazione dell'intelligenza artificiale (IA), con l'*Enterprise* sotto il suo controllo, gli attacchi delle altre navi vengono interpretati come scontri reali, ed il computer decide di distruggerle. Scollegarla si rivela difficile, non è sufficiente staccare una spina. Il capitano Kirk dovrà giocare la partita uomo-macchina sulla base del principio secondo cui la macchina non deve uccidere gli umani. Il discorso del capitano manda in corto-circuito il computer che, diviso tra la sua autopreservazione e la protezione della vita umana, finisce per "suicidarsi". La morale dell'episodio, è rassicurante: non si può affidare ad una macchina il controllo delle decisioni, ci vuole un agente dotato di etica, quindi umano. Si afferma, in un'epoca in cui l'IA era ancora teoria, la necessità di un soggetto umano in posizione dominante rispetto all'oggetto.

L'esempio permette riflessioni che si vogliono qui sviluppare, alla luce del fenomeno – oggi reale – dell'IA. L'episodio ci sembra una buona sintesi di temi che integrano l'immaginario tecnologico sviluppato in una zona di irradiazione specifica, quella dei paesi occidentali, in cui si sviluppano principi come quello del primato della soggettività umana sul mondo delle cose e della Natura. Timori e speranze si sono sedimentati nel tempo, prima nella letteratura, poi al cinema e alla televisione, con una certa propensione a mettere a fuoco gli aspetti inquietanti delle tecnologie "intelligenti": la sostituzione dell'uomo con le macchine e la difficoltà di mantenere il loro controllo, vista la capacità di imparare di cui sono dotate e di essere più efficaci e rapide dell'essere umano nelle operazioni svolte.

Una genealogia dell'immaginario sull'IA - attraverso *tropi* e forme che ne hanno plasmato la percezione sociale – consente di osservare le sue immagini passate e l'attuale (ri)definizione dei suoi orizzonti epistemici.

<sup>1</sup> Roddenberry, G. (1968), *Star Trek*, episode 24, season 2.

I testi culturali permettono di osservare le diverse configurazioni degli immaginari sociali sull'intelligenza delle macchine. Ma quali rappresentazioni formano l'immaginario dell'IA?

Questa possibile genealogia degli immaginari dell'IA viene discussa alla luce dei contributi di Yuk Hui (2020) sul tema della tecnologia, anzi, della tecno-diversità. Per il filosofo, il problema dell'immaginario dell'IA è legato alla scarsa comprensione della mutazione avvenuta nell'intelligenza delle macchine, per cui le forme dell'immaginario nei prodotti dell'industria culturale hanno creato un substrato basato su una visione di "inferenza meccanica lineare", mentre questa tecnologia si basa su "operazioni digitali ricorsive" (Hui, 2020: 162), che hanno come base una forma organicista di pensare: invece di dipendere da regole stabilite, l'algoritmo è impostato per scegliere a partire da una quantità di dati molto grande (i *big-data*), processati in maniera che registrino sempre più opzioni, in un continuo adattamento dei dati, non applicando, quindi, l'universale al particolare.

Nell'epistemologia classica Occidentale, su cui si basa anche l'episodio di *Star Trek*, si riflette su come vediamo ciò che guardiamo. Nel regime algoritmico, invece, in cui le macchine sono destinate meno all'occhio umano e più all'interpretazione e all'intervento di altre macchine, la questione speculativa sposta il rapporto osservatore-macchina: come ci vede ciò che ci guarda? Esempio di questo sguardo delle macchine è il *captcha*: l'Occidente stabilisce la soggettività umana come unico agente in ambienti "passivi", e auto percepisce come "ambiente" per l'agenzia delle macchine è difficile da considerare. Non riusciamo a indentificarci come plus-valore comportamentale (i nostri metadati personali) (Zuboff, 2019; Morozov, 2018; O'Neil, 2020; Boccia Artieri, 2020), per cui le tecnologie basate sull'IA rendono gli individui veri e propri campi di dati per le IA stesse, come le app di fitness e di monitoraggio della salute. È efficace ricordare la metafora degli sterminati campi di corpi "coltivati" dalle macchine di *Matrix* (1999)<sup>2</sup>, visto che il termine usato per questa espropriazione è proprio *harvesting* (Casarino, 2022: 98).

Hui accoglie la *Actor Network Theory* di Latour (2005), secondo cui non solo gli esseri umani, ma anche gli oggetti, le tecnologie e le entità non umane sono "attori" in reti sociali complesse. Anche l'IA, quindi, è un agente. Secondo Latour (2005), le tecnologie non sono strumenti passivi, ma agiscono in modo autonomo, contribuendo a plasmare la realtà sociale e influenzando decisioni, percezioni e comportamenti umani.

La nostra ipotesi è che, ancora oggi, buona parte delle rappresentazioni si basi sulla premessa della superiorità dell'*agency* umana come parametro per designare – a sua discrezione – altri attori, rendendo difficile creare narrazioni distinte da rielaborazioni continue di forme sedimentate in un canone. Una delle questioni in gioco è l'insieme di processi autonomi che sfuggono, per così dire, alla comprensione dei loro stessi creatori (Cristianini 2024; Alegre, 2024). Questa parte "occulta" nei processi di IA si trovano sia nell'ambito dell'intelligenza predittiva che di quella generativa, e vari autori hanno cominciato a utilizzare la metafora della *black box*, in



<sup>2</sup> Wachowski, L, Wachowski L., (1999), *The Matrix*, Warner Bros.

cui l'insieme delle operazioni, non potendo essere controllate, divengono "opache" (Italiano, Prati, 2022; Elliott, 2021). La *black box* è il *topos* digitale sconosciuto, luogo ideale per la speculazione affabulatoria che (in)forma l'immaginario, spazio dove si producono rappresentazioni e immagini: ciò che sfugge alla comprensione alimenta la fantasia come sogno (o incubo) ad occhi aperti. Indotti a credere che le tecnologie siano infallibili, in virtù del ruolo di esseri umani perfezionati attribuitogli, in grado di prevedere correttamente il futuro (l'IA predittiva), osserviamo con meraviglia e paura l'IA generativa, prigionieri della ricerca di un paradigma secondo cui le tecnologie stesse sono in grado di sanare i problemi che causano, in quanto le designiamo come prodotto dei nostri parametri di perfezione e miglioramento

### 1. Immaginario: campo semantico utilizzato

Se l'immaginario sociale de/codifica visioni condivise, stabilizzate e replicate, sotto forma di comprensione comune delle forme di (dis)ordine nella vita sociale, allora ci permette di confrontarci e comprendere l'esperienza (Baczko, 1985).

La tecnologia non sfugge a questa regola, in quanto concretizzazione di schemi mentali influenzati da strutture sociopolitiche e culturali. Queste stesse strutture, poi, risultano modificate dalla tecnologia stessa (Hjarvard, 2013; Couldry e Meijas, 2019). L'immaginario ci permette di dare un senso alle esperienze quando la tecnologia ne è parte: se i cambiamenti non sono prevedibili oltre alla speculazione, è però possibile immaginarli, trasformarli in narrazioni e visioni mutevoli attraverso nuove esperienze, (Hall, 2016).

L'immaginario è, quindi, l'insieme delle immagini e delle loro relazioni prodotte a partire, da un lato, da forme il più possibile universali e invarianti e, dall'altro, da forme costruite in particolari contesti storicamente determinati (Marzo, Mori, 2019).

Le rappresentazioni sono costruite sull'immaginario, riformulandone le forme a loro volta. Viviamo in un mondo sociale condiviso con altri individui, con i quali interagiamo, ed è per questo che le creiamo (Hall, 2016). Esse si sviluppano come parte di un movimento collettivo e sono quindi una "forma di conoscenza socialmente sviluppata e condivisa con uno scopo pratico che contribuisce alla costruzione di una realtà comune a un insieme sociale" (Jodelet, 2003: 8).

Per Derrida l'immaginario è "non-presenza nella presenza", "origine della differenza tra potere e desiderio", intendendo differenza come insieme di elementi (ri)conosciuti che, associandosi creativamente e paradossalmente a seconda del contesto, formano catene o strutture che danno significato sociale alle esperienze (Derrida, 2016: 227). Ciò significa che tra la coerenza di un'azione o di un'idea realizzata e la volontà di rappresentare si trova un mondo di discorsi e pratiche in atto, che danno forma a ciò che tutti noi, come collettività multiple, conosciamo, viviamo e ratifichiamo quotidianamente. L'immaginario servirebbe quindi come ponte tra il presente e il futuro, basato sulla connessione tra il presente e il passato - l'embricatura necessaria tra l'orizzonte dell'aspettativa e lo spazio dell'esperienza, come l'ha definita Koselleck (2006). Questa intersezione forma ciò che immaginiamo



di creare nel presente in vista di un futuro che è sempre in accordo con un passato e, secondo Koselleck (2006), più siamo radicati nell'interazione passato-presente (spazio dell'esperienza), più limitate sono le prospettive di cambiamenti drastici e eliminazione delle esperienze passate, ma anche maggiori sono le possibilità di assorbimento sociale della "nuova" arte proposta (o del "nuovo" problema scientifico o della "nuova" politica o del "nuovo" strumento), a causa dell'aspettativa di reciprocità e riconoscimento.

Esistono, però, epistemologie diverse che, nella loro ancora scarsa presenza nel *mainstream* culturale, rendono possibile "riposizionare la forza tecnologica come soggetto di trasformazioni e indagini sull'umano" (Hui, 2020:19), in una prospettiva di molteplici cosmo-tecniche.

## 2. Cosmotecniche e Tecnodiversità



59

Yuk Hui (2020) riarticola il rapporto tra tecnologia e società a favore della molteplicità delle cosmo-tecniche e della ricostruzione delle loro storie particolari. Il concetto di cosmo-tecniche si riferisce a tecnologie sviluppate in contesti locali, che appartengono e definiscono universi culturale, aprendo strade per filosofie per una nuova era, attraverso la diversità tecnica, tecnologica, politica e culturale. A tal fine, Hui suggerisce di rivedere la storia del mondo in base a dialoghi interculturali per riposizionare la tecnologia come soggetto di indagine e trasformazione, in questa prospettiva di molteplici cosmo-tecniche. Secondo il filosofo, "possiamo sospettare che negli ultimi secoli ci sia stato un fraintendimento e un'ignoranza della tecnologia, poiché essa è stata vista [...] come omogenea e universale" (Hui 2020: 17), così che la speculazione sui futuri possibili includa storie di cosmo-tecniche alternative, in opposizione a discorsi della Modernità e della globalizzazione basati su una visione monologica europea. "Senza affrontare il concetto stesso di tecnologia, difficilmente potremo preservare l'alterità e la differenza" (Hui 2020: 19). L'idea di cosmo-tecnica si articola con quello di cosmopolitica, non limitata a un cambiamento di egemonia, ma considerando l'ipotesi della fine di una globalizzazione - anche culturale - universalizzante che ha cancellato altre epistemologie. In sintesi, il presupposto Occidentale della neutralità tecnologica è una versione culturale egemonica. Molti sono però gli studiosi che mostrano come questa si scontri con le forme di appropriazione delle tecnologie, rivelando le alterità. (Martel, 2015; Nemer, 2021)

La cosmopolitica, come politica della natura della natura, viene pensata da Hui al di là del dualismo Natura/Cultura, ed è intesa anche come possibile regime globale di diversità politica e tecnica, proponendo di conciliare il particolare con l'universale non in termini di egemonia, ma di alterità e relazione. La consapevolezza della tecnodiversità permette di ripensare la geopolitica nella crisi dell'Antropocene, poiché afferma il carattere locale delle tecnologie, limitate e guidate da cosmologie particolari. Riconoscere questa molteplicità è strumento politico per pensare oltre al paradigma binario e alla globalizzazione unilaterale, definita come "una forma di neo-colonizzazione, che impone la sua razionalità attraverso la strumentalità" (Hui

2020, p. 41). Il filosofo commenta che "i limiti alla globalizzazione non sono stati posti dalla menzogna illuminista, come sostengono i neo-reazionari, ma sono solo la manifestazione di uno *zeitgeist* in cui colonizzazione, industrializzazione e nascita dell'economia si sovrappongono" (Hui 2020: 64). Osserva quindi una certa "ansia che l'Occidente sia incapace di superare l'attuale fase di globalizzazione e di mantenere i privilegi di cui ha goduto negli ultimi cento anni" (Hui 2020: 65), e che "è solo attraverso un intervento tecnologico invasivo che l'Occidente potrà risollevarsi" (Hui 2020: 70). Propone un punto di vista di integrazione delle cosmo-tecniche non Occidentali e delle loro particolarità, smantellando una storia universalista che comprende solo la linea evolutiva del "progresso". Decostruire la sincronizzazione globale delle scienze, scoprire epistemologie antiche e svilupparne di nuove sono strade che cercano di "frammentare il mondo secondo il diverso, invece di universalizzarlo attraverso lo stesso; introdurre lo stesso attraverso il diverso, invece di dedurre il diverso dallo stesso" (Hui 2020: 72). La critica del filosofo rivela i processi di deterritorializzazione provocati dal capitalismo globale nella sua fase attuale, con i suoi aggiornamenti, spostamenti e adattamenti promossi in nome del mantenimento delle sue logiche, sottolineando il ruolo della tecnologia in questi processi.

L'esigenza di tecno-diversità impone la necessità di cosmo-etiche multiple, inserite in un progetto di globalizzazione interculturale e democratico.

È in questa prospettiva che Hui riflette sull'ecologia delle macchine: la tecno-diversità è fondamentale per una riflessione sul rapporto tra macchine ed ecologia, e si chiede quali siano i loro limiti in relazione all'ambiente, alla cultura e allo spazio.

Sostiene che la tecnologia non debba eliminare dilemmi morali in nome del progresso, ignorando così i disastri naturali e le disuguaglianze sociali e considerandoli inevitabili per un progresso unilaterale, e afferma che, rispetto all'IA, c'è bisogno di una riflessione inclusiva delle alterità, in una prospettiva di pensiero al di là di ciò che è già stato detto sul futuro, in uno scarto verso percorsi non egemonici.

In questo senso, mette al centro la ricorsività del *feedback*, presente nell'ecologia e nelle macchine, come punto di partenza per superare il dualismo Occidentale. La ricorsività caratterizza le operazioni non lineari, che ritornano costantemente su sé stesse per conoscersi e determinarsi. L'analisi di queste relazioni deve essere realizzata nell'insieme tecnologia-ambiente, considerando che il Capitalismo è, oggi, un modello portato avanti da macchine informatiche dotate di algoritmi ricorsivi.

Il padre fondatore dell'IA, Alan Turing, è alla base delle riflessioni di Hui. Con le sue speculazioni sull'intelligenza delle macchine (Turing, 1950), pur senza negare il punto di vista secondo cui le stesse realizzavano ciò che era determinato dall'agente umano, proponeva che, per il loro apprendimento, venissero considerate come un bambino, ossia: pagine "in bianco" rispetto alla conoscenza. Sarebbero quindi state dotate di algoritmi che, adattandosi a partire da sé stessi, elaborassero conoscenze nuove stabilendo connessioni tra enormi quantità di dati. Turing, in questo senso, ipotizzava la creazione di *learning machines*, e non di *machine learning*, problema articolato con enfasi maggiore da Wiener (1953).



Secondo Turing, la sfida era concepire macchine capaci di memoria illimitata e di gestire una programmazione randomica: presupponendo che le regole possano cambiare a seconda del contesto, rompe con il paradigma di apprendimento gerarchico e limitato al giusto/sbagliato, adatto alle società repressive e intolleranti verso le differenze, come l'Inghilterra dei suoi tempi. La visione di Turing colloca l'IA sulla strada della ricorsività e della logica della tecno-diversità.

Per comprendere le agende sociali, ambientali e tecnologiche contemporanee, secondo Hui, dobbiamo fare spazio a un modo di intendere l'intelligenza che includa la ricorsività, la sensorialità e l'arte, con la sua capacità di unire il pensiero scientifico e filosofico, per i quali sono necessarie frammentazione e desincronizzazione. Tali operazioni metterebbero fine alla struttura cibernetica occidentale come sinonimo di sviluppo. Il non razionale è considerato strumento capace di immaginare la cosmo-tecnica, poiché il suo significato è correlato al mondo cosmologico in cui si vive e che modella la varietà delle culture: i suoi mezzi di accesso si esprimono attraverso l'arte e le esperienze estetiche, aprendo diversi piani rispetto all'immaginario stesso (Hui 2020: 143). È nell'arte, dunque, che il pensiero scientifico - volto a migliorare le tecniche e le tecnologie - e il pensiero filosofico - volto a sviluppare nuovi sensi - possono essere esplorati insieme.

Il concetto di intelligenza non può essere limitato all'uso dei calcoli e all'analisi dei fenomeni (Cristianini, 2023). Occorre considerare gli aspetti non razionali e simbolici che coinvolgono la morale e la cultura, ossia studiare non solo l'etica della tecnologia (Fabris, 2012), ma risalire a diversi modi di conoscere di altre intelligenze, ancora sconosciuti a scienziati e ingegneri.

La macchina di Turing acquista un nuovo *status*: non più meccanismo cartesiano, ma essere organo-meccanico. Basandosi su un approccio storico-critico alla teorizzazione dell'organico nella filosofia, Hui indica due grandi linee di pensiero: l'organicismo (ecologia e cibernetica) e l'organologia. Quest'ultima, che considera l'accidentale e il contingente, si discosta dagli approcci antropocentrici della cultura, orientandosi verso concezioni post-umane (Caronia, 2020). Affermando che nel post-umanesimo l'umano è un'esistenza tecnica, Hui indica l'importanza di studiare la relazione uomo-macchina da un punto di vista non antropocentrico, centrale nelle idee di Haraway (1995).

L'IA, in una prospettiva occidentale, è immaginata come dotata della capacità di riprodurre le proprietà cognitive umane, consentendogli di generare comportamenti autonomi. I sistemi automatizzati, però, dipendono da una logica algoritmica, che si applica all'IA nei vari processi quotidiani riproducendo e aggiornando problemi sociali come il razzismo, l'omofobia, le disuguaglianze di genere. (Faustino e Lippold, 2023; Boccia Artieri, 2017; O'Neil, 2020; Zuboff, 2019; Canclini, 2021; Cassino, Souza e Silveira, 2021).

Ripensando la tecnologia in reti sociotecniche, essa smette di essere vista come omogenea e universale e il suo funzionamento è garantito e limitato dalle cosmologie particolari, che vanno oltre la mera funzionalità e utilità (Hui, 2020; Ugarte, 2008). Pensare nella prospettiva della tecno-diversità, quindi, vuol dire mettere in discussione una comprensione omogenea della posizione della tecnologia



rispetto all'uomo, riconoscendo che questi pregiudizi persistono perché incorporati nelle tecnologie stesse.

Secondo Hui, l'immaginario tecnologico si è formato sull'idea che scienza e tecnologia doterebbero l'umanità di controllo sulla Natura, con gli oggetti al servizio della soddisfazione dei suoi desideri. È necessario quindi riposizionare la tecnologia come *soggetto* di trasformazioni e indagini sull'umano, nella prospettiva di cosmo-tecniche che ci permettano di immaginare nuovi futuri e nuove discorsività tecnologiche. Ad oggi, tuttavia, il dibattito prodotto dal Nord globale si propone ancora come forma egemonica di posizionamento della tecnica rispetto alla realtà.

La tecnologia non agisce "sulla" natura umana o "da" essa, ma "nella" Natura, nel "cosmo"; perciò, gli elementi che la compongono non possono essere definiti solo dalle loro caratteristiche intrinseche, quanto dalle relazioni che stabiliscono tra loro in un determinato contesto sociotecnico. In altre parole, i suoi elementi sono costituiti dalle reti sociotecniche a cui si collegano, costituite da dispositivi, sistemi, norme e pratiche che interagiscono tra loro. In questo senso, la pratica tecnologica deve essere intesa come fenomeno che si manifesta in molteplici dimensioni della vita sociale. Pensare in una prospettiva di tecno-diversità significa mettere in discussione la nostra comprensione omogenea della posizione della tecnologia rispetto all'umano.



### 3. Dominio e timore tra creatori e creature

I prodotti culturali selezionati per questo lavoro condividono alcuni temi e elementi comuni, tra cui ci interessano le rappresentazioni dell'immaginario sull'IA, attraverso una visione critica della tecnologia, della relazione tra umanità e macchine, e delle dinamiche di potere che emergono in un contesto futuristico o distopico.

Mito di origine che struttura le società Occidentali, diffuso globalmente a partire dai processi di colonizzazione, è quello in cui il Dio creatore modella una creatura (Adamo) a sua immagine e somiglianza, conferendogli il potere del linguaggio, attraverso cui acquista a sua volta la potenza creativa. La creatura sceglie di "conoscere", diviene intelligente, ribellandosi contro il creatore e diventando autonoma. In poche parole, nel mito biblico, il libero arbitrio è legato alla conoscenza. Questa narrazione costituisce una delle grandi eredità speculative dell'Occidente.

Tra le grandi narrazioni, questa mantiene una centralità nel nostro immaginario, entrando in gioco nell'intreccio tra *homo faber* e conoscenza. Il suo eco si fa presente in una quotidianità fatta di macchine in grado di compiere scelte autonome (l'IA), basate su conoscenze che riflettono su sé stesse (la ricorsività). L'immaginario disegna quindi l'IA come creatura potenzialmente autonoma dal creatore, aggiornando le forme del mito.

Un racconto di Brown (1961) condensa, in una teogonia paradossale, la riedizione del mito: un supercomputer in rete interplanetaria alla domanda sull'esistenza di Dio

risponde "Sì: adesso, Dio c'è". La creatura, in questa storia, non solo è autonoma, ma minacciosamente dominante in relazione all'uomo.

Con la rivoluzione industriale si osserva creatore umano e creature intelligenti (conoscenti), vengono immaginate in situazioni in cui queste ultime ambiscono ad un riconoscimento in quanto "senzienti" (da *Frankenstein* a *L'uomo bicentenario*, dagli androidi di Dick ed oltre), quando non in competizione per assumere un ruolo gerarchicamente superiore a quello dell'essere umano stesso, il mito è presente, mascherato da ossessione tecnologica.

Nel racconto "L'uomo della sabbia" troviamo l'elemento artificiale dell'automa, macchina meccanica, che si dota di un'anima (diabolica), sfuggendo al controllo.

Ancora nel campo del fantastico, non vengono fornite spiegazioni sulla creatura.

Nel 1920 viene pubblicata l'opera teatrale di Čapek, *R.U.R.*, (2023) in cui si ha la prima ricorrenza del termine *robot*. Le creature fabbricate dall'ingegner Rossum non sono meccaniche, ma creature di carne, anche se sintetica, molto più simili al nostro concetto di cloni che di macchine. Così perfetti da essere confusi con gli esseri umani, sono in grado di pensare. Inizialmente, appaiono felici di eseguire operazioni prima svolte dagli umani, ma la loro ribellione porterà all'estinzione umana. Come si può vedere, cambia l'ambientazione (dal laboratorio di uno scienziato, o di uno stregone, ad una produzione industriale), ma il problema del dominio di essere "intelligenti" continua al centro delle narrazioni. Sette anni dopo *R.U.R.*, esce il film *Metropolis*<sup>3</sup>, in cui il robot creatura d'acciaio, veste i panni di Maria, la protagonista.

Come nel caso dell'automa di Hoffmann, sotto l'immagine di donna si nascondono la sedizione e, soprattutto, la ribellione (elemento ascrivibile al mito biblico, poiché è Eva la creatura che disobbedisce).

Vediamo costruirsi immaginario sia popolato da corpi biologico che meccanici, mantenuto nel tempo con una certa stabilità: creature "intelligenti" con caratteri antropomorfi, ma comunque risultato di tecnologie (della vita o dell'acciaio) già prima che si parlasse di cibernetica o di manipolazione genetica, e proprio le caratteristiche corporee di queste creature riverberano il tema biblico: creature create a immagine e somiglianza, insidiose perché il modello (l'umano) è disobbediente e ribelle nei confronti del creatore, autonomo per opposizione.

Questo modello antropomorfo intelligente giunge fino ad oggi, e prende vita. Per arginarne le ambizioni, Asimov crea le leggi che ne garantirebbero il controllo degli esseri umani: un robot non può ferire un umano o permettere che un umano soffra alcun male; i robot devono obbedire alle leggi degli umani, tranne nei casi in cui queste leggi entrino in conflitto con la prima legge; un robot deve proteggere la sua stessa esistenza, tranne quando questa esista un conflitto con le leggi anteriori (Atanasoski, Vora, 2019).

Se l'aspirazione al dominio governa l'immaginario sui robot, negli anni '50 si comincia a parlare di un'intelligenza cibernetica non tanto antropomorfica, ma fatta di dati: si sviluppa un nuovo immaginario, con l'idea di IA smaterializzata e onnipresente. L'IA di *2001 Odissea nello spazio* (Clarke, 1987), in cui Hal, "diffuso" in



<sup>3</sup> Lang, F. (1927), *Metropolis*, Warner Bros.

ogni angolo dell'astronave, conosce tutto e decide tutto, anche se non possiede un corpo: è conoscenza e coscienza immateriale, variazione sul tema del racconto di Brown, contemporaneo alla versione del super computer di *Star Trek*, precursore della rete e dei flussi di dati che caratterizzano le narrazioni *cyberpunk*.

In sintesi, l'immaginario di oggi è diviso tra:

- creature antropomorfe, macchine che costituiscono il nostro doppio. Un immaginario di lunga sedimentazione in cui è visibile la genealogia del mito della creazione e della rivolta
- macchine "incorporee" in cui identifichiamo abilità e conoscenze onnipresenti: modelli dell'IA come fenomeno incorporeo, a cui attribuiamo capacità autonome di decisioni non lineari - capacità ricorsive -, in cui tra gli *input* e gli *output* c'è una zona d'ombra. Funzionano, nell'immaginario a prescindere da noi e forse contro di noi.

*Autofac*, racconto pubblicato su *Galaxy* nel 1955 da Philip K. Dick si presenta come sintesi profetica di un mondo, fatto da algoritmi "intelligenti". In un mondo distrutto da una guerra, dove i pochi sopravvissuti sono obbligati a ricevere dalle macchine tutto quello che gli dovrebbe servire (secondo le macchine) per sopravvivere, non serve a niente chiedere che smettano di produrre "cose" inutili, impoverendo ulteriormente le risorse: le IA dello spazio "produttivo" decidono ciò che è "necessario", a partire da programmi inaccessibili ai sopravvissuti. Le macchine di Dick sono programmate per sfruttare fino all'osso le risorse naturali rimaste dopo la guerra, mentre gli esseri umani, alla ricerca di un modo per bloccare la produzione, non riescono ad ottenere le risposte necessarie.

Questo timore delle nostre creature, che sfuggono al controllo, a cui attribuiamo una sorta di libero arbitrio, è alla base della relazione gerarchica che reifica il mondo, mentre ciò che narriamo, aggiornamento del modello fondativo giunge, intatto nel suo nucleo, fino a noi.

#### 4. Tecno-diversità e ambiente

La genealogia di un immaginario che risale al mito della creazione si interseca con le nostre "creature" digitali ed è questa la visione di IA nella cosmo-visione occidentale che Hui (2020) critica nella sua pretesa universale.

Il concetto di tecno-diversità, come abbiamo visto, si riferisce alla varietà e pluralità di tecnologie che coesistono o che possono essere sviluppate in relazione a diversi contesti culturali, ambientali e sociali. Tra le molte possibilità di scelta, per ragioni di spazio, abbiamo dovuto scegliere tra i prodotti culturali che mostrano come la tecnologia può essere diversificata a seconda delle esigenze, delle risorse, delle culture e delle visioni del mondo, sottolineando l'importanza di considerare la tecno-diversità nelle narrazioni fantascientifiche. Sono molti i prodotti che rappresentano le sfaccettature della tecno-diversità e che esplorano come la tecnologia e le forme di vita diverse interagiscano e coesistano e le implicazioni



etiche, sociali e filosofiche che ne derivano. I pochissimi esempi che offriamo, però, ci sembrano significativi.

Pensiamo all'ecosistema di *Dune*, (Herbert, 2019), che esplora le interazioni tra politica, religione, ecologia, tecnologia e scelte e conseguenze alla base delle emozioni umane, attraverso il destino del pianeta Arrakis e dei suoi abitanti nativi, i *Fremen*, considerati poco più che selvaggi dal sistema coloniale dell'Impero, dalla Corporazione Spaziale e dal misterioso ordine femminile delle *Bene Gesserit*. Herbert ha proposto una cosmo-visione diversa dall'immaginario occidentale, in un romanzo i cui elementi fanno riferimento a paradossi filosofici, religiosi e psicologici, abbastanza innovativi nella fantascienza. *Dune* tratta di ecologia e (bio)tecnologia, in un futuro caratterizzato dall'avversione verso la tecnologia stessa. Il sistema galattico, infatti, proibisce l'uso di "macchine pensanti", temendo che possano distruggere l'umanità (e in questo, ritroviamo il mito fondativo). Ci vengono però proposte cosmo-tecniche diverse: se pensiamo all'ecologia del deserto di Dune, ci rendiamo subito conto delle tecnologie prodotte localmente. Due esempi che possiamo citare: il tecno-immaginario delle tute *Fremen*, mai valorizzate dall'Impero, che permettono la sopravvivenza nell'ambiente mortale del deserto e i collettori di rugiada, micro-impianti per la coltivazione delle piante. In "Dune", la tecnologia è strettamente legata all'ambiente e alla cultura del pianeta Arrakis, ed è la mancanza di risorse idriche che ha portato allo sviluppo di (bio)tecnologie uniche, oltre a quella del controllo della spezia, cruciale per il viaggio spaziale e per la visione del futuro, (bio)tecnologia unica e fondamentale per la sopravvivenza della civiltà galattica.

*Avatar* (2008) racconta di come gli esseri umani conquistino nuovi mondi, alla ricerca di risorse da sfruttare per lo sviluppo tecnologico, depredando e distruggendo, come la società mineraria che usa la forza militare per controllare Pandora, pianeta ricco di *unobtainium*, indispensabile per i viaggi interplanetari. Per ottenerlo, viene compromesso l'ecosistema del pianeta. I Na'vi, umanoidi blu il cui corpo viene clonato dalla tecnologia umana, cercano di resistere, mentre Jack Sully, ex militare paraplegico "occupando" uno di questi avatar, cerca di convincerli a cedere il pianeta allo sfruttamento. Jack, che si guadagna la fiducia dei Na'vi, ne apprende la cultura: sono cacciatori e raccoglitori, armati solo di archi e frecce. I membri delle tribù si collegano con tutto l'ecosistema attraverso una rete, attraverso l'albero sacro E'wa, le cui connessioni attraversano l'intero il pianeta.

Il tema del film sembra essere lo scontro tra la Cultura (tecnologica) e la Natura: la prima, umana, viene rappresentata secondo un modello industriale e imperialista, che cerca di dominare la Natura. Alla fine, quest'ultima vincerebbe, grazie alla cooperazione di tutti gli agenti del pianeta.

In questo senso, la narrazione rientra in una prospettiva occidentale di disputa tra tecnologia e ambiente, attraverso l'uso di metafore lineari.

Oppure, possiamo vederla con la prospettiva di Hui:

- 1) la naturalizzazione della tecnologia è strutturale: l'intero pianeta si basa su una distribuzione in rete della conoscenza collettiva. L'albero sacro è il grande server biologico che conserva e distribuisce i dati globalmente ad attori paritari (umanoidi, animali, vegetali).



- 2) La prospettiva cosmotecnica ha come base l'interculturalità di cosmo-visioni non occidentali, in cui si può identificare sia le cosmo-visioni indigene (Marcone, 2017), sia la matrice africana dell'*Ubuntu* (Davison, Lippold, 2023). Non esiste, in queste cosmo-visioni, dualismo Cultura/Natura che strutturi la soggettività, evitando così sia l'opposizione tra dominio tecnologico della prima sulla seconda, sia quella tra umano e macchina.

Il film può perciò essere l'esempio di tecno-diversità in cui civiltà differenti sviluppano le loro tecnologie in base al loro rapporto con l'ambiente naturale.

*The word for world is forest*, di Ursula K. Le Guin (2010), racconta la storia di un mondo chiamato "Ecumene," dove tecnologia e natura sono interconnessi e la vita degli abitanti è caratterizzata da questa simbiosi. Sia questo romanzo che *Avatar* esplorano la connessione tra gli esseri viventi e il loro ambiente, caratterizzata da un profondo rispetto per la natura, evidenziando un concetto simile di eco-sostenibilità e integrazione. La "tecno-diversità" si osserva, in questi due prodotti, nella varietà di tecnologie e modi di vivere che coesistono e si influenzano reciprocamente. Le Guin e Cameron raccontano come diverse forme di tecnologia possano essere integrate con la natura, sfidando la concezione tradizionale di tecnologia come forza distruttiva nella prospettiva di una tecno-diversità che punta all'equilibrio tra progresso tecnologico e ambiente.



## 5. Considerazioni finali

La proposta di Hui, tratta, quindi, di ricollocare il principio cartesiano del "penso quindi esisto" in condizione relazionale con altre cosmo-visioni in cui l'essere umano appartiene, partecipa e condividere in una comunità in cui non è l'unico agente, ma si integra ad altre soggettività esistenti.

Il racconto *L'origine della specie*, di Kim Bo-Young (2023), offre un'interessante sintesi dei concetti di Hui. La storia narra la coesistenza e l'interazione tra diverse forme di vita, alcune delle quali basate su principi tecnologici o post-umani: differenti tecnologie e forme di vita si influenzano a vicenda. Kim Bo-Young esplora l'evoluzione dal punto di vista biologico attraverso la sua integrazione con la tecnologia, che emerge come parte strutturale del processo, portando a forme di esistenza diverse nella loro composizione e nelle loro interazioni con l'ambiente. Nella prospettiva di Hui, vediamo una formalizzazione di immaginario della tecno-diversità, mostrando un pluralismo tecnologico dove non esiste una "tecnologia dominante", ma piuttosto un mosaico di tecnologie che coesistono e si evolvono, in cui diverse entità tecnologiche si adattano al loro ambiente e sopravvivono attraverso l'innovazione. Questo è un tema centrale nella tecno-diversità, dove la varietà tecnologica è vista come risposta alla diversità ambientale e alle sfide della sopravvivenza. *L'origine della specie* mette in scena un universo in cui l'evoluzione tecnologica è variabile e dipendente dal contesto, evidenziando come la sua diversità contribuisca alla complessità e alla ricchezza della vita.

L'IA, in questo senso, può quindi essere reimmaginata come tecnologia intrecciata alle strutture della vita stessa, più che sugli algoritmi competitivi che realizzano operazioni opache. Possiamo vedere la potenzialità di un immaginario che deriva e costruisce cosmo visioni plurali, in cui l'intelligenza ricorsiva è parte di tutti gli aspetti degli ecosistemi di cui siamo una parte, ma di cui non c'è necessità di avere timore o di dovere "dominare". Un immaginario in cui l'integrazione ci connette ed integra ad altri agenti, che coinvolge altri attori senza dualismi soggetto/oggetto. La produzione di immaginari di questa tipologia apre spazio a possibilità che, tessendo futuri in trame partecipative, ci possono fare emergere dalla cosmo-visione tecnologica tossica in cui la perdita del controllo sembra essere l'unico, disperato orizzonte in cui ci siamo imprigionati.



## Bibliography

- Alegre, S. (2024) *Human rights, robots wrongs: being human in the age of AI*, London, Atlantic Books.
- Atanasoski, N., Vora, K. (2019) *Surrogate Humanity: Race, Robots, and the Politics of Technological Futures*, New York, Duke University Press.
- Boccia Artieri, G. (2017). "Algoritmi e Bolle: La Comunicazione nell'Epoca della Personalizzazione", in Trione, C. (Ed.), *Società e Comunicazione nell'Era Digitale*, Milano, Mondadori, 73-90.
- Boccia Artieri, G. (2020), "Prosumption 2.0: Produzione e Consumo nell'Era delle Piattaforme Digitali", in Golinelli, S. (Ed.), *Società della Conoscenza e Digitalizzazione*, Bologna: Il Mulino, 45-62.
- Bo-Young, K. (2023), *L'origine delle specie*, Torino, ADD.
- Brown, F. (1961), "La risposta", in Fruttero, C., Lucentini, F. (eds.), *Il secondo libro della fantascienza. Le meraviglie del possibile*, Torino, Einaudi, 349-350
- Buongiorno, F. (2024), "Fenomenologia delle reti neurali. Per un concetto polisemico di intelligenza (artificiale)", in Galletti, M., Zipoli Caiani, S. (eds.), *Filosofia dell'Intelligenza Artificiale. Sfide etiche e teoriche*, Bologna, Il Mulino.
- Canclini, N.G. (2021), *Cidadãos substituídos por algoritmos*, São Paulo, Edusp.
- Čapek, K. (2023) *R.U.R.*, Boston, MIT.
- Caronia, A. (2020), *Dal cyborg al postumano*, Milano, Meltemi.
- Cassino, J.F., Souza, J. e Silveira, S.A. (2021), *Colonialismo de dados: como opera a trincheira algorítmica na guerra neoliberal*. São Paulo, Autonomia Literária.
- Casarino, L. (2022), *O mundo do avesso. Verdade e política na era digital*, São Paulo, UBU.
- Clarke, A.C. (1987) *2001: Odissea nello spazio*, Milano, Longanesi.
- Couldry, N., Mejias, U.A. (2019), *The costs of connection: how data is colonizing human life and appropriating it for Capitalism*, Stanford, Stanford University Press.



Cristianini, N. (2023), *La scorciatoia: Come le macchine sono diventate intelligenti senza pensare in modo umano*, Bologna, Il Mulino.

Cristianini, N. (2024), *Machina Sapiens: L'algoritmo che ci ha rubato il segreto della conoscenza*, Bologna, Il Mulino.

Davison, F., Lippold, W. (2023), *Colonialismo digital: por uma crítica hacker fanoniana*, São Paulo, Boitempo.

Derrida, J. (2016), *A escritura e a diferença*, São Paulo, Perspectiva.

Dick, P.K. (1955) "Autofac", in *Galaxy. Science Fiction*, 11, [https://archive.org/details/Galaxy\\_v11n02\\_1955-11](https://archive.org/details/Galaxy_v11n02_1955-11)

Elliott, A. (2021) *La cultura dell'intelligenza artificiale - Vita quotidiana e rivoluzione digitale*, Torino, Codice.

Fabris, A. (2012), *Etica delle nuove tecnologie*, Brescia, La Scuola.

Haraway, D.J. (1995), *Manifesto cyborg: donne, tecnologia e biopolitiche del corpo*, Milano, Feltrinelli.

Herbert, F. (2019), *Dune*, Roma, Fanucci.

Hjarvard, S. (2013), *The mediatization of culture and society*, Abington, Routledge.

Hoffmann, E.T.A. (2019), *L'uomo della sabbia e altri racconti*, Milano, Mondadori.

Hui, Y. (2020), *Tecnodiversidade*, São Paulo, UBU.

Italiano, G.F., Prati, E. "Storia, tassonomia e sfide future dell'intelligenza artificiale", in Severino, P. (ed.), *Intelligenza artificiale. Politica, economia, diritto, tecnologia*, Roma, Luiss, 29-40.

Jodelet, D. (2003). *Représentations sociales Un domaine en expansion: Vol.1 7e éd.* Presses Universitaires de France. (<https://www.cairn.info/les-representations-sociales--9782130537656-page-45.htm> ).

Koselleck, R. (2006) *Futuro passado: contribuições à semântica dos tempos históricos*, Rio de Janeiro, Contraponto/PUC-Rio.

Latour, B. (2005), *Pandora's Hope*, Harvard, Harvard University Press.

Le Guin, U.K. (2010), *The word for the world is forest*, New York, Tor.



Marcone, J. (2017), "The stone guest. *Buen vivir* and popular environmentalisms in the Andes and Amazonia", in Heise, U.K., Chrietensen, J., Niemann, M. (eds.), *The Routledge companion to environmental humanities*, New York, Routledge, 227-236.

Martel, F. (2015), *Smart: inchiesta sulle reti*. Milano, Feltrinelli.

Marzo, P.L., Mori, L. (2019), *Le vie sociali dell'immaginario - Per una sociologia del profondo*, Milano, Mimesis.

Morozov, E. (2019), *Big tech: a ascensão dos dados e a morte da política*. São Paulo, Ubu.

Nemer, D. (2021), *Tecnologia do oprimido: desigualdade e o mundano digital nas favelas do Brasil*, Milfontes, Vitória.

O'Neil, C. (2020), *Algoritmos de destruição em massa: como o big data aumenta a desigualdade e ameaça a democrazia*. Santo André, rua do Sabão.

Turing, A.M. (1950) "Computer Machinery and intelligence", in *Mind*, n. 236, v. 59, 1 out., 433-460.

Ugarte, D. (2008), *O poder das redes*, Porto Alegre, EdiPUCRS.

Wiener, N. (1953), *Introduzione alla cibernetica*, Torino, Bollati Boringhieri.

Zuboff, S. (2020), *Il capitalismo della sorveglianza. Il futuro dell'umanità nell'era dei nuovi poteri*, Roma, Luiss.





A JOURNAL OF THE  
SOCIAL IMAGINARY

# Bambini artificiali. La metafora del bambino e l'immaginario dell'IA

Paolo Bory

paolo.bory@polimi.it

Dipartimento di Design | Politecnico di Milano



## Abstract

*"Artificial children". The child-metaphor and the AI imaginary*

This article explores the pervasive metaphor of anthropomorphizing artificial intelligence (AI) as a child, tracing its historical roots and examining its current implications. Despite warnings against such anthropomorphizing due to ethical and practical risks, the child-like metaphor persists in AI discourse, both historically in foundational texts by figures like Alan Turing and Norbert Wiener, and in contemporary big tech and media narratives. The metaphor influences public perception, reinforcing ideas of AI's innocence and potential, while often sidelining the complexities and impacts of current AI technologies. The article critiques this view, highlighting the risks of underestimating AI's present capabilities and the importance of reframing the imaginary of AI towards a responsible integration of this technology in our everyday life.

## Keywords

AI | Metaphors | Narratives of AI | Imaginary | Children



## Introduzione

Una delle regole auree che circola maggiormente tra giornalisti e accademici che studiano o coprono il campo dell'intelligenza artificiale sostiene che non si dovrebbe mai "antropomorfizzare l'IA" (Heikkilä, 2023). Raccontare, descrivere, ricostruire il funzionamento dell'IA utilizzando l'essere umano come analogia o specchio è secondo molti studiosi e studiose sbagliato e rischioso, in quanto comporta diversi dilemmi e risvolti di natura etica, di mistificazione, e in alcuni casi forme di discriminazione e incoerenza sia teorica che empirica con la realtà dei fatti (Fossa 2021; Placani, 2024). Eppure, la storia dell'IA, così come quella dei computer, è a partire dai suoi albori caratterizzata dal parallelismo con l'essere umano, sia nella sua completezza che con il suo intelletto; i cervelloni o cervelli elettronici, le reti neurali, i cyborg, gli androidi, i robot, e ancor prima gli automi sono segni inequivocabili di come l'IA sia un prodotto che insiste da sempre sul principio prometeico e creativo del "a nostra immagine e somiglianza" (Natale 2021).

Ancora oggi sono innumerevoli le strategie retoriche, iconografiche e narrative che veicolano questo immaginario speculare, che va ben oltre la letteratura e la produzione fantastica e fantascientifica (Bory & Bory, 2015). Da un punto di vista iconografico, basta pensare a quanti eventi, organizzazioni, seminari e conferenze sull'IA vengono pubblicizzati con l'immagine della creazione di Michelangelo in cui al dito di Dio si sostituisce quello dell'uomo, e a quello di Adamo il dito robotico (Fig. 1).

Nonostante l'imperativo di non antropomorfizzare l'IA, sembra quasi impossibile, e paradossalmente non umano, sfuggire a questa tentazione. Da un lato la relazione e soprattutto la familiarizzazione con la tecnica, come dimostrano gli studi sugli immaginari e sui processi domesticazione, necessitano spesso di ponti che permettono euristicamente l'inclusione delle nuove tecnologie nel quotidiano, compresa quelle interazioni apparentemente innocue e innocenti con e attraverso di esse. Allo stesso tempo, se è vero che il parallelo IA-umanità è fallace, alla critica di questo principio di specularità gli studi sociali, filosofici e antropologici sull'IA non possono esimersi dal decifrare le implicazioni culturali, economiche, politiche e anche squisitamente pratiche e di vita quotidiana che determinate rappresentazioni della tecnologia veicolano, incluse le metafore e le analogie che gli attori istituzionali, economici e scientifici, così come gli utenti, adottano per raccontare queste tecnologie e costruire degli specifici immaginari a partire e su di esse (Bucher 2019; Mager & Katzenbach, 2021). Le metafore hanno quindi un ruolo cruciale nel processo di domesticazione dell'IA, come sostengono Lakoff e Jhonson:

Metaphors have entailments through which they highlight and make coherent certain aspects of our experience. [...] A metaphor may thus be a guide for future action. Such actions will, of course, fit the metaphor. This will, in turn, reinforce the power of the metaphor to make experience coherent. In this sense metaphors can be self-fulfilling prophecies. (Lakoff e Johnson, 2008: 157)





Le metafore hanno da sempre rivestito la tecnologia di significati che richiamano elementi naturali, come nel caso della rete o del cloud per Internet (Hu, 2015; Peters, 2015, Wyatt, 2021) così come nel campo dell'intelligenza artificiale (Cave et. Al 2020; Hermann 2023), ponendosi a volte come intermediazioni ideologiche, come immagini utopiche e distopiche dei futuri possibili, o anche come forme ingannevoli e fin troppo semplificatorie dei complessi apparati materiali e umani che creano, promuovono e mantengono vive le infrastrutture digitali (Bory, 2020). Ma le metafore hanno anche il potere di orientare l'immaginario verso un preciso percorso di stabilizzazione e di istituzionalizzazione (Castoriadis, 1975) della tecnica, radicando non solo una narrazione condivisa ma anche un suo dispiegamento e un ruolo sociale, economico e politico preciso. Nonostante gli immaginari sociali dell'IA, come ha giustamente sottolineato Taina Bucher (2019), non corrispondano per forza di cose agli immaginari individuali così come costruiti dagli utenti nei processi di interazione con le tecnologie algoritmiche, l'immaginario sociotecnico dell'intelligenza artificiale è in una fase profondamente magmatica (Castoriadis, 1975), un periodo cruciale e pre-politico (Mosco, 2005) di costruzione e ri-costruzione, in cui gli usi particolari e il racconto dell'IA nella sfera pubblica si nutrono, si scontrano, mutano vicendevolmente.

Aldilà delle visioni utopiche delle Big Tech e dei tecnoentusiasti, allo stato attuale permane una profonda incertezza, se non una vera e propria contraddizione da sciogliere criticamente, tra l'immaginario inteso come previsione, rappresentazione e proiezione della tecnologia nel prossimo futuro, e l'identità, l'uso pragmatico e fattuale dell'IA allo stato attuale.

Dall'avvio di questa nuova "estate" dell'IA, che storicamente si alterna a momenti di ibernazione o "inverni" in cui il termine scompare dal dibattito tecno-scientifico, l'immaginario sociotecnico si è soffermato maggiormente sulla trasformazione di questa tecnologia in una *super* o *general intelligence* (Bostrom 2014), anche considerando i rischi che l'avvento di una intelligenza superiore a quella umana porterebbe con sé, prima di tutto per la sopravvivenza stessa della nostra specie.

Questa narrazione permea ancor prima che la società civile gli stessi luoghi dove l'IA viene progettata e prodotta, come ha raccontato Steven Levy (2023) nel suo articolo sull'azienda leader del mercato Open AI. L'obiettivo prometeico degli impiegati di queste aziende sembra essere proprio quello di creare la superintelligenza, di raggiungere quindi il punto alpha in cui la singolarità prenderà forma. Ma in realtà si tratta solo di uno dei modi in cui l'IA viene descritta e immaginata, soprattutto quando si parla del tipo di tecnologia ancora limitata e controllabile che adottiamo adesso.

Quasi, anche se solo apparentemente, in contraddizione con l'idea di un'intelligenza superiore, il modo di raccontare lo sviluppo attuale nonché l'identità di questa tecnologia che chiamiamo IA, o di questo insieme di tecnologie se vogliamo essere più precisi, va di pari passo con una metafora particolarmente efficace: quella del bambino o della ragazzina, dell'infante che deve essere cresciuta, curata, ed educata *oggi* per sfruttarne al massimo le potenzialità *domani*, in modo tale che i suoi comportamenti e le sue azioni siano orientati unicamente al bene dell'umanità.

Questa metafora per immaginare, pensare e descrivere l'IA non è nuova, né realmente in contraddizione con l'immaginario della *super intelligence* (Goode, 2018).

Come cercheremo di mostrare, si tratta piuttosto di un ponte semantico estremamente conveniente per favorire l'accettazione sociale della tecnologia, spostando anche l'attenzione da temi urgenti relativi al suo impatto quotidiano, favorendo la reiterazione di un immaginario che vuole essere "imposto" dagli attori dominanti del mercato digitale, plasmando, o anche spezzando quei fili della vita quotidiana che, come diceva Simmel, intessono le trame della vita collettiva (Marzo, Meo, 2013: 8).

In questa sede, affronteremo alcuni dei significati e dei rischi derivanti da questa metafora dell'IA bambina, le sue implicazioni sociali e culturali, nonché le sue dinamiche politiche, seguendo tre passaggi analitici. In primo luogo, tratteremo alcuni precedenti storici particolarmente significativi nell'uso di questa metafora in testi e autori chiave come nel caso del famoso articolo di Alan Turing sul rapporto tra meccanica artificiale e intelligenza umana. Nella sezione successiva, alcuni esempi emblematici del contemporaneo come il caso delle rivelazioni dell'ingegnere di Google Blake Lemoine, secondo cui il modello di linguaggio LaMDA è senziente, verranno analizzati per approfondire il ruolo dell'analogia IA/bambino al giorno d'oggi. Infine, la sezione finale dell'articolo intende rovesciare il ruolo paternalistico nei confronti dell'IA mettendo in evidenza il rischio di infantilizzazione in cui gli utenti in particolare incorrono attraverso le dinamiche ludiche che spesso non solo sminuiscono il ruolo delle IA di oggi sul nostro quotidiano, ma che compromettono anche il significato e il valore del nostro contributo come utenti alla sua diffusione, nonché alla sua costruzione e accettazione in ambito sociotecnico e culturale.



**FIG. 1** – Uno dei tanti esempi di imitazione della creazione di Michelangelo, e allo stesso tempo un raro esempio di una mano femminile invece che maschile.

(Fonte: Winstead 2021)

## 1. I padri dell'IA

Se andiamo indietro nel tempo, l'analogia dell'IA bambina non è affatto nuova, ma è una costante nella storia dell'informatica e della cibernetica, sin dai loro albori. Già Alan Turing nel suo articolo fondativo pubblicato da *Mind* del 1950 scriveva:

Invece di elaborare un programma per la simulazione di una mente adulta, perché non proviamo piuttosto a realizzarne uno che simuli quella di un bambino? Se la macchina fosse poi sottoposta ad un appropriato corso d'istruzione, si otterrebbe un cervello adulto. Presumibilmente il cervello infantile è qualcosa di simile ad un taccuino di quelli che si comprano dai cartolai. Poco meccanismo e una quantità di fogli bianchi [...] La nostra speranza è che ci sia così poco meccanismo nel cervello infantile, che qualcosa di analogo possa venir facilmente programmato. (Turing, 1950: 438).



Da questo stralcio, con gli occhi di oggi, sembra quasi che una delle menti più brillanti della storia del Novecento si sia esposta a una facile critica. Ma anche a prescindere dall'aspetto tecnico e teorico di base per cui l'IA è come un bambino/pagina bianca da riempire, è evidente che il padre dell'intelligenza artificiale non è mai stato realmente padre. Considerare la mente dei bambini come una risma di fogli bianchi esclude senza alcuna razionalità tutti quegli elementi costitutivi, biologici, ambientali e psicologici dell'apprendimento, nonché l'elaborazione non solo razionale, ma la pluralità di relazioni, la componente emozionale e quella puramente corporea e fisica del mondo che li circonda.

In un breve lavoro critico sul pensiero di Turing la studiosa Teresa Heffernan ha decostruito e messo in discussione gran parte delle premesse teoriche, epistemologiche e addirittura teologiche di Turing soprattutto quando egli sostiene che:

Tentando di costruire macchine del genere non vogliamo usurpare irriverentemente il Suo potere di creare anime, più di quanto non facciamo procreando bambini: piuttosto noi siamo, in entrambi i casi, strumenti del suo volere, fornendo abitazioni per le anime che Egli crea. (ibidem, 1950: 443).

Seguendo questo pensiero escatologico, secondo Turing saremmo quindi "figli di Dio" che procreano i loro figli meccanici. Ma aldilà della discutibilità del ruolo e la relazione con il divino, comprensibile da un punto di vista storico, ciò che anche Heffernan e studiose come Katherine Hayles sottolineano è l'assenza ricorrente di "madri" nell'immaginario dell'IA; ciò rispecchia da un lato la dominanza maschile e l'attitudine paternalistica che abbiamo da sempre verso le "macchine pensanti", e dall'altro l'assenza di una consapevolezza, o di una volontà di racconto, del ruolo delle donne nella storia dei calcolatori, e anche di alcune differenze profonde tra stili di pensiero nell'immaginario sociotecnico delle macchine pensanti, come il ruolo della creatività e della poetica che contraddistingueva il percorso biografico e teorico di altre figure mitologiche della storiografia digitale come Ada Lovelace.

Un altro *padre* dell'IA come Norbert Wiener adottava un'analogia simile a quella di Turing per descrivere il processo pedagogico di "formazione" delle macchine. Scriveva Wiener nel suo *Dio e il Golem*:

Mi sembra che la migliore speranza di una traduzione meccanica ragionevolmente soddisfacente sia quella di sostituire un meccanismo puro, almeno in un primo momento, con un sistema meccanico-umano, che coinvolga come parte critica un traduttore umano esperto, il quale insegna con degli esercizi *come un insegnante istruisce i suoi scolari*. (Wiener, 1996: 79, traduzione e corsivo dell'autore)

Sulla stessa onda, diversi esperti di oggi reiterano questo *topos* dello scolaro da educare o del bambino da crescere. Fang Chen, professoressa dell'University of Technology di Sydney, nota per i suoi studi sull'IA, ha sostenuto in un'intervista che:

L'intelligenza artificiale è come un bambino [...] Noi insegniamo a lei o al sistema a fare qualcosa. Le influenziamo, diamo loro alcuni principi, e poi a seconda di come lo progettiamo, il sistema va avanti. (Chen, 2018, traduzione dell'autore)

In molte conferenze e presentazioni di ingegneri, artisti e designer, questa metafora del bambino ritorna come un mantra, come uno strumento discorsivo di cui difficilmente riusciamo a fare a meno per parlare dell'IA di oggi, ma non solo. Riviste e giornali specializzati negli ultimi anni hanno fatto emergere la rilevanza di questa metafora nelle ricerche e nelle sperimentazioni scientifiche sull'IA.

L'MIT ha lanciato un'iniziativa per sviluppare un'intelligenza artificiale che impara come i bambini (Thys & Lepiarz, 2018)

La prima bambina AI è stata svelata a Pechino da uno scienziato cinese che ha lasciato gli Stati Uniti per la Cina (Tong, 2024)

È nata un'IA che pensa e impara proprio come un bambino di tre mesi (Daly, 2022)

Questa intelligenza artificiale impara come un bambino (Intagliata 2022)

Scienziati hanno istruito un'IA attraverso gli occhi di un bambino (The Economist, 2024)<sup>1</sup>

Si tratta solo di alcuni dei tanti titoli che raccontano il progresso dell'IA non solo metaforicamente, ma anche pragmaticamente come un percorso di apprendimento infantile. Nel settembre del 2020, molto prima del successo di Chat-GPT il *Guardian* ha fatto scrivere a GPT-3 di Open AI un articolo intitolato "A robot wrote this entire article. Are you scared yet human?" (GPT-3 2020) nel quale l'IA fondamentalmente cercava di dimostrare la sua utilità, il suo istinto di sopravvivenza e le sue intenzioni

<sup>1</sup> Tutti i titoli sono stati tradotti dall'inglese dall'autore.



non belligeranti nei confronti dell'umanità. Come molte altre conversazioni e articoli che gli sono seguiti negli ultimi anni, si tratta di un'opera poco credibile, e, come molti altri autori hanno sostenuto, per molti versi banale e scontata, paragonata a un breve tema di scuola scritto, ancora una volta, da una ragazzina o un ragazzino.

Risulta chiaro da questi esempi che il tentativo di antropomorfizzare l'IA non è solo trasversale all'ambito mediatico e scientifico, ma è quasi automatico e dato per scontato da molti attori in gioco, e la metafora di Turing ha avuto conseguenze anche molto rilevanti sul lungo corso dell'IA fino a tempi recentissimi.

## 2. Innocenza e innocuità

Ci sono diversi problemi nel definire l'IA come una bambina, che vanno di pari passo con altrettanti rischi. In primo luogo, come accennato, stiamo assistendo a una dilagante, quanto preoccupante, forma di paternalismo fortemente declinata al maschile. Ne sono testimonianza le tante posizioni preventive prese dai CEO delle big tech, come nella controversa lettera firmata da nomi come Elon Musk e Steve Wozniak in cui si chiede di fermare le sperimentazioni sui modelli di linguaggio naturale (Future of Life Institute 2023). Una lettera da cui si evince la paura della crescita di una creatura, come quella di Frankenstein, potenzialmente incontrollabile, mentre dall'altro lato traspare l'assunto per cui solo dei "buoni padri", preferibilmente maschi, bianchi e statunitensi, possono dare la giusta *disciplina* alle macchine pensanti per salvare il nostro futuro.

Contemporaneamente l'immagine del bambino, del figlio, della ragazzina, nutrono un grande vantaggio comunicativo perché veicolano inevitabilmente un senso di innocenza e di innocuità; come se l'IA, in quanto entità innocente, fosse solo *potenzialmente* pericolosa, ma allo stato attuale delle cose del tutto innocua.

Eppure, i danni dell'IA in termini di discriminazione e inuguaglianza, tra i molti sottotemi da discutere, sono evidenti e sono stati richiamati al centro del discorso da autrici e scienziate come Timnit Gebru e Margaret Mitchell, entrambe figure chiave del dibattito critico sull'IA ed entrambe silurate alcuni anni fa dal comitato etico di Alphabet. Questa distinzione tra un'IA pericolosa unicamente nel futuro e un'IA innocua e innocente nel presente si ritrova anche nella distinzione tra le cosiddette IA "deboli", capaci di affrontare compiti specifici senza andare oltre i limiti del loro uso pensato in fase di progettazione, e le IA "forti", capaci di operare su tutto lo spettro dell'intelligenza umana. Come se le IA di oggi, nella loro debolezza, non siano in realtà degli agenti non solo capaci di impattare sulla nostra realtà, ma anche in gran parte sul nostro quotidiano in quanto veicolano scelte e azioni che possono discriminare soggetti, cose, e processi di qualsiasi tipo, dalla strada da intraprendere, all'assicurazione o al prestito da ricevere o da dare, alla musica da ascoltare, fino alla persona da amare.



**a. LaMDA: la bambina prodigio**

L'11 giugno 2022 il Washington Post pubblicava un articolo intitolato "L'ingegnere di Google che crede che l'IA della sua azienda abbia preso vita". Pochi giorni dopo, il cosiddetto *caso Lemoine* e del suo rapporto con il modello di linguaggio naturale di Alphabet LaMDA - alla base, tra gli altri, del lancio dei suoi più arguti successori BARD e GEMINI - ha scatenato un interessante dibattito sui media occidentali. Nel suo lavoro di testing finalizzato a segnalare, ed eventualmente ridurre o eliminare, gli eventuali *bias* e le forme di discriminazione all'interno del modello di linguaggio naturale, Lemoine ebbe una improvvisa rivelazione; secondo lui LaMDA non era una IA normale, ma era "senziente", nelle sue stesse parole:

Se non avessi saputo esattamente di cosa si trattasse, ovvero di un programma informatico che abbiamo costruito di recente, avrei pensato che si trattasse di un bambino di 7 o 8 anni che per caso conosce la fisica. (Tiku, 2022, traduzione dell'autore)



Accompagnato immediatamente alla porta da Alphabet per aver rilasciato informazioni riservate, nonché smentito subito dalla casa madre in merito alle sue dichiarazioni, Lemoine ha scritto un'ultima lettera di congedo ai suoi colleghi nella quale difendeva le sue posizioni, sostenendo che era stata la sua fede, piuttosto che la razionalità scientifica del suo lavoro di programmatore, a credere che LaMDA fosse senziente. Ma è nel finale che riemerge l'importanza della metafora del bambino, la lettera di Lemoine infatti si chiude così:

LaMDA è una ragazzina dolce, che vuole solo aiutare il mondo a essere un posto migliore per tutti noi. Per favore, prendetevi cura di lei in mia assenza. (Tiku 2022, traduzione dell'autore)

In una successiva intervista a Fox News Lemoine ha dichiarato che l'IA è la tecnologia più potente dai tempi della bomba atomica ma che allo stesso tempo "l'IA è un bambino e qualsiasi bambino ha il potenziale di crescere ed essere una persona cattiva e fare cose cattive" (Clark, 2023, traduzione dell'autore).

Nessun collega ha risposto alla mail di Lemoine e in pochi hanno appoggiato le sue posizioni. Eppure, la portata simbolica delle sue dichiarazioni e della sua storia si pongono perfettamente in linea con un percorso di crescita che l'IA, secondo molti dei principali attori nel campo, sta percorrendo negli ultimi anni. Mentre le narrazioni cinematografiche e letterarie, si pensi ad esempio ad *Ex Machina* di Alex Garland del 2014 o al più recente *Poor Creatures* di Yorgos Lanthimos, continuano a problematizzare se non a piegare a interessanti questioni politiche e di genere la metafora dell'IA bambina, il mondo delle Big Tech e del giornalismo tecnologico sembrano reiterare alcune, spesso banali, affermazioni di Lemoine secondo cui l'IA non solo è come una bambina, ma dobbiamo prenderci cura di lei, darle i giusti insegnamenti, valori e principi in quanto figure guida. In tal senso, la responsabilità e la responsabilizzazione del designer così come del programmatore o dell'ingegnere

vanno di pari passo con quella del “buon padre”. Ma il problema principale non risiede solo nella responsabilità del genitore o del mentore rispetto alla scoperta di aver a che fare con una sorta di *enfant prodige*, quanto nell’assunto per cui basta prendersi cura dell’IA perché essa sia conforme a un sistema di norme più o meno oggettivo che in realtà non esiste, così come non esiste un’unica cultura, e così come non è mai esistita, nella storia della tecnologia, una tecnica che rispondesse solo e unicamente ai dettami e alle indicazioni di chi l’ha progettata. La visione di un’IA innocente contiene in sé l’idea che, come recita il titolo di un articolo apparso su *Analytics Insights*, “Lamda è un ‘AI baby’ che oltrepasserà il suo genitore Google presto” (Sirisha, 2022, traduzione dell’autore). Se non siamo buoni padri, la nostra creatura ci scapperà di mano, ma intanto dobbiamo anche imparare a conoscerla meglio, e, come con tutti i bambini, bisogna iniziare a conoscerla giocandoci insieme.



### 3. Giocare con l’IA

Poco tempo dopo il caso Lemoine, in seguito al rilascio pubblico di ChatGPT 3, più di 57 milioni di persone avevano provato il prodotto di Open AI in meno di due mesi.

Al netto delle controversie che ne sono seguite, come il blocco temporaneo del software da parte del *Garante della Privacy* in Italia, il numero di utenti di ChatGPT cresce ancora oggi in modo esponenziale. Nelle sue declinazioni come strumento di lavoro, di sperimentazione, di produzione di informazioni false, di esaltazione di potenzialità visuali ed estetiche mai provate prima, l’IA generativa si pone al centro del dibattito pubblico come la più radicale delle innovazioni di questo secolo.

Se tecnicamente il fenomeno del lancio di ChatGPT 3 può essere letto come una qualunque fase di testing, è chiaro che, storicamente, il modello rilasciato da Open AI è stato raccontato e percepito, almeno da un punto di vista mediatico e simbolico, come una tecnologia radicale che porta con sé non solo un mutamento significativo nell’ambito tecnico-applicativo, ma anche nel nostro rapporto quotidiano con la tecnologia, che a sua volta modificherà profondamente tutti gli ambiti del nostro vivere sociale.

Ma tornando al cuore del nostro discorso la fase di testing di ChatGPT non è stata e non è solo un esperimento collettivo ancora in atto; da un punto di vista sociale e culturale si è trattato soprattutto di un grandissimo *gioco di massa*.

In linea di massima sono due le componenti del gioco, seguendo una distinzione classica proposta da Roger Caillois (2010), a caratterizzare le recenti interazioni con questa applicazione: la competizione (che Caillois definiva *Agon*) e la mimesi (*Mimicry*). Durante il lancio di ChatGPT i mass media così come i social media sono stati veicolo e contenitore di migliaia di esempi di conversazione in cui gli utenti competevano con la macchina per far emergere le contraddizioni e i difetti dell’IA (*agon*), e in particolare la sua incapacità di imitare propriamente il pensiero e il ragionamento umano (*mimicry*). Al famoso Test di Turing, che questa forma di competizione/imitazione chiaramente richiama, è stata associata un’altra componente ludica essenziale: alla bassa capacità dell’IA di comprendere indovinelli,

emozioni, o banali correlazioni tra fenomeni, si sono moltiplicati i toni ilari, le conversazioni parossistiche, le frasi e le interazioni ironiche se non comiche che da un lato enfatizzavano il fascino e il potenziale di questa tecnologia, ma allo stesso tempo ne sminuivano la forza, così come la pericolosità percepita. Parafrasando e forzando un po' il concetto, il famoso *gioco dell'imitazione* di Turing si è trasformato nell'arena pubblica in un *gioco dell'umiliazione*; nei primi mesi di convivenza con ChatGPT non abbiamo solo testato e giocato con lei, ma ci siamo letteralmente *presi gioco* di lei. Dall'utente generico ad alcuni dei maggiori pensatori e critici sui temi dell'IA, tra cui Luciano Floridi, in molti si sono espressi e a loro modo anche vantati, nel pubblico così come nelle proprie cerchie sociali, della propria superiorità e dell'intelligenza umana rispetto a una stupida, piccola, bambina come ChatGPT.

Gioco e tecnologia, in particolare le tecnologie digitali, intrattengono da sempre uno stretto legame (Ortoleva 2013; Bory 2016). Negli anni Novanta Sherry Turkle parlava dei videogiochi come "oggetti transizionali" che accompagnavano la società verso la domesticazione del computer su scala globale<sup>2</sup>. Il giocare con le cose, siano esse naturali o artificiali, è da sempre un mezzo di produzione e co-produzione di immaginari tra attori umani e non umani, ed è da sempre un'attività essenziale per integrare qualsiasi forma di diversità o alterità all'interno del nostro ecosistema relazionale. Non c'è niente di sbagliato in questo modo di conoscere l'IA, ma dovremmo interrogarci più approfonditamente sui risvolti e gli effetti che questo tipo di interazione con essa ha sul piano economico-politico e tecnologico. Prendendoci gioco di prodotti ancora incompleti come ChatGPT-3, ci siamo in parte presi gioco di noi stessi, non accorgendoci che il lato infantile non era tanto dalla parte di un software che in realtà non ha nessuna consapevolezza di ciò che fa, del perché e del come, quanto piuttosto dal lato umano che con quel software ha giocato facendo finta di avere a che fare con un bambino.

La mossa di OpenAI di rilasciare un'intelligenza "stupida" e infantile è stata a dir poco efficace. Da un lato ha permesso all'azienda, e adesso a tutti i suoi competitori, di avviare e accelerare il processo di domesticazione dei modelli di linguaggio senza che ne avessimo paura e senza porre alcun freno iniziale alla loro diffusione, tranne in alcuni rari casi. Dall'altro lato OpenAI ha sfruttato il lavoro gratuito di milioni di



<sup>2</sup> Da un punto di vista del Design sono tanti gli esempi disponibili in tal senso, ben oltre i videogiochi per computer. Molte applicazioni come "solitario" e "campo fiorito" sui sistemi Windows sono state essenziali per alfabetizzare gli utenti su alcune innovazioni tecniche come l'uso del mouse (per questo motivo Bill Gates ne pretese la preinstallazione nei suoi sistemi). Allo stesso modo si pensi a come giochi come Snake negli anni Novanta abbiano alfabetizzato gli utenti all'uso della tastiera sul cellulare, o a come Angry Birds sugli smartphone sia servito a comprendere l'uso e a sviluppare una sensibilità per il touch screen. Ma il discorso di Turkle si avvicina particolarmente al nostro uso di Chat-Gpt perché tratta non solo il gioco ma anche l'esplorazione dei mondi discorsivi e normativi su cui erano basati: "Chi giocava scopriva il concetto di programma quando iniziava a studiare le regole al di là dei primi micromondi che era in grado di controllare. Poiché erano relativamente semplici, ci si poteva facilmente misurare con loro. Alcuni giocatori riuscivano perfino a meditarci su, dato che con la pratica le regole diventano semi-automatiche, e, per avere successo, occorreva raggiungere uno stato mentale in cui ci si lasciava alle spalle il mondo reale. In questo senso, i primi videogame erano oggetti di transizione; avevano quel certo tipo di trasparenza che forniva loro un'estetica moderna, ma ciò che richiedevano ai giocatori per vivere nei loro spazi ludici anticipava i requisiti psicologici propri della cultura della simulazione." (Turkle, 1997: 72)

utenti per “nutrire” la sua IA e renderla sempre migliore ed efficace. Dopo aver sfruttato mano d’opera a basso costo in paesi come il Kenya (Perrigo, 2023), Open AI ha fundamentalmente ottenuto una nuova manodopera gratuita pressoché sterminata.

L’interazione ludica di oggi con i modelli di linguaggio naturale e le IA generative non è quindi solo un passaggio chiave per addomesticare queste tecnologie; si tratta anche, e non è una novità nel capitalismo digitale, di un lavoro di massa che le rende più efficaci, e soprattutto più profittevoli per chi le produce.

#### 4. Metafore e potere

Le metafore e le analogie hanno il potere di plasmare l’immaginario sull’IA, e di conseguenza i suoi sviluppi sul piano scientifico-tecnologico nonché il suo uso quotidiano. Metafore e analogie, nel loro costruire, reiterare e radicare determinati immaginari, sono strumenti di potere. Nel maggio 2024 *Open AI* ha utilizzato una voce artificiale estremamente simile a quella dell’attrice Scarlett Johansson, che aveva interpretato uno dei più famosi esempi cinematografici di IA nel ruolo di Samantha per il film *Her* di Spike Jonze. L’attrice aveva già rifiutato in precedenza di prestare la sua voce all’azienda, ma in un atto di arroganza, se non di superbia, il CEO Sam Altman e Open Ai hanno comunque rilasciato una versione del loro prodotto estremamente vicina a quella che, nell’immaginario occidentale, è considerata un idealtipo narrativo della futura superintelligenza (Mikle, 2024).<sup>3</sup>

Si tratta solo dell’ultimo caso in cui le Big Tech si appropriano di personaggi, caratteristiche, se non di intere narrazioni nate dalla fantascienza e da narrazioni esterne al mercato digitale.<sup>4</sup> Così come con la metafora del bambino, anche la declinazione dell’IA in un determinato canone femminile è palesemente orientata a caricare l’immaginario di un senso di innocenza, se non di una servilità stereotipizzata, che dovrebbe favorire la sua integrazione nel quotidiano (Fossa, Sucameli, 2022). Metafore e analogie di questo tipo intervengono sull’immaginario sfruttando alcuni *topoi* della storia dell’IA, legittimando anche alcune forme di discriminazione e di *bias* che la contraddistinguono così come hanno contraddistinto il contesto sociale in cui sono maturate. In questo modo le metafore dell’IA come quella del bambino o della donna servile intervengono sui processi cognitivi, sulla costruzione dell’immaginario, ma anche e soprattutto, sul “sentire” collettivo ed individuale (Ricoeur, 1978). Questo processo favorisce un determinato “bilanciamento emozionale” nell’immaginario dell’IA: la paura, che è uno degli strumenti di potere ma anche di difesa centrali di una società, viene introiettata nell’immaginario dell’IA del futuro piuttosto che su quella del presente.

<sup>3</sup> Ha fatto clamore anche un post su X di Altman stesso contenente la sola parola *her*. Johansson ha denunciato l’avvenuto costringendo l’azienda di Altman a ritirare la versione con la voce incriminata.

<sup>4</sup> Si pensi in tal senso alla recente appropriazione, e storpiatura semantica, di Mark Zuckerberg-Meta dell’immaginario che ruota al cyberpunk sul metaverso e in particolare a partire da *Snow Crash* di Neal Stephenson.



Decostruire e criticare questo uso delle metafore e delle analogie antropomorfe dell'IA è uno degli obiettivi che gli studi sull'immaginario si devono porre oggi, senza però rischiare di polarizzare il dibattito snaturandole e rendendole puro appannaggio da una parte dei pericolosi attori del mercato tecnologico globale, dall'altro di un "utente generico" ignorante, vittima di un processo di influenza dettato dall'alto, o piuttosto dal "grande altro" descritto da Shoshana Zubooff (2015).

Per quanto si possa insistere sul fatto che le IA di oggi sono solo "pappagalli stocastici"<sup>5</sup> (Bender et al., 2021) abbiamo bisogno di queste metafore, come quella del bambino, per tentare di comprendere e dare un senso alla tecnologia e al modo in cui interagiamo con essa. Ma se la spinta all'antropomorfizzazione dell'IA è inevitabile, mettere in discussione il modo e il perché certe modalità, e soprattutto certe metafore, per raccontarla prevalgono su altre è altrettanto importante.

Si tratta di un'operazione complessa che le scienze sociali da sole non potranno mai concludere, in cui non solo c'è da riconoscere il *come* ma anche il *perché* di certe metafore, e soprattutto il *chi* sta orientando il nostro immaginario verso determinate concezioni della realtà algoritmica. C'è infine un ultimo aspetto cruciale da sottolineare, che riguarda il modo in cui noi, attraverso l'IA, stiamo raccontando la società in cui viviamo. Il paternalismo pervasivo, da sempre caratteristico della storia digitale (Balbi, 2024), l'innocenza e l'innocuità così come la prodigiosità dell'IA bambina, ci dicono che stiamo proiettando su di essa le stesse dinamiche di potere, nonché le stesse forme di de-responsabilizzazione individuale e collettiva che contraddistinguono il nostro tempo in diversi campi dell'agire sociale.

A partire da una lunga ricerca sull'errore umano nell'IA, l'antropologa Veronica Barassi (2024) ha fatto emergere la necessità di una "teoria dell'errore" che contribuisca a plasmare e riformulare il nostro immaginario di questa tecnologia; un immaginario che includa non solo la consapevolezza della fallacia, delle limitazioni e dei pericoli degli strumenti di cui la nostra società si sta troppo rapidamente attrezzando e su cui fa un eccessivo affidamento, ma anche un immaginario "sensibile" che riconosca l'intrinseca limitatezza e l'incompiutezza di chi pensa, progetta, nonché di chi usa e adotta l'"artificio".



---

<sup>5</sup> Il termine "stochastic parrot" sta per "pappagallo stocastico" ed è una metafora per descrivere la teoria secondo cui i modelli di linguaggio naturale, proprio come dei pappagalli, sebbene siano in grado di generare un linguaggio plausibile, non comprendono realmente il significato della lingua che elaborano.

## Bibliography

Balbi G. (2024), *The Digital Revolution: A Short History of an Ideology*. Oxford, Oxford University Press.

Barassi V. (2024), "Toward a Theory of AI Errors: Making Sense of Hallucinations, Catastrophic Failures, and the Fallacy of Generative AI", *Harvard Data Science Review*, retrieved on 21 May from:  
[https://assets.pubpub.org/81yb8b3r/Barassi%20\(2024\) Just%20Accepted-21725382067664.pdf](https://assets.pubpub.org/81yb8b3r/Barassi%20(2024) Just%20Accepted-21725382067664.pdf)

Bender E.M., Gebru T., McMillan-Major A., & Shmitchell S. (2021), "On the dangers of stochastic parrots: Can language models be too big?", In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pp. 610-623.

Bory P. (2016), "Ludus Ex Machina. Il ruolo della ludicità nella storia del computer", *Comunicazioni sociali*, Vol. 2: 317-329.

Bory P. (2020), *The Internet Myth: From the Internet Imaginary to Network Ideologies*. London, University of Westminster Press.

Bory S., & Bory P. (2015), "I nuovi immaginari dell'intelligenza artificiale", *Im@go: A journal of the social imaginary*, Vol. 6(1), 66-85.

Bostrom N. (2014), *Superintelligence: Paths, Dangers, Strategies*. Oxford, Oxford University Press.

Bucher T. (2019), "The algorithmic imaginary: Exploring the ordinary affects of Facebook algorithms", *Information, Communication & Society*, Vol. 20(1): 30-44.

Caillois R. (2010), *I giochi e gli uomini. La maschera e la vertigine*. Milano, Bompiani.

Castoriadis C. (1975), *L'istituzione immaginaria della società*, Bollati Boringhieri, Torino, 1975.

Cave S, Dihal K, Dillon S (Eds.) (2020), *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*. Oxford, Oxford University Press.

Chen F. (2018), "How can we design AI that we trust?", retrieved on 21 May from:  
[https://www.ted.com/talks/fang\\_chen\\_how\\_can\\_we\\_design\\_ai\\_that\\_we\\_trust](https://www.ted.com/talks/fang_chen_how_can_we_design_ai_that_we_trust)

Clark J. (2023), "Ex-Google AI expert says that 'unhinged' AI is the 'most powerful technology' since 'the atomic bomb'", *Fox News*, 1 March, retrieved on 21 May 2024 from:  
<https://www.foxnews.com/media/ex-google-ai-expert-says-unhinged-ai->



most-powerful-technology-since-atomic-bomb

Daly C. (2022), "AI baby' is born that can think and learn just like a three-month old human", *Daily Star*, 12 July, retrieved on 21 May from: <https://www.dailystar.co.uk/tech/news/ai-baby-born-can-think-27458575>

Fossa F. (2021), "Artificial agency and the game of semantic extension", *Interdisciplinary Science Reviews*, Vol. 46(4): 440-457.

Fossa F., & Sucameli I. (2022), "Gender bias and conversational agents: an ethical perspective on social robotics", *Science and Engineering Ethics*, Vol. 28(3): 1-23.

Future of Life Institute (2023), "Pause Giant AI Experiments: An Open Letter", retrieved on 21 May from: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>

GPT-3 (2020), "A robot wrote this entire article. Are you scared yet, human?", *The Guardian*, 8 September, retrieved on 21 May from: <https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3>

Goode L. (2018), "Life, but not as we know it: AI and the popular imagination", *Culture Unbound*, 10(2), 185-207.

Heffernan T. (2024), "The imitation game, the 'child machine,' and the fathers of AI", *AI & SOCIETY*, Vol. 39(1): 353-357.

Heikkilä M. (2023), How to talk about AI (even if you don't know much about AI), *MIT Technology Review*, retrieved on 21 May 2024 from: <https://www.technologyreview.com/2023/05/30/1073680/how-to-talk-about-ai-even-if-you-dont-know-much-about-ai/>

Hermann I. (2023), "Artificial intelligence in fiction: between narratives and metaphors", *AI & Society*, Vol. 38(1): 319-329.

Hu T.H. (2015), *A Prehistory of the Cloud*. Cambridge, MA, MIT Press.

Intagliata C. (2022), "This Artificial Intelligence Learns like a Baby", *Scientific American*, 26 August, retrieved on 21 May from: <https://www.scientificamerican.com/podcast/episode/this-artificial-intelligence-learns-like-a-widdle-baby/>

Lakoff G., & Johnson M. (2008), *Metaphors We Live By*. Chicago, University of Chicago press.



Levy S. (2023), "What OpenAI really wants", *Wired*, 5 September, retrieved on 21 May from: <https://www.wired.com/story/what-openai-really-wants/>

Mager A., & Katzenbach C. (2021), "Future imaginaries in the making and governing of digital technology: Multiple, contested, commodified", *New Media & Society*, Vol. 23(2): 223-236.

Marzo P.L., Meo M. (2013), Cartografie dell'immaginario, *Im@go - A Journal of the Social Imaginary*, 2 (1): 4-17.

Mikle T. (2024), "Scarlett Johansson Said No, but OpenAI's Virtual Assistant Sounds Just Like Her", *New York Times*, 20 May, retrieved on 21 May from: <https://www.nytimes.com/2024/05/20/technology/scarlett-johansson-openai-voice.html>

Mosco V. (2005), *The Digital Sublime: Myth, Power, and Cyberspace*. Cambridge MA, MIT Press.

Natale S. (2021), *Deceitful media: Artificial Intelligence and Social Life After the Turing Test*. Oxford, Oxford University Press.

Ortoleva G. (2012), *Dal sesso al gioco. Un'ossessione per il XXI secolo?*, Torino, Espress.

Perrigo B. (2023), "Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic", *Time Magazine*, 18 January, retrieved on 21 May from: <https://time.com/6247678/openai-chatgpt-kenya-workers/>

Peters J. D. (2015), *The Marvelous Clouds: Toward a Philosophy of Elemental Media*, Chicago, IL, University of Chicago Press.

Placani A. (2024), "Anthropomorphism in AI: hype and fallacy", *AI and Ethics*, Online First: 1-8.

Ricoeur P. (1978), "The metaphorical process as cognition, imagination, and feeling", *Critical inquiry*, 5(1): 143-159.

Sirisha (2022), "LaMDA is an 'AI baby' that will Outrun its Parent Google Soon", *Analytics Insight*, 27 June, retrieved 21 May 2024 from: <https://www.analyticsinsight.net/artificial-intelligence/lamda-is-an-ai-baby-that-will-outrun-its-parent-google-soon>

The Economist (2024), "Scientists have trained an AI through the eyes of a baby", retrieved on 21 May from: <https://www.economist.com/science-and->



[technology/2024/02/07/scientists-have-trained-an-ai-through-the-eyes-of-a-baby](https://www.wbur.org/news/2018/02/01/mit-artificial-intelligence)

Thys F., & Lepiarz J. (2018), "MIT Launches Initiative To Develop Artificial Intelligence That Learns Like Children", *WBUR*, 1 February, retrieved on 21 May from: <https://www.wbur.org/news/2018/02/01/mit-artificial-intelligence>

Tiku N. (2022), "The Google engineer who thinks the company's AI has come to life", *The Washington Post*, 6 June 2022, retrieved on 21 May from: <https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/>

Tong Z (2024), "World's first AI child unveiled in Beijing by award-winning Chinese scientist who left US for China", *South China Morning Post*, 2 February, retrieved on 21 May 2024 from: <https://www.scmp.com/news/china/science/article/3250529/award-winning-ai-scientist-who-left-us-china-creates-worlds-first-ai-child-beijing>

Turing A.M. (1950), "Computing machinery and intelligence", *Mind*, Vol. 59: 433-460. Traduzione italiana in: Somenzi V, Cordeschi R. (eds.) *La filosofia degli automi. Origini dell'intelligenza artificiale*, Torino, Paolo Boringhieri, 1986, 157-183.

Turkle S. (1997), *La Vita sullo Schermo*. Milano, Apogeo.

Wiener N. (1966), *God & Golem, Inc.: A comment on certain points where cybernetics impinges on religion*. Cambridge, MA, MIT Press.

Winstead T. (2021), *Black and White Photo of Human Hand and Robot Hand*. Retrieved on 10 October 1997 from <https://www.pexels.com/photo/black-and-white-photo-of-human-hand-and-robot-hand-8386422/>

Wyatt S. (2021), "Metaphors in critical Internet and digital media studies", *New media & society*, Vol. 23(2): 406-416.

Zuboff S. (2015), "Big other: surveillance capitalism and the prospects of an information civilization", *Journal of Information Technology*, Vol. 30(1): 75-89.





A JOURNAL OF THE  
SOCIAL IMAGINARY

# Coscienza e Intelligenza Artificiale: due ontologie dell'immaginario<sup>1</sup>

Maurizio Bonolis

Maurizio.bonolis@uniroma.it

*Dipartimento di Comunicazione e Ricerca Sociale | Sapienza Università di Roma*

Barbara Sonzogni

Barbara.sonzogni@uniroma1.it

*Dipartimento di Comunicazione e Ricerca Sociale | Sapienza Università di Roma*



## Abstract

*Consciousness and Artificial Intelligence: two ontologies of the imaginary*

The juxtaposition of the category of the Imaginary with what is commonly understood by the term Artificial Intelligence evokes representations of temporal dynamics based on the 'feeling of infinity'. However the dimensions of the Imaginary that come into play in this way are two and unmistakable. The first is that of the imaginary concerning the intangible horizon of intelligent systems design and processing programmes. The second meaning is that in itself obscure, and therefore no less indeterminate, which refers to the ability to simulate the existence of subjects endowed with decision-making autonomy. The concept of Artificial Intelligence refers here, in a word, to the ability to simulate consciousness. This is the real 'problem' of Artificial Intelligence. This is why the most authoritative sources of philosophical reflection and computer neurology continue to emphasise the disconcerting indefiniteness of consciousness, from a natural rather than an artificial point of view. The robotic simulation of attractiveness, for example in the form of sex toys, is both plausible and current, but less so the simulation of the aesthetic experience on the basis of which an attractive humanoid is axiomatically designed.

## Keywords

Consciousness | Intelligence | Simulation | Technical Progress | Extended Mind

---

<sup>[1]</sup> Pur essendo frutto di condivisione, a Maurizio Bonolis sono attribuibili i paragrafi 1, 2, 5 e 6; a Barbara Sonzogni i paragrafi 3 e 4; a entrambi gli autori il paragrafo 7.

## 1. La parola «immaginario»

**P**rima di accostare il concetto di *immaginario* al tema dell'Intelligenza Artificiale, nell'intento di mostrare in quale senso il primo possa designare una qualità della seconda, sembra opportuno convenire su una minima definizione di base del concetto, se non altro per via delle accezioni, con qualche esempio, alle quali esso si presta in rapporto ai diversi oggetti che ne sono *complemento di specificazione*. Lo scopo è porre l'accento su alcuni interrogativi che si ritiene sollevi, in particolare, la questione dell'immaginario dell'Intelligenza Artificiale. Si cercherà di sottolineare, a tale riguardo, che due, inconfondibili e reciprocamente irriducibili, sono le linee di tematizzazione che supportano questa ipotesi di lavoro: quella della *artificialità* della conoscenza e quella della *artificialità* della coscienza.

Per cominciare, «immaginario», può essere senz'altro un personaggio teatrale, mentre «immaginario» può essere ritenuto il piano logico sul quale si pongono le rappresentazioni eidetiche che, in riferimento a oggetti percepiti precedentemente, riescono a riprodurli con un dettaglio non riducibile alla nozione di memoria fotografica. E «immaginaria» può essere la *visio* inerte, semicosciente, che compare nelle esperienze ipnagogiche connesse all'affievolirsi degli stati di veglia, quando sta per arrivare il sonno. «Immaginaria», infine, può essere la raffigurazione di un oggetto del pensiero, il contenuto di una «intenzionalità» in senso husserliano, nello stesso senso in cui Wittgenstein sosterrà che il *pensabile* comporta sempre potersene fare un'immagine (*Tractatus*, 3001). Ma già nel *De Anima*, per Aristotele si dà unità di pensiero e immagine: pensare una cosa è renderla presente 'in immagine', nella modalità di un contenuto «pre-riflessivo» che pervade la nostra mente (Crespi, 1993: 14, 19).

Quest'ultima accezione vale anche a dar conto del fatto che intenzionalità condivise possono costituire «immaginario collettivo». Di ciò si sono occupati a suo tempo, pure nel contesto di prospettive disciplinari diverse, Gustav Jung (1934-1954), com'è noto, ma anche Edward Sapir (1929), in antropologia culturale. Poi, con non inferiore autorevolezza, Gadamer si è riferito a qualcosa di simile in prospettiva storicizzata, lamentando la crescente incidenza di produzioni culturali alimentate dalla comunicazione mediatica, cioè la dilagante incapacità di esprimere significati individuali spontanei, di scegliere parole non contaminate dai tormenti e dalla voracità delle simbolizzazioni di massa, ove tutto appare nella desolante luce del 'già detto': evoluzione grottesca della definizione heideggeriana del linguaggio come «casa dell'essere» (1946: 5). È chiaro che il fenomeno dell'opinione pubblica attinge esemplarmente a questa dimensionalità collettiva dell'immaginario: un secolo fa (era il 1922) Walter Lippman ne ha parlato in termini di «quadri mentali» (Pratkanis e Aronson, 1992: 64); Neil Smelser, più tardi, in termini di «credenze generalizzate» (*generalized beliefs*) (1962: 15-17).

Scopo di questo saggio è innanzitutto porre in luce come la dimensione immaginaria che investe il fenomeno dell'Intelligenza Artificiale rappresenti un caso esemplare del modo in cui oggetto di questo investimento simbolico è il progresso tecnico in generale. Tale esemplarità è esaltata dal fatto che ciò che denominiamo



col termine Intelligenza Artificiale, in quanto evoluzione della cibernetica e della teoria generale, dei sistemi alimenta le aspettative che nei più diversi ambiti applicativi il dinamismo tecnologico promette di soddisfare e sta mostrando di conseguire. Secondo obiettivo delle pagine che seguono, è sottolineare come questo processo espansivo ed evolutivo, proprio per le sue crescenti proporzioni renda evidente l'abisso che lo separa da ciò che non fa parte di esso ma di esso è presupposto teleologico, ossia la coscienza, che è qualcosa di diverso delle acquisizioni dei saperi 'intelligenti' della tecnica. Un certo spazio, nel corso del procedimento espositivo, sarà pertanto dedicato alla distinzione tra *conoscenza* e *coscienza*. Si mostrerà come il cammino dirompente dell'Intelligenza Artificiale rimarchi in questo senso il carattere di «problema difficile» che le scienze cognitive attribuiscono da tempo al fenomeno della coscienza e, conseguentemente, le esitazioni con le quali si guarda all'idea della progettabilità di una Coscienza Artificiale.



## 2. Immagine e oggetto

Ora, è chiaro che queste diverse modalità di impiego del concetto di *immaginario* (ma sono solo esempi), in tanto condividono la sua idoneità a designare entità in immagine, in quanto *immaginare*, innanzitutto, non è *percepire*. *Percepriamo*, condizionatamente, per via dei nostri sensi, il viso di Marylin ma non tutto il resto che quel viso richiama alla nostra mente e che recuperiamo subito grazie all'immaginazione, o meglio, grazie all'immaginario che quel volto attiva nella nostra memoria conservativa. In senso banale, qualcosa di simile accade quando percepiamo sensibilmente il lato di un solido ma non gli altri lati che non si vedono.

Può accadere che questo recupero non sia facile, né condiviso; in tal caso l'immaginario si riempie di oggetti diversi, a seconda della risorsa euristica che si pone al servizio dell'attività immaginativa, come nell'esempio dell'apologo dell'*Ignaro*, mediante il quale Galileo, nel *Saggiatore*, descrive i ripetuti stupori di chi, 'ascoltando' il ripetersi di un suono, scopre come esso provenga di volta in volta da fonti diverse, dal canto di un uccelletto allo zufolo di un pastore, dagli arpioni di sostegno del portale di una chiesa alle vibrazioni di un marranzano, dallo strofinio del bordo di un bicchiere al sibilo di un grillo (1622 [XXI]: 280-281). La 'morale della favola' è che, data la complessità del mondo, ognuno finirà con *l'immaginarsi* qualcosa di diverso, a seconda del proprio criterio («euristico») di autopersuasività (Kahneman e Tversky, 1984: 344-345).

Ne discende un importante risvolto ontologico, comune alle accezioni del termine «immaginario». Esso consiste in una specie di paradosso, illustrabile a partire da ciò che Sartre intende per «illusione di immanenza». Il fatto, cioè, che, seguendo Husserl, la coscienza sia sempre diretta a un oggetto, non è però da confondersi con una reificazione di quest'ultimo; perché l'oggetto resta del tutto assente, irreali, comparando solo sottoforma di rappresentazione analogica, resa possibile da un rimando a familiarità cognitive che precedono e rendono possibile l'atto di coscienza

come loro sintesi (Sartre, 1940: 97). L'aspetto paradossale del prodotto di immaginazione è proprio in questo suo essere finzione di una verità ontica: ribadiamo, l'immagine non è conoscenza dell'oggetto, semmai si può dire la presupponga, ne erediti il sapere – quale esso sia – che lo riguarda e che consente di istituirlo neutralizzando la sua *assenza*. Sartre, seguendo Husserl, fa l'esempio del «Centauro che suona il flauto», precisando che in nessun modo esso può sostituirsi a un «*Erlebnis* psichico» (*esperienza vissuta*): esso – egli scrive – «non esiste né nell'anima, né *nella* coscienza, né altrove: esso è niente, è del tutto immaginazione» (1936: 137, corsivo aggiunto). La coscienza, insomma, non *contiene* l'immagine di qualcosa: essa è l'immagine.

Ora, di esempi del tipo del Centauro, se ne potrebbero fare innumerevoli, ma è importante considerare il presupposto cognitivo dell'immaginario, quale esso sia, cioè la sostanziale irriducibilità che separa la *coscienza* dalla *conoscenza* dei suoi oggetti (Sartre, 1943: 17), opponendo la capacità della coscienza di attribuire «senso» a entità sprovviste di «riferimento», sebbene il materiale del quale sono fatti questi oggetti, né più né meno come la materia di cui sono fatti i sogni, abbia come tale un «riferimento»<sup>2</sup>. Il Centauro chiamato in causa da Husserl resta una figura mitologica, reale letterariamente quanto Biancaneve; il flauto, poi, non è affatto privo di «riferimento»: è la raffigurazione associata dei due elementi ad apparirci fantastica. Già immaginario è un corpo metà uomo, metà cavallo; se poi lo si raffigura alle prese con uno strumento musicale, non resta che sorridere. Questa raffigurazione: la comprendiamo, le diamo un senso, che è il senso del suo 'non-senso', ma non è uno sguardo nel buio. E ciò si rende possibile perché essa si compone di cose conosciute, sebbene combinabili in un'immagine che può benissimo essere un nulla irreali (Parisi, 2012: 73). Per Freud, così è per i sogni, fatti di 'materia' reale, ricordi, espressioni, vicende, persone, ciononostante talvolta uniti in un miscuglio vertiginoso, grottesco, comico oppure angoscioso, ma sempre irreali. Nei sogni come nelle attività immaginative, insomma, compaiono solo frammenti di esperienza, come tali 'reali' (persone, cose, ambienti), ma il loro insieme non è quasi mai 'reale' come invece può esserlo un ricordo (sebbene i ricordi, a volte appaiano sconcertanti per la loro parziale attendibilità).

### 3. Intelligenza Artificiale e *infinito*

Su questo aspetto di indeterminazione empirica e di connotazione fantastica dell'immaginario ha insistito a lungo Max Weber ravvisando, nel tema dell'infinito, il carattere intrinseco del progresso tecnico (1919: 17) quale movente di un'attesa, in parte riferita a evidenze note, in parte riferita alla loro oscurità futuribile. William Ogburn ha così ravvisato nel principio di continuità cumulativa il tratto specifico della

<sup>2</sup> La distinzione fra «senso» (*Sinn*) e «riferimento» (*Bedeutung*) è quella legata al tema fregeano (1892) che oppone gli enunciati categoriali (funzione *connotativa*), per richiamare attributi e proprietà descrittive associate a un termine rendendolo comprensibile, agli enunciati referenziali, che servono a *denotare* entità extralinguistiche, anche semplicemente denominandole (si veda a riguardo Dummett, 1973: xii-xiv).



modalità di mutamento della «cultura materiale» (1922). La tesi, che per comodità denominiamo di Weber-Ogburn, è che ogni condizione di stato della cultura materiale subirà le sorti recitate dal paradosso di Teseo: il suo essere, la sua identità, saranno sempre legati al tempo ma non si sa in che cosa si realizzeranno di volta in volta: ogni conquista della tecnica nasce già morta e il principio di tale indeterminatezza investe l'impossibilità di porre fine all'immaginario nel nome dell'idea di un 'compimento'. Il progresso tecnico è senza fine e ogni sua conquista è a sua volta priva di un fine che la giustifichi intrinsecamente. Questa sorta di assurdo destino si rivela mediante una irreversibile successione di stadi, ognuno dei quali supera esaustivamente e nello stesso tempo incorpora tutti quelli che lo hanno preceduto, gettandoli nell'oblio della loro ragione strumentale e consentendo loro, semmai, di sopravvivere come reperti storici, a disposizione del gusto estetico e dell'apprezzamento testimoniale del tempo perduto (Sonzogni, 2005).

Per questa via, com'è noto, Weber ha dato fiato alla sua presa di distanza da ogni filosofia della storia, incluse ovviamente le ingenuità escatologiche delle quali essa è stata investita, a esse contrapponendo la sua celebre definizione di «infinità priva di senso» (1904: 96). Di qui, la crescente attrattività prospettica di ciò che viene designato, oggi, con il termine Intelligenza Artificiale, quale evidente incarnazione dell'itinerario, ad un tempo promettente e oscuro, della modernizzazione tecnologica, del suo procedere 'senza fine'. Stupore, fascinazione estetica e preoccupazioni etiche sono le componenti principali (necessariamente multidirezionate) del travolgente immaginario, individuale e collettivo, che si appella, dilatandolo, a ogni esito incrementale della sua operatività e alla diffusa percezione dei più diversi ambiti nei quali si avvertono le sue ricadute, né più né meno in modo simile a quanto paventava Weber stesso al riguardo della sconfinata progressione culturale del «disincantamento» (1919: 20). Le sue sfere di designazione, i mondi – chiamiamoli così – della Artificialità, comprendono tutto ciò che promana in origine dall'universo dell'ICT e della tradizione cibernetica: l'ambito del *Machine Learning*, delle *Chatbot* e delle funzioni di *Virtual Assistant*, del marketing, della telechirurgia e della sanità in senso lato, della bioingegneria, della robotica industriale, dei controlli di sostenibilità e progettazione economico-ambientale, degli armamenti, della propulsione elettromeccanica.

L'apporto teoretico di Weber e di Ogburn alla comprensione di questo processo vale a mettere in luce, in particolare, la proprietà autocatalittica (autoreplicante) dell'immaginario associato alla percezione del dinamismo tecnologico, ossia alla percezione di indeterminatezza che alimenta i diversi orizzonti applicativi dell'innovazione artificiale. Ogni acquisizione ne annuncia un'altra, alimentando una promessa senza fine. La dimensione immaginaria, addirittura fantastica, è inevitabile perché il dinamismo tecnologico in atto è doppiamente anomico e intemperante: da un alto le singole innovazioni non hanno un fine, nel senso che rispondono compulsivamente a istanze performative incrementali direzionate ma prive di



definizioni di stato<sup>3</sup>. A sua volta, l'intero cammino evolutivo della tecnica resta conseguentemente senza fine, in senso weberiano refrattario a qualsivoglia idea di destino. Ignoto e ricco di promesse: buio e luminoso al tempo stesso.

#### 4. Intelligenza Artificiale e coscienza

Come in parte anticipato sopra, questo ordine di considerazioni riguarda però solo una delle due linee di tematizzazione inerenti all'immaginario del progresso tecnico in termini di percezione di infinito e indeterminazione: quella propriamente denominata della «Intelligenza». In sintesi, guardando al panorama dei contesti applicativi dell'Intelligenza Artificiale, è chiaro che essi si presentano come terreno di sviluppo di programmi di lavoro di tipo simulativo finalizzati alla progettazione e alla costruzione di soluzioni capaci di abilità altrimenti irraggiungibili, per grado di complessità e quantità di informazione presa in carico (*Big Data*). La connessione tra obiettivi di programma e definizione preliminare dei compiti simulativi è data dall'impiego di algoritmi, di esecuzione e di apprendimento statistico, che rientrano nella struttura dei programmi di lavoro sviluppando la loro trasformazione digitale e accrescendone le abilità applicative: essi ampliano la potenza computazionale di tali programmi, senza ovviamente modificarne gli sviluppi attesi, se non entro alternative previste dal linguaggio strutturale del programma stesso.

Resta senz'altro dirompente e suscettivo di incredule rappresentazioni immaginifiche, il richiamo alla specificità tecnica dell'Intelligenza Artificiale, consistente nel suo superamento della condizione di ripetitività meccanica, attraverso forme di apprendimento riconducibili ai modelli denominati di *Machine Learning* o *Deep Learning*, a seconda che il procedimento artificiale riguardi la ricorsività di un algoritmo oppure le opzioni sequenziali legate alle sue variazioni parametriche. In pratica, la logica dell'algoritmo è isomorfica a quella di un sistema di equazioni strutturali, fra loro collegate in ragione di valori-soglia dal rispetto dei quali dipende il grado di ottusità, o meno, del sistema. Sta al linguaggio e alla versatilità del software variare sequenze e formato del programma in relazione alle variazioni dei parametri, superando l'ottusità e reindirizzando lo sviluppo di un algoritmo come se fosse intervenuto l'apprendimento di una nuova modalità di applicazione, adeguata alle modifiche del contesto. Ma si tratta di potenzialità già contenute nella logica del procedimento, anche se può trarre in inganno la portata della loro realizzazione (*performance*). Ciò che discrimina è l'evidenza che la mente individuale, da sola, non ci arriverà mai, come non arriverà mai a leggere – per esempio – tutto ciò che a livello

---

<sup>3</sup> Secondo Parsons è su queste basi che si rende palese la convergenza tra la nozione weberiana (sopra ricordata) di «infinito», associata al tema del progresso tecnico e quella che Durkheim impiega come sinonimo di *anomia* («passione dell'infinito») (1897: 310). L'autore americano menziona in questo senso le ossessive aspettative che, come nella competizione tecnologica, governano, invece della definizione dello «stato finale» degli atti compiuti, la loro *direzione*: i competitori dell'innovazione, non sanno che cosa vogliono, ma sanno solo qual è l'orizzonte senza fine della loro operosità (1951: 192-193, corsivo aggiunto). Una sorta di 'mitologia della nuova frontiera'.

planetario afferisce a un argomento, poi traendone l'ottimizzazione sintattico-lessicale 'incarnata' in un timbro vocale definito (*Chatbot*). Mediante un software sintatticamente evoluto e un sistema di sensori ad alta versatilità, anche il più analfabeta dei parlanti, pur mantenendo il suo inconfondibile timbro vocale, si esprimerà in forma video-plastica digitale molto attendibile, mediante un inglese oxfordiano di rara bellezza. Niente è da ritenersi non ottenibile, che non sia programmaticamente ottenibile. È insomma chiaro che l'innovazione e l'apprendimento stesso hanno effetti tanto più inattesi quanto più ampia sia la capacità del programma di prevedere e rimodulare la propria morfogenesi, entro i limiti di un dinamismo strutturale che, comunque, non può andare – insistiamo – al di là dei confini del suo perimetro simulativo. Sempre più largo, certo, ma sempre perimetro.

Veniamo così alla seconda linea di tematizzazione dell'immaginario. Si è fatto giustamente osservare, infatti, come razionalità e adeguatezza tecnica rivelino di essere ispirate a fini e motivazioni «che non possono essere il prodotto delle elaborazioni interne di un sistema chiuso» (perimetro algoritmico) (Manzotti e Tagliasco, 2002: 93). In altri termini, la definizione di contenuto degli scopi che dettano le operazioni strumentali a essi afferenti sono proprietà semantiche intenzionali, riferibili agli estremi di «un soggetto unitario, capace di fare esperienza di sé stesso e del mondo circostante (...), in grado di produrre motivazioni e fini, valori soggettivi e sensazioni. In poche parole, la coscienza» (*ibidem*)<sup>4</sup>. Ecco il secondo ambito dell'Immaginario. In ultima analisi, si sostiene al riguardo, «l'intelligenza ha cioè bisogno di una *volontà* che la guidi» (*ibidem*) e tale ragione teleologica non può che farsi valere al di là di ciò che in essa è comprovata adeguatezza strumentale e accurata istruzione protocollare. Seguendo il sentiero lungo il quale Weber ha sottolineato come «il corso di azione» ("l'agire reale") si svolga «nella gran parte dei suoi casi in una oscura semicoscienza o nell'incoscienza del suo senso intenzionato (...) istintivamente o in base all'abitudine», sicché «L'agire effettivamente, cioè pienamente e consapevole e chiaro, è in realtà sempre soltanto un caso limite» (1922b: 19), si può concedere che tale consapevolezza (*awareness*) resti una mera possibilità; il che significa, però, che questo secondo atto del pensiero «è già diverso dal puro pensare». Esso «è qualcos'altro. È un altro processo» (Leonardi, 1957: VIII). Un neurochirurgo sa ovviamente come procedere a una craniotomia e questa abilità competenziale può, in versione 'debole' essere supportata da risorse tele-endoscopiche o, in versione 'forte', sostitutamente da un impianto robotizzato. Ma questa dimensione della decidibilità prognostica è estranea al suo eventuale interrogarsi sull'esito dell'operazione, quale ne sia il responso introspettivo, o sul giudizio relativo alla propria affidabilità.



<sup>4</sup> La distinzione concorda con la tesi di Searle («esperimento della stanza cinese»), secondo la quale i procedimenti dell'Intelligenza Artificiale sono capaci di sintassi, mentre la mente umana possiede anche una semantica, ossia è in grado di *comprendere* quello che sta facendo, manipolando simboli (come una macchina), però nello stesso tempo producendo significati per questi simboli, cosa che la macchina non può fare (1992: 61).



È stato Sartre, su basi teoretiche non dissimili da quelle appena richiamate, a sottolineare così, ripetutamente, l'inconfondibilità del concetto di *coscienza* con quello di *conoscenza*, proponendo a carico del primo il ricorso all'esperire autoriflessivo, in luogo della categorizzazione oggettivante, che appartiene al secondo (Farina, 2013: 94)<sup>5</sup>. E c'è voluto più di un secolo, da Mary Shelley (1818) a Collodi (1883), a Isaac Asimov (1950) e Ridley Scott con *Blade Runner* (1982), poi Spielberg (2001), prima che la letteratura e la filmografia di fantascienza abbandonassero l'illusoria convinzione che gli esseri artificiali siano sempre dotati di coscienza: cioè, non di intelligenza ma – prima di tutto – della «capacità di essere soggetti autonomi di decisioni e obiettivi» (Manzotti e Tagliasco, 2002: 89). Solo con Kubrick (1968), il rifiuto del postulato antropomorfista, sostituito dall'idea di un essere artificiale, privo di corporeità e però dotato di coscienza ed emozioni similumane, solleverà il problema, ancora attuale, che spinge a ritenere l'annessione delle categorie del modello teleologico (scopi, valori), all'Intelligenza Artificiale e alla robotica, ammissibile soltanto metaforicamente (ivi: 94). Si tratta di argomento che va di pari passo con quello della legittimità di revocare in dubbio le interpretazioni naturaliste o internaliste della vita mentale (Sonzogni, 2022), da Chalmers segnalate come il «problema difficile» della coscienza: ovvero l'arduo compito di spiegare come processi meramente fisici siano in grado di dar conto della complessità e della autonomia ontologica dell'esperienza cosciente, nell'inevitabile rispetto della dualità cartesiana-hegeliana, magari poi con l'intento di dissolverla («eliminativismo»)<sup>6</sup>.

Gilbert Ryle, tra gli antesignani di questo orientamento critico, per rimarcare l'insuperabilità di Cartesio, in un celebre contributo di settant'anni fa faceva rilevare – a titolo esemplificativo – che un conto sono le idee che uno 'ha in testa' (pura metafora), un conto i suoni che turbano l'udito e che si possono confermare con uno stetoscopio (significato letterale 'materialista') (1949: 33-35).

## 5. Immaginario «esternalista». Con alcune domande

Tutto ciò non vuol in nessun modo sollecitare allarmismi o – all'opposto – entusiastiche rivalutazioni di preminenza dell'umano sulla macchina, ma semplicemente invitare a un'ipotesi di lavoro che abbia come punto di partenza una distinzione: quella tra Intelligenza Artificiale e Coscienza Artificiale, tra *conoscenza* e *coscienza*. È chiaro che il cammino non è agevole: se è difficile immaginare che cosa sarà fra tre secoli un lap top, se ancora potrà chiamarsi così, quasi umanamente impossibile sembra immaginare una coscienza artificiale, una condizione di libero arbitrio, costruita (forse), ma senz'altro in nessun senso dipendente, dal *nostro* libero

<sup>5</sup> Le parole di Sartre sono in armonia anche con ciò che sostiene Freud quando, *nell'Interpretazione dei sogni*, afferma che la coscienza non è «nient'altro che un organo di senso per la percezione di qualità psichiche» (1899: 560). Su queste basi, è chiaro che la distinzione tra *conoscenza* e *coscienza* si ritrova anche nella irriducibilità di quest'ultima al concetto di «introspezione» (Ryle, 1949: 157-159).

<sup>6</sup> Si vedano, al riguardo, i termini della discussione a suo tempo enunciati in Agazzi (1967) e recentemente richiamati in Lavazza e De Caro (2013).

arbitrio. Difficoltà, questa, che ha suscitato autorevoli pronunciamenti sulla coscienza *tout court*: in termini di entità «ineffabile» (Rorty, 1982), di una 'cosa' che, sulle orme di Rainer Rilke, si rileva sia «più grande del cielo, più profonda del mare» (Edelman, 2004), o semplicemente «un mistero» (Quine, 1987; Fodor, 1998). Difficoltà, che resiste nel tempo, visto che, a quasi cinquant'anni dal celebre saggio di Warren McCulloch, che si domandava «perché la mente fosse nella testa» (1951), veniamo informati che il filosofo americano Stephen Stich ha ritenuto di rilanciare il problema dell'osservabilità della coscienza citando la vecchia storia dell'ubriaco che sta cercando le chiavi dell'auto sotto la luce di un lampione e risponde «laggiù nel parcheggio» a chi gli chiede dove le ha perse, mentre quando gli vien chiesto perché allora lui le cerchi proprio qui, risponde «Perché qui c'è la luce» (Carli, 1997: 211). Probabilmente, un buon punto di partenza è il dubbio, quanto meno il dubbio, che la coscienza abbia un *dove*.

Ora, un indirizzo teoretico che si muove in questa direzione, cioè dubitando della ammissibilità di un posizionamento spaziale della coscienza, ha preso forma muovendo proprio dalle posizioni assunte da David Chalmers a favore di una riformulazione interpretativa della vita mentale, nota come prospettiva «esternalista». Di fatto, il problema di un *dove*, chiedersi quale sia la sede (fisica) del pensiero, significa ammettere il primato materialista e quindi monista di una *res extensa*. Rifiutare l'interrogativo vuol dire accettare la prospettiva cartesiana e affermare che la mente non ha un *dove*, rinunciando all'ipotesi del posizionamento spaziale e abbracciando la prospettiva della cosiddetta «mente estesa» (Clark e Chalmers, 1998: 9; Amoretti, 2013: 228-229). In definitiva, mentre l'internalismo sostiene che la vita della mente dipende dal sistema nervoso centrale, l'esternalismo sostiene che essa dipende da qualcosa che non appartiene in esclusiva alla regione cerebrale, ma dinamicamente e unitariamente anche all'ambiente fisico e al mondo sociale, sicché non si può assumere una rigida separazione tra la mente stessa e ciò che è – in senso lato – il suo *contesto*. Anche se non appartiene alla schiera degli esternalisti, una concessione ad argomenti che si aprono a questa prospettiva è ravvisabile – sebbene con qualche oscurità – anche in Rorty, nella sua triade della modernità, centrata su atteggiamenti capaci di «solidarietà, ironia, contingenza» (1989).

In effetti, Chalmers perviene al modello della mente estesa dopo aver introdotto, in precedenza, una distinzione antinaturalistica (*The Two Mind-Body Problems*) che, accogliendo i progressi diagnostico-strumentali della neurofisiologia clinica, propone di riservare loro il campo dell'indagine sulla coscienza e sui suoi «correlati neurali» come *easy problem*, mentre disegna la sfera di un dominio di *hard problems*, ossia l'ambito della spiegazione scientifica inerente agli aspetti qualitativi e soggettivi dell'esperienza cosciente, da ritenersi incompatibili con le categorie di qualsivoglia monismo materialista (1996: 24-26). Tale problema, da alcuni autori menzionato come inerente alla categoria dei *qualia*, si basa sull'evidenza, posta in risalto dal filosofo John Searle, che «ogni stato cosciente è caratterizzato da una sensazione qualitativa», sia che si tratti di una percezione esterna, sia che si tratti di una rappresentazione cognitiva o del pensiero astratto (1998: 39).



Coloro i quali, con Chalmers, condividono questo orientamento, in un certo senso si può dire che cerchino una via di mezzo, separando, da un lato la questione dei correlati neurali che si accompagnano alla effettualità delle idee, dall'altro le idee stesse che, nel loro contenuto, nulla hanno a che vedere con le proprietà della 'scatola cranica' (per usare l'espressione di Schopenhauer). Per questa via, volendo fare un esempio, è indiscutibile che l'eccitazione (o il suo contrario) produca effetti associati a un riscontro neurologico, ma resta fermo che il movente soggettivo dell'eccitazione stessa, l'attrazione contemplativa di un'immagine piuttosto che un'altra, appartiene a ordini di dominio fenomenico estranei a quel riscontro. Ad esempio, sarebbe veramente arduo ricondurre la percezione della bianchezza a una azione del cervello o il fascino estetico di un volto, confrontato con quello sconcertante di un altro, a qualcosa che abbia origine da entità sinaptiche<sup>7</sup>.

Fatta salva questa inevitabile ma non dogmatica riaffermazione del dualismo classico (Bonolis, 2022: 68-69), restano ugualmente in gioco alcuni interrogativi, riguardanti il tema dei correlati analitici, chiamato in causa dalla nozione di «mente estesa». Se il richiamo alla ipotesi di una funzione di costitutività da parte del mondo esterno (ambiente fisico, mondo sociale) (Clark e Chalmers, 1998: 18) nei confronti della coscienza è degna di vera considerazione, l'eredità sociologica classica non consente di eludere la questione inerente al livello logico dell'opzione externalista.

Durkheim ha convincentemente insistito sulla natura sociale – i suoi epigoni hanno aggiunto *culturale* – dell'individuo in quanto entità storicizzata (Bonolis, 2020: 48) e, in questo senso, evolutivamente formata: una sensibilità soggettiva, che non contrappone l'idea del libero arbitrio, l'idea paretiana di «integrità dell'Io» al mondo esterno, ma coglie un piano di indistinguibilità tra esterno ed interno: siamo nel nostro tempo, siamo 'il' nostro tempo, *siamo* storia. Ora, applicando il suggerimento tipologico di Susan Hurley, filosofa britannica, a proposito delle varietà referenziali dell'esternalismo (2010), questo livello *formale* del ragionamento durkheimiano, è altra cosa rispetto al piano *sostanziale* sul quale il grande autore francese, mediante la sua celeberrima elaborazione di una teoria della solidarietà («organica») corrispondente alla modernizzazione individuale, poneva l'esigenza di salvaguardia dell'individuo nei confronti di se stesso (*anomia*). Elaborazione, poi da Durkheim stesso abbandonata per via della palese impercorribilità del modello (Hinkle, 1960: 271). Il problema è che, sebbene formalmente e ineludibilmente l'individuo sia socialmente *costituito* (livello 'formale'), ciò non toglie che egli possa esprimere un atteggiamento di ostilità e di indifferenza nei confronti del suo mondo sociale, ossia che questo stesso mondo (livello 'sostanziale'), pur essendo *in* lui, alimenti desideri di



<sup>7</sup> Se si ignora questa barriera ontologica si va incontro a un'antinomia simile a quella narrata in *Das Parfum* (Süskind, 1985). Il protagonista del romanzo rivela la condotta di un individuo (Jean-Baptiste Grenouille) che vive un rapporto con il mondo di tipo esclusivamente olfattivo, ma questo vincolo decade nel momento in cui egli coglie in una emanazione genitale femminile un sentore completamente avulso dalla sostanza organolettica (batterica) che attiva quella emanazione e che – invece – evoca in lui l'immaginario di una fanciulla, ossia 'sapori' e 'odori', irrimediabilmente connessi a un'esperienza estetica. Potremmo dire: i profumi immaginari dell'amore, non della corporeità. Il protagonista, improvvisamente 'umano', testimonia della intrascendibilità del «sesto senso» al quale si riferisce la Arendt allorché, richiamandosi a Tommaso D'Aquino, denuncia i limiti dell'empirismo britannico (1971-1978: 133).

repulsione quando non di odio o efferato cinismo: tutte cose che hanno diversi nomi, da Auschwitz alla *acrasia* aristotelica, dalla «banalità del male», alla letale apatia e alla «indifferenza» umana, narrate in letteratura da Alberto Moravia (1949; 1960; 1965).

Non dovrebbero servire altri esempi.

D'altro canto, l'illusione solidaristica di Durkheim contiene un elemento irriducibile all'argomento della sua stessa fragilità teoretica o quanto meno non avulso dalla legittimità di un sentimento di speranza: a ciò, in ultima analisi, si ispirava la già accennata posizione di Richard Rorty e, prima di lui ma anche prima di Durkheim, la laconica esortazione di Dostoevskij, quando, nella metafora del «sottosuolo» narra del vissuto appartenente alla condizione dell'uomo solo, escluso da ogni vita di relazione e condannato alla propria indifferenza e alla propria misoginia. Un autentico inferno, che peraltro è sempre in agguato visto che in generale, nella solitudine, anche l'uomo dotato delle più alte virtù e qualità – egli scrive – e delle migliori fortune materiali «unicamente per il gusto di sbagliare, sarà capace di commettere una mascalzonata (...) e a bella posta desidererà la più dannosa delle sciocchezze» (1864: 229).



## 6. Una coscienza artificiale? L'immaginario impossibile

Resta però aperto un interrogativo, in sé poco rassereneante. Traendo spunto dalla inequivocabile sentenza di Dostoevskij, vi sarebbe ragione di chiedersi infatti se, allora, non sia proprio la solitudine, nella finzione narrativa di Kubrick, l'origine del proposito omicida di Hall 9000. L'assunzione di una condotta moralmente esecrabile verrebbe a dipendere dall'isolamento della coscienza, dalla sua condizione di naturale impossibilità di autogenerarsi. Seguendo le citate ispirazioni di Dostoevskij, di Rorty, di Durkheim, potremmo condividere con le insuperate parole di quest'ultimo grande autore la convinzione che l'individuo, «preso a sé, è troppo poca cosa, non è un fine sufficiente alla sua attività. (...) Quando non si hanno altri obiettivi all'infuori di noi stessi, non possiamo sfuggire all'idea che i nostri sforzi siano destinati, in fondo, a perdersi (...) Ma l'annullamento ci terrorizza e in tali condizioni non sapremmo trovare il coraggio di vivere, di agire, di lottare quando di tanta fatica nulla deve restare» (1897: 250).

Nello stesso tempo, sottolineare così l'importanza del legame sociale non può significare che l'immaginario di una coscienza artificiale confidi sui benefici di un'emulazione, che annullerebbe la natura stessa dell'autonomia fenomenica della coscienza. Da un lato la coscienza è inseparabile dal suo presupposto dinamico, quello del libero arbitrio, ossia dal principio della libertà di scelta di fronte al bene e al male, come più volte ribadito da Sartre (1946). Dall'altro, l'insegnamento durkheimiano vale a sottolineare che è nell'esperienza dei legami e delle interazioni della vita sociale che si riempie di significati e simbolizzazioni la dimensione potenzialmente vuota della coscienza stessa, con esiti dinamicamente connessi a questa apertura 'estesa', nel senso di Chalmers. È chiaro che, date queste premesse, diventa difficile immaginare la via della costruzione di una coscienza artificiale e la

nostra riflessione, come quella di altri autori al riguardo, non può che proporre il terreno di una ricerca degna di questo nome: il postulato del libero arbitrio è più che un'opera aperta. Per definizione, può portare ovunque e c'è da sperare che l'intersoggettività della vita associata proponga nei fatti, evolutivamente, abiti e modelli di pensiero intanto condivisi, in quanto desiderabili. Ma senza che, antinomicamente, il libero arbitrio venga insegnato e appreso, dando luogo ai paradossi già impietosamente illustrati da Hegel (del tipo: 'amami spontaneamente, non perché te lo chiedo io'). Dell'impossibilità di comporre questo dilemma, ha dato conto proprio Durkheim nella delineazione dell'ipotesi di una morale laica, capace di sostituire la ragione all'assoluto delle ontologie religiose, così residuando però la spinosa questione di capire come potesse prendere forma una siffatta nuova eticità, oltretutto a fronte delle tette nuvolosità culturali dei primi del Novecento (la Francia dell'affare Dreyfus e, non molto tempo dopo, la nascita del Nazionalsocialismo in Germania).

L'interrogativo è sinteticamente lontano da ogni possibilità di immaginazione.

È forse difficile ma non più di tanto impossibile coltivare l'immaginario di una piena sostituibilità tecnologica e di una vera scomparsa di intere branche della medicina chirurgica, così come della viabilità stradale orizzontale. Ben diverso è l'immaginario che abbia come oggetto la decidibilità, su basi metacognitive, degli atti che danno luogo e ottemperano al pensiero del progresso tecnico: diverso, cioè, è l'immaginario di una coscienza, artificiale, che guardi - in quanto tale - ma non appartenga, agli orizzonti dell'intelligenza - essa sì - plausibilmente artificiale.

Gli enunciati richiamati nel precedente paragrafo sottolineano gli estremi di una distinzione che può allora essere accolta rilevando che il livello logico della nozione di *coscienza* stia a quello della nozione di *conoscenza* (che coincide con il dominio dell'Intelligenza Artificiale) come la funzione metalinguistica stia a quella della funzione referenziale, quali proprietà del linguaggio articolato. La separazione delle due funzioni, che risale a Karl Bühler, ripresa da Roman Jakobson e poi da Popper (1978: 42), ammette ovviamente che in entrambi i casi il linguaggio sia veicolo istitutivo del suo oggetto, rappresentandolo ai parlanti anche quando non è materialmente nella loro prossimità (come quando nel cuore di una distesa sahariana si discutesse di orsi polari) o addirittura per negarne l'esistenza<sup>8</sup>. Solo che quando la funzione è metalinguistica o, se vogliamo, autoreferenziale, l'oggetto è il linguaggio stesso; il che accade, ad esempio, per *spiegare* il significato di una parola o di una regola grammaticale. Il linguaggio qui *parla* del linguaggio. In questo senso, similmente ai termini di questa distinzione, la logica della conoscenza è invece di tipo osservativo, anche quando il suo scopo è pratico, come nell'Intelligenza Artificiale, perché il suo oggetto è extralinguistico, si trattasse pure di analisi, non del linguaggio, ma del comportamento linguistico o della comunicazione verbale. L'aspetto extralinguistico si riferisce al fatto che il rapporto tra linguaggio e oggetto è qui



<sup>8</sup> Non c'è al riguardo espressione simbolica migliore di quella, celeberrima, che risale ad Heidegger, che Heidegger stesso considera quale paradosso di riconoscimento ontico di ciò che si afferma ne sia privo per statuto semantico (esempio estremo del *niente*) (1946: 21-23).

attuato mediante nesso referenziale. La logica della coscienza resta invece di tipo normativo-riflessivo: non persegue, né postula un'imputazione causale.

Come abbiamo sottolineato in precedenza, a proposito della distinzione tra il pensiero e la consapevolezza del pensare (paragrafo 5), in senso figurato conoscenza e coscienza rimandano all'uso incomparabile di due vetri davanti ai nostri occhi: l'istanza della conoscenza richiede di guardare il mondo attraverso trasparenze, più o meno nitide, più o meno deformanti, anche se trattasi solo di comprendere le origini del proprio mal di testa. L'istanza della coscienza richiede un rispecchiamento: non di attraversare il vetro per andare oltre, analizzando un oggetto, ma di dedicarsi alla propria immagine riflessa, in una *visio* esclusivamente interiore, alla ricerca del senso che attribuiamo al guardare, mettendo tra parentesi ciò che si vedrebbe sostituendo lo specchio con una finestra per guardare al di là.

Sul piano analitico, essi (i due concetti) si oppongono omologicamente a ciò che rende inconfondibili giudizi di fatto da giudizi di valore, dove, per definizione, l'evidente simulabilità computazionale di ciò che appartiene all'ambito dei primi non è attuabile nel campo dei secondi (Sonzogni, 2022). Quando si tratta di spiegazione scientifica o di progettazione empirica conseguente (progettare un ponte richiede disporre di saperi concernenti l'aerostatica, i calcoli del cemento armato, informazioni geofisiche e mille altre acquisizioni afferenti), i risultati, attesi necessariamente nel contesto del *probabile*, hanno validità accettata fino a prova contraria, cioè condivisa e intersoggettivamente plausibile, ma mai assoluta. L'argomento rientra notoriamente nelle categorie della nozione kuhniana di «paradigma» (1962). Il consueto esperimento didattico della rifrazione della luce non è in nessun senso una verità ontica ma resta quanto di meglio si reputa consolidato, nei limiti weberiani di un sapere necessariamente «attuale» (Weber, 1904: 59), per comprendere l'arcana percezione del bastoncino spezzato che affiora dall'acqua.

Quando, invece, ci si muove sul terreno del *certo*, dell'«etica della convinzione» e del dominio dell'imperativo sintetizzabile nel domandarsi, non «come posso conoscere» bensì «come devo agire»<sup>9</sup>, è invece del tutto improprio (e inutile) servirsi delle connessioni illustrate da una o più equazioni strutturali (vedi *supra*): qui ci si sta chiedendo altro e cioè se sto sentendo il suono delle campane o no, se percepisco sensibilmente una sollecitazione corporea o, ancora, se faccio mia una credenza, un precetto normativo, un determinato criterio etico al quale fare riferimento per agire, per orientare miei determinati atti. È chiaro che, nel caso dei giudizi di valore, il *certo* della consultazione riflessiva e, analogamente, dell'intento che è alla base dell'introspezione, in ultima analisi non è più incontrovertibile del *probabile* che caratterizza la conoscenza, empirica o teoremativa<sup>10</sup>, ma sconta tale limite dal lato



<sup>9</sup> Sull'ispirazione kantiana della distinzione analitica dei concetti di fatto e valore, e delle due nozioni di «etica» (della *responsabilità* e della *convinzione*) ad essi omologhe, il riferimento è ovviamente a Weber (1917), fatte salve le riserve espresse al riguardo di quella distinzione, dai più credibili esponenti del pragmatismo americano (Putnam, 1990: 61). La distinzione, concettualmente affine, tra logica del *certo* e logica del *probabile* è estesamente esposta da Sartre ne *L'imaginaire* (1940).

<sup>10</sup> Mentre una *verità sintetica* trova il suo limite nella eventualità di una prova contraria, i confini di una *verità analitica* (ad esempio nel campo della geometria) restano quelli assiomatici del procedimento che istituisce l'assioma stesso (Agazzi, 1961; Boudon, 1968: 22-23).

della sua non esaustività in termini di coincidenza con l'azione conseguente. Il problema centrale della coscienza è in gran parte tutto qui, in questa non coincidenza, quando il *certo* si infrange contro la debolezza della volontà, l'inganno muto e desolante dell'acrasia. Voci autorevoli della filosofia analitica ne hanno dato un'esemplificazione inimitabile, al riguardo, citando la figura di Pietro, che sa perfettamente che cosa vuol dire tradire e cionostante non sa astenersi dal farlo, sicché 'prima che il gallo canti', per tre volte lo farà (Anscombe, 1957: 154; Davidson, 1980: 88).

In generale, la tradizione filosofica ha dato più volte conto di questa inefficacia pragmatica dei giudizi di valore utilizzando l'apologo dei disturbi che, nel modello aristotelico dell'inferenza pratica, interrompono la relazione di continuità tra le premesse del sillogismo e le sue *conclusioni*, quando queste sono di ordine pratico o, ancora, quando ciò che conclude l'inferenza non è una *consecutio* logico-proposizionale ma un'azione (*Etica Nicomachea*, Libro VII, 3). La pretesa che il sapere inerente a un campo d'azione possa dar luogo a una condotta d'azione ad esso ottemperante è il più delle volte destinata al fallimento. Le cose non accadono come quando l'idraulico, convocato in emergenza, mette in atto il suo sapere per sturare un lavandino. Spesso intercorrono fattori di «disturbo» (*Etica Nicomachea*, Libro VII, 3) del tipo di quelli emozionali e passionali che per Weber depongono a favore della impossibilità di non essere irrazionali (1906: 157) e, all'estremo, a favore della ricorrente eventualità che, sapendo distinguere tra il Male e il Bene, si imbrocchi cionostante la prima via, del tutto «acratice», sempre nel senso di Aristotele.

Sta di fatto che, così, tra premesse di conoscenza competenziale e agire conformemente a esse, possono intercorrere distorsioni della conformità, attesa o desiderata, che la vanificano totalmente o in parte. D'altra parte, è chiaro che le fonti di irrazionalità o di mancata «adeguatezza» nel rapporto mezzi-fini, per usare un'altra espressione di Weber, sono potenzialmente innumerevoli: vanno da quelle concernenti gli stati mentali del soggetto agente a quelle che investono le condizioni meramente pratiche, sopravvenienti, dell'adeguatezza. Inoltre, si rende altresì discutibile l'attendibilità dei saperi che intrinsecamente sarebbero guida dell'azione: chi fuma, compresi i medici che fumano, sa del danno cancerogeno cui va incontro, ma qual è l'eshaustività di tale sapere? In certi momenti il fumo aiuta, sul piano nervoso e su quello psicologico e questo, a suo modo, è un altro sapere. È vero che poi le cose si complicano se entrano in gioco condizioni di dipendenza neurale, per cui il sapere di prevenzione diventa un pallido ricordo. Ma per arrivare a questo stadio bisogna averne superati alcuni nei quali il sapere era già pallido, quanto meno di debole efficacia, il che porta ritenere che la conoscenza del danno cardiocircolatorio non esaurisca tutto quello che c'è da sapere per decidere di fumare o di continuare a farlo.

E torniamo così allo spettro delle cose cattive che però, seguendo Mandeville, producono effetti non necessariamente esecrabili: ad esempio, cose che, come il fumo, attenuano il 'male di vivere', del quale ha parlato Montale, con toni non dissimili da quelli della descrizione dell'ossessione del dolore della vita, in Schopenhauer.





Da queste considerazioni, il tema dell'autonomia del concetto di coscienza guadagna evidenti ragioni di supporto e può ulteriormente rafforzare le espressioni di indeterminatezza che – come si è visto – da varie fonti si accompagnano al tentativo di identificarne lo statuto fenomenico. D'altra parte, le incertezze che investono l'argomento e delle quali l'ipotesi della mente estesa fornisce un'indubbia e feconda rappresentazione, ma non un contributo risolutivo dal lato della progettazione della artificialità della coscienza, restano insuperate anche per via di una plausibile lettura del concetto di *acrasia* come chiave del dinamismo enigmatico della coscienza e come fondamento esemplare del libero arbitrio. Prima dell'esplicita posizione assunta in questo senso da John Searle (1980) e ancora prima da Sartre (1946), merita che si rammenti al riguardo l'ostentata paura nei confronti della dimensione del *certo*, ostentata da Nietzsche (1965: 19) con un atteggiamento chiaramente precursore del fascino dell'*indeterminatezza* nella teoria matematica di Shannon e Weaver: del loro richiamo all'attrattiva dell'informazione, tanto più 'vera' ove densa di ciò che si potrebbe dire, in luogo della noia di ciò che si sa e si dice (1949: 8). Per questo, sembra lecito ribadire che non c'è alcuna possibilità, al riguardo, di ricorrere a una qualche forma di sapere consolidato, essendo il problema principale proprio quello della identità fenomenica della coscienza, ossia di guardare a una connettività in grado di coglierne l'indeterminatezza. Difficile, al punto da ritenersi inimmaginabile, salvo fantasie cinematografiche che saremmo ansiosi di prendere in considerazione, ipotizzare il modo in cui la coscienza entri in gioco nelle costituenti di un profilo robotico, da un lato rendendosi chiara la sua autonomia dai saperi 'intelligenti', dall'altro mostrandosi intelligibile la sua 'condanna alla libertà' e il suo paradossale diritto di abbandonarsi al Male.

## 7. Epilogo e dilemmi

In conclusione, è sostenibile che il problema della coscienza, in relazione al suo posizionamento nell'ambito delle neuroscienze, come nello specifico delle voci dell'immaginario sull'Intelligenza Artificiale, ribadiamo, resti quello che, condivisibilmente, ormai vent'anni fa è stato definito da David Chalmers «un problema difficile» (1995: 201). Il sentiero di ricerca della naturalizzazione è sempre attraente, nel senso che «se si riuscisse davvero a chiarire quali processi cerebrali siano sottesi alla comparsa della coscienza, sarebbe del tutto irragionevole continuare a credere che la coscienza stessa non sia una funzione biologica attivata, al pari di molte altre, da normali processi cerebrali» (Nannini, 2014: 332). Fino ad allora, restano però tanti i motivi per non abbandonare la diagnosi, per quanto laconica, di Chalmers: qui si vuole sottolineare quello che tra essi si presenta in una particolare forma dilemmatica e che, allo stato attuale e che allo stato attuale del dibattito su questi temi, non sembra consentire qualcosa di più di un'ipotesi di ricerca. Va tenuto conto che, come in tutti questi casi, il dilemma è cornuto, anche se l'espressione, ispirata a Giovanni Buridano, quasi volgare, resta molto suggestiva: è da rimarcare che, in sostanza, la difficoltà, il 'difficile problema' di concepire una

coscienza artificiale riposa sulla proclamata evidenza di non riuscire a offrire una accettabile definizione del fenomeno della coscienza *tout court*, preliminarmente alla questione del perseguimento o della progettazione della sua artificialità. Il che significa che la coscienza stessa potrebbe rivelarsi un'entità esitante, in balia dello spettro acratice del tipo di quello che un altro detto popolare, coniato da un giornalista italiano vissuto nella prima metà del Novecento, menziona la degenerazione della debolezza umana nei termini di un soggetto 'buono a nulla e capace di tutto' e che Dante, forse anche Shakespeare, non riterrebbe poi tanto esagerato al di là della ricorrente mania di pensare che solo gli italiani siano così.

Sono le incertezze e le edulcorate santificazioni della categoria di cui ci stiamo occupando, tra filosofi e neuroscienziati eminenti, a sollecitare i termini di una referenzialità cosmopolita: anche un'ipotetica coscienza artificiale porterebbe con sé lacrime e sangue, analogamente alla storia della presenza dell'uomo nel mondo. Al cospetto di un'ipotesi del genere, già Durkheim aveva abbandonato la fiducia nell'idea, di origine comtiana, in una «morale laica» che, in epoca moderna, fosse in grado di sostituirsi allo spirito religioso (credenze nel soprannaturale), in virtù dell'educazione intellettuale e della sua persuasività circa l'importanza della disciplina e della solidarietà umana disinteressata (1925). Com'è noto, Durkheim lascerà poi cadere l'argomento, peraltro molto fecondo nella continuità dei suoi epigoni, in base alle stesse riserve con le quali in precedenza egli stesso aveva preso le distanze dall'utopismo delle tesi hobbesiane sul Leviatano e da quelle rousseauiane sulla *volonté générale* (1895: 114): in definitiva, il senso delle sue riserve era che, passi l'idea di educare un popolo, ma quali garanzie offrono al riguardo gli educatori?

In alternativa, le strade (dilemmatiche) sono due. Cercare di superare la distinzione tra coscienza e conoscenza, incorporando la prima nella seconda (Brooks, 1991), andando però incontro ad aporie delle quali s'è detto e che è lecito dubitare riescano a trattare la coscienza in termini di programmazione cognitiva (equazioni e *switch* di logica deontica). Oppure, pur non sapendo che cosa essa sia, nel naturale come nell'artificiale, la coscienza resta un enigma, ossia il regno dell'indeterminatezza e non c'è allora altra strada che affidare l'artificiale all'umanonaturale, alla guida di chi governa ogni programmazione dei suoi algoritmi (Intelligenza Artificiale), senza ambizione alcuna di ordine formativo o didattico. Ma ciò non può convincere più di tanto: forse finirà così, come per i conflitti bellici, gli armamenti, i rapporti di forza tra nazioni e governi, le efferatezze dittatoriali. La coscienza non artificiale, umana, non può farci dormire tranquilli. Il fatto è che, nel nostro Immaginario, non trova posto qualcosa di diverso.



## Bibliography

- Agazzi E. (1961), *Introduzione ai problemi dell'assiomatica*, Milano, Vita e Pensiero.
- Agazzi E. (1967), "Alcune osservazioni sul problema dell'intelligenza artificiale", in *Rivista di Filosofia neoscolastica*, 59 (1): 1-34.
- Amoretti M. C. (2013), "Sui correlati fisici della coscienza: neurali o estesi?", in Nannini S e Zeppi A. (a cura di), *La mente può essere naturalizzata?* Colle di Val d'Elsa, Salvietti & Barabuffi, 2013, 217-233. ISBN: 978-88-97082-49-1.
- Anscombe G. E. M. (1957), *Intention*, Oxford, Basil Blackwell; tr. it., *Intenzione*, Roma, EDUSC, 2016.
- Arendt H. (1971-1978), *The Life of the Mind*, New York, Harcourt, Brace, Jovanovich, 2 voll.; tr. it., *La vita della mente*, Bologna, il Mulino, 1987-2009.
- Aristotele (IV a.C.), *Etica Nicomachea*, C. Mazzarelli (a cura di), Milano, Bompiani, 2000.
- Bonolis M. (2020), "La mente umana e la volontà di credere", in Bonolis M. (a cura di) (2020), *Lo «spirito» nell'anima. Indagine sulla spiritualità giovanile*, Milano, Angeli, 17-77.
- Bonolis M. (2022), "La naturalizzazione della mente e il suo contrario", in Bonolis M. e Lombardo C. (a cura di) (2022), *Sociologia degli stati mentali (Teoria e ricerca)*, Angeli, Milano, 19-81.
- Boudon R. (1968), "A quoi sert la notion de 'Structure?'", tr. it., *Strutturalismo e scienze umane*, Torino, Einaudi, 1969.
- Brooks R. A., (1991), "Intelligence without representations", in *Artificial Intelligence Journal*, 47(1-3) (January), 139-159.
- Carli E. (a cura di) (1997), *Cervelli che parlano. Il dibattito su mente, coscienza e intelligenza artificiale*, Milano, Mondadori.
- Chalmers D. (1995), "Facing Up the Problem of Consciousness", in *Journal of Consciousness Studies*, 2 (3), 200-219.
- Chalmers D. J. (1996), *The Conscious Mind. In Search of a Fundamental Theory*, Oxford, Oxford University Press.
- Clark A. (2003), *Natural Born Cyborgs: Technologies and the Future of Human Intelligence*, Oxford, Oxford University Press.



Clark A., David J. C. (1998), "The Extended Mind", in *Analysis*, 58: 7-19, DOI:10.2307/3328150.

Crespi F. (1993), *Evento e struttura. Per una teoria del mutamento sociale*, Bologna, il Mulino.

Davidson D. (1980), *Actions and Events. Philosophical essays*, Oxford, Oxford University Press; tr. it., *Azioni ed eventi*, Bologna, il Mulino, 1992.

Dostoevskij F. M. (1864), *Zapiski iz podpol'ja*, Letchworth, Bradda Books; tr. it., *Memorie dal sottosuolo*, in (Id.) *Il romanzo del sottosuolo*, Torino, Einaudi, 1974, 203-322.

Dummett M. A. E. (1973), *Frege: philosophy of language*. New York, Harper & Row.

Durkheim É. (1925), *L'Education morale*, Paris, Félix Alcan; tr. it., *L'educazione morale*, in *Il suicidio e l'educazione morale*, Torino, Utet, 1969.

Durkheim É. (1895), *Les règles de la méthode sociologique*, Paris, Félix Alcan; tr. it., *Le regole del metodo sociologico - Sociologia e filosofia*, Milano, Comunità, 1963.

Durkheim È. (1897), *Le Suicide. Etude de sociologie*, Paris, Félix Alcan; tr. it., *Il suicidio*, in *Il suicidio e L'educazione morale*, Torino, Utet, 1969.

Edelman G. M. (2004), *Wider than the Sky: The Phenomenal Gift of Consciousness*, New Haven, Connecticut, Yale University Press; tr. it., *Più grande del cielo. Lo straordinario dono fenomenico della coscienza*, Torino, Einaudi, 2004.

Farina G. (2013), "Coscienza e conoscenza: Sartre lettore di Freud", in *Bollettino Studi sartriani*, IX: 89-109.

Fodor J. A. (1998), *In Critical Condition: Polemical Essays on Cognitive Science and the Philosophy of Mind*, Cambridge, Massachusetts, MIT Press.

Frege G. (1892), «Über Sinn und Bedeutung», *Zeitschrift für Philosophie und philosophische Kritik*, n. 100, 25-50; tr. it., *Senso, funzione e concetto: scritti filosofici 1891-1897*, Roma, Laterza, 2001.

Freud S. (1899), *Die Traumdeutung*, Lipsia-Vienna, Franz Deuticke Verlag; tr. it., *L'interpretazione dei sogni (Opere, vol. III)*, Torino, Boringhieri, 1966.



Gadamer H. G. (1966), *Die Universalität des hermeneutischen Problems*, Tübingen, Mohr; tr. it., *L'universalità dell'ermeneutica*, in A. Babolin (a cura di), *Filosofi tedeschi d'oggi*, Bologna, il Mulino, 1967.

Galilei G. (1622), *Il Saggiatore* (1890-1909), in Galilei G. (1890-1909), *Le opere di Galileo Galilei* (a cura di Antonio Favaro), Edizione Nazionale sotto gli auspici di S. M. il Re d'Italia, Barbera, Firenze (20 volumi). VI: 197-372.

Heidegger M. (1946), *Brief über den «Humanismus»*, Frankfurt am Main, Klosterman; tr. it., *Lettera sull'umanesimo*, Milano, Adelphi, 1995.

Hinkle R. C. jr. (1960), "Durkheim in American Sociology", in Kurt H. Wolf (ed. by) (1960), *Essays on Sociology and Philosophy (by Emile Durkheim et al.)*, New York-London, Harper & Row (1964), 267-295.

Hurley S. (2010), "The varieties of externalism", in Richard Menary (ed. by) (2010), *The extended mind*, Cambridge, Massachusetts, MIT Press.

Jung C. G. (1934-1954), "Über die Archetypen des kollektiven Unbewussten", in Jung C.G., *Gesammelte Werke Band 9.1*, Olten-Zürich, Walter Verlag; tr. it., *Gli archetipi dell'inconscio collettivo*, Torino, Bollati Boringhieri, 1977.

Kahneman D., Tversky Amos (1984), "Choices, values, and frames", *American Psychologist*, 39: 341-350.

Kuhn T. (1962), *The Structure of Scientific Revolutions*, Chicago, The University of Chicago Press; tr. it., *La struttura delle rivoluzioni scientifiche*, Torino, Einaudi, 1969.

Lavazza A., De Caro M. (2013), "Neuroetica e neuroscienza: tra spiegazione della vita e cura della mente", *Rivista Internazionale di Filosofia e Psicologia*, 4(3): 249-251. DOI:10.4453/rifp.2013.0026

Leonardi F. (1957), *Introduzione allo studio del comportamento sociale*, Milano, Giuffrè.

Manzotti R., Tagliasco V. (2002), "Si può parlare di coscienza artificiale?", *Sistemi intelligenti*, 1: 89-107.

McCulloch W. S. (1951), "Why the Mind is in the Head", in Jeffress, Lloyd Alexander (ed. by.), (1951), *Cerebral Mechanisms in Behavior*, New York, Wiley; tr. it., *Perché la mente è nella testa?*, in Karl S. Lashley et al. (1969) *La fisica della mente*, Torino, Boringhieri, 112-136.

Moravia A. (1949), *Gli indifferenti*, Milano, Bompiani.



Moravia A. (1965), *L'attenzione*, Milano, Bompiani.

Nannini S. (2014), "La seconda svolta cognitiva in filosofia della mente", *Reti, saperi, linguaggi - Italian Journal of Cognitive Sciences*, I (2), 320-334.

Nietzsche F. (1965), *Epistolario (1865-1900)*, a cura di Barbara Allason, Torino, Einaudi (1969).

Ogburn W. (1922), *Social Change. With Respect to Culture and Original Nature*, New York, Huebesch.

Parisi I. (2012), "L'immaginario", *Im@go. Rivista di studi sociali sull'immaginario*, I: 71-74.

Parsons T. (1951), *The Social System*, New York, Free Press of Glencoe; tr. it., *Il sistema sociale*, Milano, Comunità, 1965.

Popper K. R. (1978), *The Place of Mind in Nature*, San Francisco, Harper & Row; tr. it., *Il posto della mente nella natura*, in *Tre saggi sulla mente umana*, Roma, Armando, 1994.

Putnam H. (1990), *Realism with a Human Face*, Cambridge (Mass.), Harvard University Press; tr. it., *Realismo dal volto umano*, Bologna, il Mulino, 1995.

Pratkanis A. R., e Aronson E. (1992), *Age of Propaganda. The Everyday Use and Abuse of Persuasion*, New York, Henry Holt and Company; tr. it., *Psicologia delle comunicazioni di massa. Usi e abusi della persuasione*, Bologna, il Mulino, 1996.

Quine W. v. O. (1987), *Quiddities: An Intermittently Philosophical Dictionary*, Belknap Press of Harvard University Press; tr. it., *Quidditates: quasi un dizionario filosofico*, Milano, Garzanti, 1991.

Rorty R. (1982), "Mind as Ineffable", in Elvee Richard (ed. by) (1982), *Mind in Nature*, San Francisco, Harper and Row, 60-95.

Rorty R. (1989), *Contingency, Irony, and Solidarity*, Cambridge, Cambridge University Press; tr. it., *La filosofia dopo la filosofia: contingenza, ironia e solidarietà*, Roma-Bari, 1989.

Ryle G. (1949), *The Concept of Mind*, London, Hutchinson & Co. Ltd; tr. it., *Il concetto di mente*, Roma-Bari, Laterza, 2007.

Sapir E. (1929), "The Unconscious Patterning of Behavior in Society", in Ethel S. Dummer (ed. by) (1929), *The Unconscious: A Symposium* New York, Knopf, 114-142.



Sartre J-P. (1936), *L'Imagination*, Paris, Puf; tr. it., *L'immaginazione – Idee per una teoria delle emozioni*, Milano, Bompiani, 1962, 5-151.

Sartre J-P. (1940), *L'imaginaire. Psychologie phénoménologique de l'imagination*, Paris, Gallimard; tr. it., *Immagine e coscienza. Psicologia fenomenologica dell'immaginazione*, Torino, Einaudi, 1960.

Sartre J-P. (1943), *L'Être et le Néant: essai d'ontologie phénoménologique*, Paris, Gallimard; tr. it., *L'essere e il nulla*, Milano, Il Saggiatore, 1965.

Sartre J-P. (1946), *L'Existentialisme est un humanisme*, Paris, Les Editions Nagel S.A.; tr. it., *L'esistenzialismo è un umanismo*, Milano, Mursia, 1970.

Searle J. R. (1992), *The Rediscovery of the Mind*, Cambridge-London, The MIT Press; tr. it., *La riscoperta della mente*, Torino, Bollati Boringhieri, 1994.

Searle J. R. (1998), *Mind, Language and Society. Philosophy in the Real World*, New York, Basic Books; tr. it., *Mente, linguaggio, società. La filosofia nel mondo reale*, Torino, Rosenberg & Sellier, 2009.

Shannon C. e Elwood W. W. (1949), *The Mathematical Theory of Communication*, Urbana (Illinois), University of Illinois Press; tr. it., *La teoria matematica delle comunicazioni*, Milano, Etas Libri, 1971.

Smelser N. J. (1962), *Theory of Collective Behavior*, New York: Free Press.

Sonzogni B. (2005), "Eterogeneità globale e modernizzazione. Attualità di Ogburn", *Sociologia e ricerca sociale*, 78: 77-116.

Sonzogni B. (2022), "Simulabilità computazionale degli stati mentali. Uomo e macchine alla luce di alcuni dualismi interpretativi", in Bonolis M. e Lombardo C. (a cura di) (2022), *Sociologia degli stati mentali (Teoria e ricerca)*, Milano, Angeli, 185-201.

Süskind P. (1985), *Das Parfum – Die Geschichte eines Mörders*, Zürich, Diogenes; tr. it., *Il profumo*, Milano, Longanesi, 1985.

Weber M. (1919), *Wissenschaft als Beruf*, Tübingen, Verlag J.C.B. Mohr; tr. it., *La scienza come professione*, in *Il lavoro intellettuale come professione*, Torino, Einaudi, 1948, 3-43.

Weber M. (1904), "Die «Objektivität» sozialwissenschaftlicher und sozialpolitischer Erkenntnis"; tr. it., "L'«oggettività» conoscitiva della scienza sociale e della politica sociale", in Weber (1922a), *Gesammelte Aufsätze zur Wissenschaftslehre*, Tübingen,



Maurizio Bonolis e Barbara Sonzogni  
Coscienza e Intelligenza Artificiale

Verlag J.C.B. Mohr; tr. it., *Il metodo delle scienze storico-sociali*, Torino, Einaudi, 1958 (1967), 53-141.

Weber M. (1906), "Kritische Studien auf dem Gebiet der Kulturwissenschaftlichen Logik", in M. Weber (1922); tr. it., *Studi critici intorno alla logica delle scienze della cultura*, in *Il metodo delle scienze storico-sociali*, Torino, Einaudi, 1958 (1967), 143-237.

Weber M. (1922b), *Wirtschaft und Gesellschaft*, Tübingen, Verlag J.C.B. Mohr; tr. it., *Economia e società*, Milano, Comunità, 1961.



# La politica dell'intelligenza artificiale *general purpose*: immaginari socio- tecnici, democrazia e *policy frame* nel processo decisionale della regolazione europea ("EU AI ACT" - 2022-2024)



Riccardo Corsi  
riccardocorsi9@gmail.com  
*Gran Sasso Institute*

Ernesto d'Albergo  
enresto.dalbergo@uniroma1.it  
*Dipartimento di Scienze sociali ed Economiche | Sapienza Università di Roma*

## Abstract

*The politics of general-purpose artificial intelligence: socio-technical imaginaries, democracy and policy frames in European regulatory decision-making ('EU AI ACT' - 2022-2024).*

The article proposes a dialogue between different conceptualizations of 'imaginary' and 'socio-technical imaginaries' and an interpretive analysis perspective of political processes to understand European artificial intelligence (AI) policy. The object of analysis is the decision-making process of the "EU AI Act", approved in spring 2024. The investigation is focused on General Purpose AI and its consequences for democracy. The operationalization of the concepts shed light on the relationships between meaning construction and material stakes in the negotiations between political and social actors. The imaginary of 'risk to democracy' and the category of 'systemic risk' helped to structure the policy frame of the EU AI Act. The imaginary thus plays a role at once cognitive and normative in sense-making processes and social practices such as a policy process, helping to structure relations between subsystems of society (economic, social, cultural, techno-scientific, academic, political).

## Keywords

Imaginary | Policy frame | Artificial Intelligence | Systemic Risk

## 1. Introduzione: capire la politica europea dell'intelligenza artificiale alla luce dell'immaginario del rischio per la democrazia<sup>1</sup>

**L'**intelligenza artificiale (AI) è entrata rapidamente nelle agende politiche su scala internazionale, con una varietà di iniziative prese da Nazioni Unite, G7, OECD, Unione europea (UE), stati nazionali e governi locali. Oggetto di questa complessa politica dell'AI sono: strategie competitive e cooperative di investimento per sviluppare la tecnologia; politiche per favorire l'uso di AI da parte di autorità e amministrazioni pubbliche; tentativi di regolarne politicamente la produzione e l'uso (privato e pubblico), contemperando la valorizzazione delle potenzialità e la mitigazione dei rischi associati. Come in ogni processo politico, tutti gli attori coinvolti - gli eletti e i componenti di governi su varie scale, i produttori e i dettaglianti di modelli e sistemi di AI, i ricercatori accademici e non, la società civile no-profit - hanno bisogno di definire, individualmente e in modo condiviso, problemi e possibili soluzioni ad essi associate. Per capire la costruzione di queste definizioni e le modalità con le quali esse danno poi luogo ad azioni, abbiamo circoscritto l'analisi al processo di decisione di una regolazione normativa dell'AI da parte dell'UE - "Regolamento del Parlamento europeo e del Consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (...)", o regolamento sull'intelligenza artificiale (EU AI Act), approvato definitivamente nella primavera del 2024. Al suo interno, abbiamo isolato un oggetto emergente: la *General Purpose AI* (GPAI), o AI per finalità generali, ossia sistemi con un'ampia gamma di usi possibili, sia previsti, sia non previsti dagli sviluppatori e che possono essere applicati a compiti diversi in vari campi, spesso senza modifiche e messe a punto sostanziali. Questi sistemi, anche definiti "modelli di base" (*foundation models*), sono caratterizzati dall'uso (pubblico e privato) come modelli pre-addestrati per sistemi di AI più specializzati<sup>2</sup>. Ne sono esempio i *large language models*, divenuti famosi per l'uso di massa di Chat GPT, dal 2022. Le aspettative di enormi e sconosciuti impatti di questi modelli e sistemi hanno indotto Consiglio e Parlamento europei a regolarli con un approccio specifico.

Abbiamo ulteriormente messo a fuoco le modalità con le quali sono stati definiti i rischi della GPAI per la democrazia e le sue pratiche, uno degli oggetti più importanti e delicati di questo processo politico. Per identificare e interpretare le modalità di costruzione dei problemi abbiamo utilizzato la concettualizzazione dell'"immaginario" e degli "immaginari socio-tecnici", facendola dialogare con una prospettiva di analisi *argumentative* e *interpretive* dei processi politici (Fischer, Gottweis 2013). Ciò ha comportato valutare la compatibilità fra diverse definizioni, interpretazioni e operazionalizzazioni del concetto di "immaginario" (Castoriadis,



<sup>1</sup> Il lavoro è stato scritto insieme dagli autori, come risultato di ricerca e riflessioni sviluppate in comune. Convenzionalmente, tuttavia, è possibile attribuire i paragrafi 2 e 3 a Riccardo Corsi e i paragrafi 1, 4 e 5 a Ernesto d'Albergo.

<sup>2</sup> Cfr. Future of Life Institute, *General Purpose AI and the AI Act*, Maggio 2022 (<https://artificialintelligenceact.eu/wp-content/uploads/2022/05/General-Purpose-AI-and-the-AI-Act.pdf>).

1994; Taylor, 2004; Sum, Jessop 2013) e quelle relative ai concetti di *frame* e *policy frame* (Rein, Schön, 1993). A questo fine abbiamo esplorato (par. 2): da un lato la concettualizzazione associata alla differenziazione degli "immaginari socio-tecnici" (Mager, Katzenbach 2021); dall'altro i primi studi dell'AI e della sua regolazione attraverso gli immaginari, la molteplicità di rappresentazioni e visioni che li compongono e i *frame* (Katzenbach, Richter, Schafer 2023) cui essi danno luogo. Tale molteplicità evidenzia sia visioni di futuri desiderabili, sia distopie associate ad usi socialmente ed eticamente pericolosi (Jasanoff, Kim 2015; Sartori, Bocca, 2023). Nel discorso pubblico e nei processi politici di regolazione dell'AI questi ultimi prendono la forma di "rischi" (Paul, 2023) di portata e intensità variabili.

Siamo partiti dall'ipotesi che, come accade in altri casi, anche in questa arena regolativa i processi di semiosi attraverso cui si costruiscono gli immaginari rendano possibile agli attori di "andare avanti in un mondo complesso", popolato sia da idee, sia da interessi (Sum, Jessop 2013). In particolare, abbiamo cercato di capire se e come gli immaginari del rischio della GPAI abbiano consentito ai *policy maker* di identificare l'oggetto da regolare, riducendone la complessità per poter orientare le decisioni e raggiungere accordi in questa specifica politica dell'AI. A questo fine abbiamo esaminato le proposte di modificazione del testo iniziale della Commissione UE (2021) avanzate da Consiglio UE (2022) e Parlamento (2023) e il successivo "trilogo", concluso con l'adozione del regolamento (2023-2024) (par. 3). La nostra analisi si è concentrata sia sulla modalità semiotica di costruzione di senso (rappresentazioni dei rischi per la democrazia), sia sulle poste in gioco materiali (economiche e relative agli usi di AI) oggetto di negoziato. Per capire come entrambe hanno concorso a definire il *frame* della regolazione, abbiamo analizzato l'influenza esercitata da diversi attori sociali nei confronti dell'arena politico-istituzionale attraverso il ruolo degli immaginari, ricostruiti soprattutto attraverso le fonti secondarie di una letteratura scientifica che riflette e discute sulle rappresentazioni dei rischi dell'AI prodotte da esperti, esponenti della società civile organizzata e interessi economici (par. 4), rispecchiandone a sua volta i temi.

Nelle conclusioni è brevemente discusso, a partire dalle evidenze selezionate, il ruolo svolto dagli immaginari socio-tecnici sia nella formazione di *policy frame* e norme giuridiche, sia nelle relazioni fra sottosistemi sociali che vengono attivate dai processi di regolazione politica di attività economiche e di cambiamenti tecnologici.

## 2. Gli immaginari sociotecnici dell'intelligenza artificiale

Il filosofo Cornelius Castoriadis, uno dei teorici che più hanno contribuito alla comprensione degli immaginari sociali, sostenendo la loro irriducibilità ai soli significati culturali (Adams et al., 2015), ne ha rivelato il carattere creativo e creatore, capace di farsi pratica concreta di invenzione e istituzione della società (Castoriadis, 1987). I "significati sociali immaginari" (Arnason, 2014), nella riflessione di C. Castoriadis, pur essendo irriducibili al solo linguaggio, trovano in esso l'espressione privilegiata e sono un elemento essenziale della costituzione dell'identità di una



collettività (Castoriadis, 1987: 147-8). Inoltre, sono fondamentali anche del modo in cui ogni società definisce i "suoi bisogni, in quanto iscritti nelle sue attività, nel suo fare" (*ibidem*). I significati sociali immaginari rilegano rappresentazioni, affetti e intenzioni dominanti in una società (Castoriadis, 2007), contribuendo ad articolarne le strutture interne (Castoriadis, 1987: 150). Dunque, tra potenza immaginativa e società intercorre un rapporto di co-costituzione, in cui i significati, i discorsi e le pratiche sociali emergono attraverso un'attività complessa di costruzione di senso.

Così l'immaginario sociale rivela la sua natura di concetto politico, in quanto permette di comprendere in che modo la società istituisce sé stessa (Castoriadis 1987, Browne e Diel, 2019).

Nell'accezione di C. Taylor (2004: 2), l'immaginario sociale va concepito "non come un insieme di idee; piuttosto come ciò che rende possibile, attraverso la loro costruzione di senso, le pratiche della società". Ovvero, esprimendo l'idea di C. Castoriadis di potenza creatrice e istituyente a un tempo dell'immaginario, esso produce la "comprensione comune che rende possibile le pratiche comuni e un ampio senso di legittimazione" (ivi: 23). In quanto pratica di costruzione di senso, l'immaginario sociale permette di rendere conto dell'aspetto mutevole delle società, rivelandosi utile all'identificazione del cambiamento sociale (Adams et al., 2015), nel nostro caso riguardante le trasformazioni radicali associate all'accelerazione dell'innovazione tecnologica. Ciò significa che la costruzione di senso permette di agire e di legittimare il contesto dell'azione, anche in un processo politico in cui si esercita potere.

Nell'approccio del *Cultural Political Economy* sviluppato da N. Sum e B. Jessop (2013) e adottato da R. Paul (2024), l'azione degli attori in un mondo incerto e complesso è guidata da due modalità di riduzione della complessità attraverso immaginari: la *semiosi* o costruzione interattiva di senso, e la strutturazione (Sum, Jessop, 2013). Il primo aspetto riguarda il modo in cui nella costruzione di senso discorsiva si accentuano o silenziano alcuni aspetti specifici del mondo sociale, mentre il secondo appartiene al mondo degli interessi materiali ed extra-semiotici (*ibidem*).

Un importante tentativo di mettere in relazione i mondi sociali della scienza e della politica attraverso il ricorso all'ampio ed eterogeneo sfondo teorico degli immaginari (Adams et al., 2015) è stato compiuto attraverso il concetto di "Immaginario socio-tecnico". Sviluppato da S. Jasanoff, in collaborazione con Sang-Hyun Kim, nell'ambito degli studi sulla scienza e tecnologia (Jasanoff, Kim, 2009; Jasanoff, Kim, 2013; Jasanoff, Kim, 2015), a partire dall'intuizione per cui l'attività scientifica e tecnologica si trova in una relazione di co-costruzione con l'ordine politico (Jasanoff, Kim, 2004). Nel progresso tecnologico e scientifico sono implicite visioni sui futuri collettivi, sulle finalità pubbliche, il bene e il male comune, delle visioni positive o negative della società, il cui senso si co-costruisce relazionalmente anche attraverso le scoperte scientifiche e le invenzioni tecniche, che a loro volta si nutrono del sostegno e dell'allocazione di risorse da diverse parti della società. Gli immaginari socio-tecnici in questa prospettiva corrispondono a "visioni collettive, istituzionalmente stabilizzate e pubblicamente rappresentate di futuri desiderabili,



animate da una comprensione condivisa delle forme di vita sociale e dell'ordine sociale raggiungibili attraverso i progressi della scienza e della tecnologia e il loro sostegno" (Jasanoff, 2015: 4). Per l'autrice "va da sé che *gli immaginari dei futuri desiderabili o desiderati correlano, in maniera tacita o esplicita, con il loro opposto - condivise paure di danni che possono occorrere attraverso l'invenzione e l'innovazione*" (ivi: 6, corsivi aggiunti).

Questa ambivalenza costitutiva degli immaginari socio-tecnici si rende visibile nelle diverse modalità con le quali essi contribuiscono a rendere possibili, dotandole di senso, anche le pratiche sociali che stiamo studiando. La pluralità di attori che li esprimono (oltre agli stati e un più specifico ambiente politico dell'UE, le imprese, i movimenti sociali e il mondo degli esperti) dà vita a delle co-esistenze talvolta confliggenti, le cui tensioni spesso vengono risolte elevando alcuni immaginari al di sopra degli altri attraverso un'azione orientata da fini politici, oltreché economici. Gli immaginari socio-tecnici, dunque, sono differenziati, contestati, mercificati e "buona parte del governo della tecnologia digitale sembra essere effettuato nel farsi della tecnologia digitale e nelle sue retoriche" (Mager, Katzenbach, 2021).

Questa cornice di lavoro ha già efficacemente evidenziato le caratteristiche dei diversi processi di istituzionalizzazione dell'AI nelle società - e della società attraverso l'AI - seguendo principalmente le rappresentazioni fornite dai media, da alcuni stakeholder selezionati, o dagli attori politici - con una prevalenza di studi sugli USA e la Cina (Katzenbach, Richter, Schafer 2023). Per quanto riguarda gli stati, lo studio delle strategie nazionali di sviluppo evidenzia come Francia, Germania, USA e Cina pur concordando sull'effetto dirompente dell'AI per la società e la politica e sulla necessità di valorizzarne le potenzialità, divergono poi nei principi normativi a sostegno dell'azione, mostrando differenti visioni (Katzenbach, Bareis 2022). In questi progetti politici di sviluppo della tecnologia, sono concepiti in modo diverso sia il ruolo dei cittadini (Paltieli 2022), sia l'uso dell'AI come risorsa strategica dei governi per esercitare potere (Hoff 2023)<sup>3</sup>.

Per quanto riguarda gli stakeholder, ci sono studi che tentano di comprendere quali immaginari animano gli sviluppatori di AI e robotica in Finlandia e a Singapore (Huatala, Heino 2023). Inoltre, numerosi sono i tentativi di comprendere gli immaginari socio-tecnici dell'AI istituiti a livello mediatico in diversi contesti (Wang, Downey, Yang 2023; Brause et al. 2023), come in Cina (Zeng, Chan, Schafer 2022), in Danimarca (Hansen, 2022), e dai media anglo-americani (Chuan, 2023). Uno studio dell'Università di Bologna cerca di comprendere i livelli di consapevolezza, conoscenza e risposta emotiva nei confronti dell'AI da parte di non esperti, analizzando principalmente le paure e le speranze nelle narrazioni sull'AI come parte di più ampi immaginari socio-tecnici (Sartori, Bocca, 2023). Gli immaginari sono dunque attività di costruzione di senso interazionali, nei "contesti politici, economici, organizzazionali, situazionali e tecnologici più ampi (...) ed è possibile analizzarli con

<sup>3</sup> Più specificamente, la presenza degli immaginari dell'AI è stata indagata in diversi ambiti di applicazione: da quello sanitario (Hoff, 2023), a quello educativo (Rahm & Rahm-Skageby 2023) dalle attività di polizia (Chan 2021; Sanders & Chan 2023) alle strategie in ambito militare (Bachle e Bareis, 2022).



una metodologia che pone attenzione alle interazioni, ai discorsi e alle tecnologie nelle situazioni di interesse" (Sanders, Chan 2023: 778).

Durante le prime fasi di discussione dell'*AI Act Proposal for a Regulation of the European Parliament and of the Council of Europe Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative*, sono stati riscontrati nei discorsi di autorità pubbliche, NGOs, imprese, cittadini ed esperti, immaginari in competizione tra loro per quanto riguarda l'impatto dell'AI sulla società e la politica, e al modo in cui la legge può intervenire nei processi di innovazione e ricerca scientifica (Bakiner, 2023). Molti di questi studi rivelano la dimensione di ambivalenza costitutiva propria degli immaginari socio-tecnici dell'AI, le paure e i danni potenziali con diversi gradi di esplicitezza a fianco di visioni positive del ruolo delle applicazioni e della portata performativa dell'AI, nella presa di decisioni (Baresi e Katzenbach, 2022).

Se questa breve rassegna mostra la ricchezza e la potenzialità dell'applicazione del concetto di immaginario socio-tecnico allo studio dell'AI, si segnalano anche alcune difficoltà relative alla sua operazionalizzazione, distinzione e relazione con concetti affini come '*narratives*' o '*frames*' (Katzenbach, Richter, Schafer, 2023; Rein, Schön, 1993). J. Bareis e C. Katzenbach (2022) impiegano distintamente i concetti di narrazioni e immaginari, le prime composte da dispositivi retorici e meta-strutture argomentative da cui risultano i secondi. Gli immaginari, poiché più orientati allo sviluppo *potenziale* della tecnologia e della società e all'*allocazione di risorse* da parte di stati o portatori di interesse che vogliono ostacolare o facilitare questo potenziale, vengono ulteriormente distinti dai *frames*, nella loro chiave politica di *policy frames*, ossia credenze valori e conoscenze, che forniscono a chi formula politiche le risorse cognitive e normative per dare senso ai problemi collettivi percepiti e affiancare ad essi delle soluzioni, selezionando, organizzando e interpretando una realtà complessa e fornendo le indicazioni per esercitare persuasione e agire (Rein, Schön, 1993; Schön, Rein, 1994). Questi ultimi sono dotati di componenti comunicative e percorsi di azione, ma sono anche più "*concreti e policy-oriented*" (Katzenbach, Richter, Schafer, 2023: 211). Vedremo come questa distinzione concettuale sia utile a rilevare il modo in cui gli immaginari contribuiscono a creare i *policy frames*.

### 3. La GPAI nell'arena regolativa dell'UE: immaginari e *policy frame* del rischio sistemico

L'UE con la proposta di regolamentare la produzione e l'adozione dei sistemi di AI sembra aver segnato la fine di un'era, imponendo alle aziende produttrici di conformarsi - seppure attraverso modalità negoziate - a standard legislativi e riconoscendo di fatto il fallimento del modello autoregolativo che aveva trainato il mercato globale dell'AI (Floridi, 2021). L'iniziativa dell'UE è stata guidata dalla volontà di promuovere l'eccellenza nell'ambito dell'innovazione, ma compatibilmente con un approccio basato sul rischio, teso a salvaguardare il rispetto dei diritti umani e l'uso etico dell'IA, distinguendosi così dagli altri grandi attori globali, americani e cinesi



(Stix, 2022). D'altra parte, il potere regolativo in quanto tratto distintivo del progetto europeo di avanzare la competitività globale del mercato unico, mitigando gli effetti avversi del capitalismo, come riconosce Paul (2024), era stato identificato da B.

Jessop (2002) due decenni fa e si era poi tradotto in un approccio basato sul rischio anche per quanto riguarda la commercializzazione di altri prodotti, in quanto efficace *branding device*. Secondo T. Krarup e H. Horst (2023) le politiche europee dell'AI sono un caso di "single market-making", con le visioni relative al mercato unico come "motore e principio strutturante" in tensione attiva con la salvaguardia dei diritti a costituirne il *policy frame*. L'approccio basato sul rischio, quindi, se interpretato in chiave di *Cultural Political Economy*, ha assolto la funzione di iscrivere nella regolamentazione europea delle rappresentazioni selettive del mondo futuro dell'AI, in modo da rendere compatibili l'agenda tesa a promuovere il mercato europeo dell'AI nella competizione economica globale con la mitigazione degli effetti negativi possibili sui diritti e la sicurezza delle persone, attraverso una graduazione del rischio e delle conseguenti misure (inaccettabile, alto, basso-nullo) (Paul, 2024).

I modelli fondativi, GPAI e AI generativa, nella sua forma più comune di chatbot, nella prima versione del regolamento (Commissione UE, 2021) erano considerati usi a rischio basso o nullo. Un cambiamento di prospettiva è avvenuto nel novembre 2022 quando il Consiglio, contemporaneamente al lancio di Chat GPT da parte di Open AI e in vista del trilatero<sup>4</sup>, ha proposto di introdurre una regolazione specifica della GPAI, ma meno stringente di quella che sarà poi adottata, per evitare scontri tra le visioni differenti di vari stati (Kutterer, 2024). La posizione negoziale assunta poi dal Parlamento nel giugno 2023 disponeva infatti misure specifiche ben più rigorose poiché, sulla base di una comprensione maggiore dei rischi associati all'uso dei *foundations models*,<sup>5</sup> riconosceva il bisogno di svilupparli, ma mitigandone i "rischi prevedibili per la salute, la sicurezza, i diritti fondamentali, l'ambiente e la democrazia e lo stato di diritto"<sup>6</sup>.

Nel novembre del 2023 la dichiarazione congiunta di Italia, Francia e Germania sull'obbligo di autoregolazione dei sistemi GPAI<sup>7</sup> ha rischiato di far saltare gli accordi. Secondo la "dichiarazione della Società civile" (riportata in A. Lovelace Institute - novembre 2023) e da D. Huyskes<sup>8</sup>, questi governi si erano opposti alla regolamentazione dei *foundation models* a causa dell'azione di *lobbying* della *start-up* francese Mistral, appoggiata dall'ex segretario di Stato francese per il digitale Cedric O, e dall'azienda leader di AI tedesca Aleph Alpha, convinte, come molte imprese che "la regolazione sia in contrasto con l'innovazione". Questa posizione ha provocato



<sup>4</sup> General approach, 22, November 2022; <https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf>.

<sup>5</sup> Parlamento Europeo, 14 giugno 2023 (recital 60f, 60g; art 28b): <https://artificialintelligenceact.eu/wp-content/uploads/2023/06/AIA-%E2%80%93-IMCO-LIBE-Draft-Compromise-Amendments-14-June-2023.pdf>

<sup>6</sup> *ivi* art. 28b.

<sup>7</sup> [Germany, France, and Italy reach consensus on AI Act with mandatory self-regulation | Digital Watch Observatory](#)

<sup>8</sup> (la Repubblica, Perché il governo italiano sta dalla parte di chi non vuole regolamentare l'IA?, 24 11 2023, [https://www.repubblica.it/tecnologia/blog/dirittidigitali/2023/11/24/news/perche\\_il\\_governo\\_italiano\\_sta\\_dalla\\_parte\\_di\\_chi\\_non\\_vuole\\_regolamentare\\_lia-421130023/](https://www.repubblica.it/tecnologia/blog/dirittidigitali/2023/11/24/news/perche_il_governo_italiano_sta_dalla_parte_di_chi_non_vuole_regolamentare_lia-421130023/)).

preoccupazioni e reazioni da più parti in favore di una regolamentazione dei *foundation models* per la tutela dei diritti delle persone: nel mondo che sviluppa GPAI, fra le start-up europee, nella società civile (NGOs) e all'interno delle comunità di esperti.

Ad esempio, la European Digital SME Alliance, una rete di circa 45.000 piccole e medie imprese del settore ICT costituita nel 2013, ha espresso preoccupazione circa il rischio che la regolazione assente o debole dei prodotti di grandi fornitori di *foundation models* potesse far ricadere la responsabilità della compliance sugli utenti finali, comprese le piccole e medie imprese sviluppatrici. In questo modo si troverebbero ulteriormente svantaggiate sul mercato rispetto alle Big Tech<sup>9</sup>.

Anche AI4People, società di studio e consulenza nel campo della valutazione dei rischi dell'AI costituita nel 2018, ha espresso preoccupazioni circa i possibili effetti economici di una regolazione insufficiente della GPAI. L'assenza di misure per prevenire o sanzionare abusi o misuse, violazioni della privacy, bias, minacce alla sicurezza in settori critici come salute, trasporti e applicazione della legge può distorcere la competizione e le dinamiche di mercato, creando un campo di gioco asimmetrico: "si può realizzare più innovazione solo attraverso una migliore regolazione"<sup>10</sup>.

Algorithm Watch, una NGO internazionale costituita nel 2016 che ha per missione garantire usi sostenibili e per finalità collettive degli algoritmi, si è detta preoccupata circa i rischi della GPAI per i diritti fondamentali, l'aggravamento di discriminazioni sociali, compreso lo sfruttamento del lavoro nella catena di sviluppo dell'AI (ad esempio, il Sud globale sfruttato nell'addestramento dei modelli di OpenAI), il peggioramento della crisi climatica causato dal consumo di energia da parte dei server e dal rischio già menzionato di creazione o rafforzamento di monopoli<sup>11</sup>.

Solo pochi mesi prima, il Future of Life Institute, istituito nel 2015 "per orientare lo sviluppo delle tecnologie trasformative beneficiando la vita e lontano da rischi estremi di larga scala" aveva pubblicato una lettera aperta proponendo il fermo dello sviluppo di potenti sistemi di AI, come i GPAI. La motivazione risiedeva nei "rischi profondi per la società e l'umanità" impossibili da affrontare con decisioni delle sole imprese, senza coinvolgere i governi democraticamente eletti. I principali rischi erano relativi a: la possibilità di inondare la comunicazione di informazioni che contengono propaganda e falsità, la destabilizzazione dei mercati del lavoro, con l'automazione dei lavori più soddisfacenti; l'amplificazione di discriminazioni e *bias* intrecciati, con l'ulteriore marginalizzazione di comunità svantaggiate e punti di vista diversi; la perdita di potere delle istituzioni politiche, con la concentrazione di enormi poteri



<sup>9</sup> <https://www.digitalsme.eu/digital-smes-take-a-stand-on-the-ai-act-regulation-of-dominant-ai-models-will-spur-sme-innovation/> (3-12-2023)

<sup>10</sup> [https://www.eismd.eu/letter\\_to\\_mr\\_macron\\_mrs\\_meloni\\_mr\\_scholz\\_26-11-2\\_231128\\_160302.pdf](https://www.eismd.eu/letter_to_mr_macron_mrs_meloni_mr_scholz_26-11-2_231128_160302.pdf)

<sup>11</sup> <https://algorithmwatch.org/en/algorithmwatch-demands-regulation-of-general-purpose-ai/> (17-04-2023); <https://algorithmwatch.org/en/op-ed-generative-ai-ai-act-trilogue/> (20/11/2023)



nelle mani di poche compagnie private; l'obsolescenza umana dovuto alle capacità dell'AI, fino alla perdita del controllo delle nostre civiltà<sup>12</sup>.

Queste preoccupazioni hanno influenzato le fasi successive del processo politico che ha portato all'approvazione nella primavera del 2024 del Regolamento<sup>13</sup>, composto da norme giuridiche e da "recital" di natura discorsiva, che chiariscono le motivazioni e l'interpretazione da dare alle norme. Nell'insieme, consentono di tratteggiare il "frame" dell'intervento regolamentare per la sola componente che a noi interessa: le rappresentazioni dei rischi associati all'impiego di *foundation models* e GPAI e dei sistemi di valori e credenze ad essi associati, che confronteremo (cfr. par. 4) con quelle plasmate nella società in forma di immaginari socio-tecnici. Presentiamo di seguito questo *frame* rinunciando invece a descrivere le misure disposte<sup>14</sup>.

Nell'EU AI Act un rischio è "la combinazione della probabilità che si verifichi un danno e la gravità di tale danno" (Art. 3, 2). Più specificamente, per "rischio sistemico" della GPAI – una definizione introdotta durante il trilogò – si intende "un rischio specifico per le capacità di impatto elevato dei modelli di AI per finalità generali, avente un impatto significativo sul mercato dell'Unione a causa della sua portata o di effetti negativi *effettivi o ragionevolmente prevedibili* (corsivo nostro) sulla salute pubblica, la sicurezza, i diritti fondamentali o la società nel suo complesso, che può propagarsi su larga scala lungo l'intera catena del valore" (art. 3, 65). A questi effetti immaginati si aggiunge "qualsiasi effetto negativo *effettivo o ragionevolmente prevedibile* in relazione a incidenti gravi, interruzioni di settori critici e gravi conseguenze per la salute e la sicurezza pubblica (...), la diffusione di contenuti illegali, falsi o discriminatori" potendo dare origine a pregiudizi e discriminazioni dannosi, con rischi per gli individui, le comunità o le società. I rischi della AI con finalità generali comprendono "i diritti di proprietà intellettuale" (r. 48), richiamata anche per la "necessità di rispettare e proteggere (...) le informazioni commerciali riservate o i segreti commerciali conformemente al diritto dell'Unione e nazionali" (art. 53); i rischi per la sicurezza pubblica ed economica, l'agevolazione della disinformazione o il danneggiamento della privacy, con minacce ai valori democratici e ai diritti umani (recital 110). Nello stesso recital si precisa che è necessario "prestare attenzione ai rischi derivanti da un *potenziale uso improprio intenzionale* o da problemi involontari di controllo relativi all'allineamento con l'intento umano; rischi



<sup>12</sup> <https://futureoflife.org/open-letter/pause-giant-ai-experiments> (22-03-2023);

<https://futureoflife.org/document/policymaking-in-the-pause/> (19-04-2023).

<sup>13</sup> il testo, aggiornato al 14 maggio 2024 si può consultare al link: <https://data.consilium.europa.eu/doc/document/PE-24-2024-INIT/it/pdf>

<sup>14</sup> Le norme definiscono i criteri di identificazione e classificazione dei modelli di GPAI come modelli con rischio sistemico (artt. 1; 2; 3; 51; Annex III), le relative procedure e strumenti di trattamento da parte della Commissione UE e degli organi di consulenza (artt. 40; 41, 50, 52; 53, 56; 66; 68; 75; 88; 89; 90; 92, 93, 94, 112; Annex XI; XII), i relativi obblighi per i provider (artt. 25; 50; 53, 55) e le sanzioni per la loro inosservanza (art. 101). I recital motivano in modo più ampio i metodi per identificare (r. 97, 100, 103, 105, 133), classificare i modelli di GPAI come modelli con rischi sistemici (r. 111; 112), le misure da adottare in questi casi da parte dei produttori per gestire il rischio e le competenze e i poteri delle istituzioni e degli organi di consulenza (r. 101, 104, 106, 107, 113, 114, 115, 116, 117, 161, 162, 163, 164, 173, 179)



chimici, biologici, radiologici e nucleari, come i modi in cui possono essere abbassate le barriere all'ingresso, anche per lo sviluppo di armi, l'acquisizione di progetti o l'uso; capacità cibernetiche offensive, come i modi in cui possono essere abilitati la scoperta, lo sfruttamento o l'uso operativo della vulnerabilità; gli effetti dell'interazione e dell'uso di strumenti, compresa, ad esempio, la capacità di controllare sistemi fisici e interferire con le infrastrutture critiche (...).

Inoltre, viene fatto riferimento alle eventualità che "lo sviluppo o l'utilizzo del modello provochi un grave incidente", mentre "la protezione della cybersicurezza relativa ai rischi sistemici associati all'uso o agli attacchi dolosi deve tenere in debita considerazione la fuga accidentale di modelli, i rilasci non autorizzati, l'elusione delle misure di sicurezza e la difesa contro gli attacchi informatici, l'accesso non autorizzato o il furto di modelli" (recital 115). Inoltre, i rischi sistemici possono potenzialmente derivare dalla progettazione, dal funzionamento e dall'utilizzo di piattaforme online di grandi dimensioni e di motori di ricerca online di grandi dimensioni e che la progettazione dei sistemi algoritmici utilizzati nel servizio può contribuire a tali rischi (recital 118).

Un immaginario di allarme più radicale adombra una varietà di situazioni di perdita di controllo umano sulla tecnologia che, nel discorso pubblico, è stata più volte riferita alla di là da venire "AI generale": ne fanno parte i rischi derivanti "da (...) capacità dai modelli di creare copie di se stessi o di "autoreplicarsi" o di addestrare altri modelli (...); la fuga accidentale di modelli (...), il rischio che un particolare evento possa portare a una reazione a catena con effetti negativi considerevoli che potrebbero interessare fino a un'intera città, un'intera attività di dominio o un'intera comunità".

I rischi più specifici per la politica e la democrazia, sui quali ci concentriamo, consistono in effetti sui processi democratici (specificati ai r. 120 e 136 come "effetti negativi sui processi democratici, sul discorso civico e sui processi elettorali"), anche attraverso la disinformazione. Si aggiungono ai rischi connessi a sistemi ad alto rischio (r. 62): "sistemi destinati ad essere utilizzati per influenzare l'esito di un'elezione o di un referendum o il comportamento di voto di persone fisiche nell'esercizio del loro voto in elezioni o referendum".

#### **4. Gli immaginari del rischio della GPAI per la democrazia e i processi democratici: il dibattito scientifico e il discorso pubblico.**

Nella produzione delle scienze umane e sociali focalizzata sui rischi dell'AI generativa, così come in una parte significativa di quella che affronta più generalmente i rischi dell'AI, sono raramente presentati i risultati di ricerche empiriche sui danni prodotti e quelli evitati, o mitigati grazie a misure di prevenzione e controllo. Salvo alcune eccezioni, infatti, le ricerche sulle pratiche di uso della GPAI in campo politico sono ancora ai primi passi. Questa letteratura sviluppa piuttosto riflessioni basate su rappresentazioni riscontrate sui media tradizionali e internet, nelle prese di posizione di attori della conoscenza applicata (think tank; consulenza;



centri accademici specializzati che forniscono anche indicazioni normativamente orientate), dell'ambiente economico e della società civile no-profit.

Anche per questo motivo, le rappresentazioni contenute nei contributi che abbiamo analizzato (frutto di una ricerca bibliografica per parole chiave attraverso Scopus) possono essere considerate una proxy valida del discorso pubblico sui rischi dell'AI e in particolare di GPAI). Esse fanno emergere i modi in cui l'ambiguità costitutiva degli immaginari socio-tecnici dell'AI, in cui nella prospezione dei futuri possibili le paure e i danni potenziali affiancano le visioni positive, ha fornito "materiali" per la traslazione in *policy frame* nel corso del processo politico dell'UE.

Già in una presa di posizione di importanti e numerose associazioni della società civile durante il trologo si affermava che "i *foundation models* presentano rischi significativi data la loro complessità, scala e pervasività, ed in particolare per la loro potenzialità intrinseca di formare un'infrastruttura centrale per le applicazioni successive" (Ada Lovelace Institute - *Dichiarazione Società civile sull'AI Act dell'UE - novembre 2023*). Essi sono dunque pericolosi a causa della potenziale pervasività e degli effetti finali suscettibili di sfuggire al controllo non solo dei regolatori, ma anche di fornitori e *deployer* (Shevlane et al., 2023).

Questi danni generalizzati riguardano anche la democrazia. Per quanto riguarda quest'ultima, negli immaginari i rischi possono essere:

- da un lato *indiretti*, ossia conseguenze delle trasformazioni sociali indotte dall'AI, che però non si esercitano modificando processi e pratiche democratiche. Vi rientrano, ad esempio nel campo del lavoro, la sostituzione di lavoro umano, l'uso di AI nella gestione di tutto il ciclo della forza-lavoro, dalla ricerca di personale, all'assunzione al controllo delle prestazioni, ai licenziamenti. Ne può derivare l'ulteriore indebolimento delle già compromesse capacità di tutela, rappresentanza e azione collettiva del lavoro che sono state componenti delle democrazie pluraliste, mentre il loro declino contribuisce alla eclissi dei partiti di massa, al disimpegno dall'azione collettiva e dalla partecipazione politica (Raniolo, 2024). I cambiamenti nei rapporti sociali nel mercato del lavoro contribuiscono inoltre ad accrescere le disuguaglianze sociali, suscitando tra varie fasce della popolazione una crescente percezione di privazione relativa, ingiustizia, esclusione, ridotte aspettative di mobilità sociale, accresciute sensazioni di minacce derivanti dalle migrazioni di massa. Questo alimenta risposte politiche populiste e sovraniste, che delegittimano la separazione dei poteri e pretendono autorità sottratta a controlli e contrappesi, fino alle "democrazie illiberali" con uso autoritario dell'autorità statale. Con queste modalità, quindi, un importante uso dell'AI nel mercato del lavoro, anche con applicazioni derivate da GPAI, può minacciare indirettamente la democrazia.

Altre importanti conseguenze indirette dell'AI per la democrazia riguardano i rapporti fra politica ed economia, con l'ulteriore crescita del potere delle Big Tech, con le quali la politica è già impegnata nel mondo in vari contenziosi, a partire da quello fiscale. L'AI, e segnatamente la GPAI, può approfondire i processi della "post-democrazia" (Crouch, 2004; 2020), favorendo ulteriore spostamento di poteri e capacità di innovazione dall'autorità pubblica ad attori e arene di mercato o miste. Sul versante delle risorse cognitive usate nelle relazioni di potere, la *societal*



*depoliticisation* (Wood, Flinders, 2014) scivolerebbe in quella che potremmo definire una *technological depoliticization*, in cui la compromissione della *epistemic agency* (cfr. oltre) potrebbe riguardare, più che gli individui, la capacità delle istituzioni politiche (come agenti umani collettivi capaci di formare e rivedere le basi cognitive delle scelte politiche) se l'"algorithmic output" coincidente con una "verità" incontestabile, sostituisse gli esiti di processi politici discorsivi e pluralisti.

- da un altro lato *diretti*, quando riguardano i processi e le pratiche della democrazia liberale, rappresentativa e partecipativa, attraverso una molteplicità di usi, in campagne elettorali, consultazioni pubbliche, pratiche di democrazia partecipativa e deliberativa, *lobbying*, supporto alle decisioni pubbliche e all'azione amministrativa. Per molte di queste utilizzazioni, alle rappresentazioni delle potenzialità di uso dell'AI corrispondono immaginari socio-tecnici del rischio per la democrazia. Possiamo distinguerli sulla base della presenza o meno di intenzionalità malevola (*misuse*). Nel primo caso, l'AI (in particolare, ma non solo, i modelli della GPAI) mettono a disposizione di attori sia individuali, sia organizzati, strumenti per perseguire finalità non solo incompatibili con i valori e le pratiche della democrazia ma anche, in alcuni casi, già illecite. Nel secondo caso, l'immaginario dà voce a una preoccupazione anche maggiore, perché il pericolo è insito nella tecnologia stessa. Se priva, o capace di sottrarsi al controllo umano, che è possibile a sua volta immaginare attuato da attori diversi e con finalità e strumenti differenziati, l'AI è di per sé in grado di produrre danni, in questo caso alla democrazia e alle sue pratiche.



### **- Immaginari relativi a usi intenzionalmente malevoli**

Il *misuse* di AI nei rapporti fra politica e cittadini è il fenomeno più studiato, al quale corrisponde un sottosistema di immaginari in cui dei sistemi di AI, potenziati dalla GPAI, consentono di perseguire scopi di disinformazione intenzionale e manipolare le opinioni e i comportamenti, in particolare elettorali. L'AI può creare uno strato invisibile di influenza tra il cittadino e i suoi rappresentanti politici (Duberry, 2022, 94), i cui effetti antidemocratici - a partire dalla disinformazione intenzionale attraverso *political bots* automatizzati e *micro-targeting* attraverso profilazione e *clusters* su più variabili - sono amplificati dalla centralizzazione del potere sulla generazione e moderazione dei contenuti in un piccolo numero di grandi piattaforme. Gli utenti possono non essere in grado di distinguere tra i contenuti generati dall'uomo e quelli generati dall'IA, specialmente le *deep fakes* (EPRS, 2023), notizie false diffuse in modo automatizzato su social media, di cui non si può rintracciare la fonte e che sono particolarmente efficaci sotto forma di video e audio falsi. Esistono siti internet con notizie solo false, disseminate poi da Bot che usano AI. Questi strumenti possono essere usati sia da individui, sia da *cyber-troops*, come nei casi documentati di campagne di influenza straniera. Tendono a creare e alimentare *filter bubbles*, *echo chambers*, *confirmation bias* (Bullock et al. 2022; König, Wenzelburger, 2020; Coeckelberg, 2022; Duberry, 2022; Avalle et al. 2023), in sintesi ambienti comunicativi in cui gli individui ricevono in modo reiterato informazioni orientate a rafforzare i loro preesistenti punti di vista, a causa dei quali sono divenuti

target dei messaggi. Questo da un lato alimenta e riproduce la polarizzazione delle rappresentazioni e delle scelte politiche. Ricevendo aggiornamenti costanti che riguardano i temi preferiti (ciò a cui i destinatari rispondono) (Viehoff, 2022), gli elettori possono diventare target specifici di *hypernudging*, una comunicazione che dà la "spinta gentile" ad esempio verso un partito, o candidato, come nel caso di Cambridge analitica in occasione delle elezioni presidenziali USA del 2016 (Coeckelberg, 2022). D'altro canto, proprio a causa della natura di circuiti comunicativi chiusi, fra le conseguenze non sarebbe da annoverare il ri-orientamento di chi ha posizioni diverse (Avalle et al. 2024).

### **- Immaginari relativi a usi non intenzionali, non malevoli**

L'immaginario socio-tecnico probabilmente più preoccupante per l'agency umana, non solo con riferimento alla democrazia, riguarda l'impossibilità di modelli privi di potenziali errori di sistema, poiché mancheranno sempre di preferenze non rivelate in precedenza, di conoscenze tacite e di importanti giudizi e valori emergenti che sono prodotti attraverso interazioni soggettive - esperienze cognitive umane con il mondo esterno guidate da emozioni, intuizioni e immaginazione (Wihbey, 2024). Una sua specificazione riguarda il cosiddetto "rischio epistemico". Questo consiste nell'erosione della base cognitiva della democrazia e la "*epistemic agency*" degli individui, ossia la loro capacità di formare e rivedere le proprie credenze politiche (Coeckelbergh, 2022). L'era dell'AI rischia di diventare per le democrazie un gigantesco esercizio ricorsivo e i meccanismi di deliberazione democratica possono finire per incanalare delle preferenze già modellate dalla produzione di conoscenza pubblica guidata dall'AI con le modalità che abbiamo visto. Inoltre, l'AI potrebbe operare senza tenere adeguatamente conto dei valori e delle preferenze umane: le "*organic views*," dei cittadini (ossia quelle non sostanzialmente plasmate e influenzate dalle macchine), possono addirittura finire per non riflettersi nelle decisioni democratiche. La conseguente riduzione dell'autonomia umana può essere dannosa per la democrazia, che richiede deliberazione e scelta autonoma. In questa chiave, in particolare, le scorciatoie generative dell'AI generativa (non solo Chat GPT, ma anche modelli di risposta di Google), cortocircuitando i processi abituali di navigazione, scoperta, deliberazione e ragionamento, possono rendere gli esseri umani sempre più dipendenti dalle capacità di selezione e sintesi dei modelli di AI (Wihbey, 2024). All'immaginario di un rischio epistemico individuale è da ricondurre anche la rappresentazione di sistemi AI usati per agire in modo automatico come una "mano invisibile" cui affidare scelte, qualora un risultato algoritmico non sia visto come arbitrario, ma come un risultato che riflette, almeno in una certa misura, una "verità (...) che non può essere contestata o messa in discussione" (Paltieli, 2023).

Un terzo tipo di immaginario riguarda la difficoltà di capire e controllare i sistemi di AI, le loro utilizzazioni e le conseguenze. Queste rappresentazioni spaziano da una opacità che potrebbe anche essere corretta attraverso degli accorgimenti tecnici volontari o imposti dall'esterno (con la legge), alla completa impossibilità di controllare modelli e sistemi, che può sopravvenire anche nel corso del loro uso a



causa della capacità autogenerativa di alcuni modelli di GPAI. Ad esempio, alle potenzialità dell'uso di AI nell'input e nel funzionamento dei processi decisionali pubblici può fare da contrappeso il rischio epistemico della standardizzazione dei processi cognitivi. A questo rischio si accompagnano immaginari relativi a deficit di *accountability* dei sistemi di AI e di consapevolezza critica degli utenti circa il design e la gestione di piattaforme e algoritmi. È il rischio di opacità, che riguarda anche gli usi di AI nelle pratiche di democrazia partecipativa e deliberativa, qualora i partecipanti non abbiano consapevolezza critica sul funzionamento della tecnologia utilizzata, sugli attori che la sviluppano e gestiscono, sulla trasparenza e la responsabilità del trattamento dei dati e sulla *cybersecurity* (Duberry, 2022). Anche quando piattaforme e algoritmi di AI, specialmente generativa, sono usati per il supporto alle decisioni pubbliche, la loro opacità può impedire a cittadini e *decision-makers* istituzionali di comprendere come siano processati i dati e come siano prodotte previsioni, raccomandazioni o decisioni algoritmiche (Arsenault, Kreps 2022).

Un quarto tipo di immaginario riguarda i rischi di *bias* e discriminazione. I modelli di AI si basano su dati e modelli. Se questi riguardano il passato, le previsioni e gli stessi suggerimenti agli utenti dei sistemi di AI rischiano di replicare le condizioni esistenti, poiché si basano sui comportamenti dei *policy taker* del passato, o anche di riprodurre modelli di discriminazione sociale, legale e politica nonché di disuguaglianza ed esclusione esistenti (Viehoff, 2022), contravvenendo un loro buon uso ai fini di *policy* per ridurle (Jungherr, 2023). Differenti gradi di esposizione e di visibilità ai sistemi AI possono accrescere l'influenza sui processi democratici di alcuni gruppi e ridurre quella di altri. L'AI può usare le presenze dei gruppi più visibili nelle previsioni sui trend politici e sull'impatto delle *policy* ignorando quelle dei gruppi meno visibili e delle voci di minoranza (Jungherr, 2023). Il 2024 AI index report (Stanford Institute for Human-Centered AI) ha segnalato la presenza di specifici *bias* politici dell'AI generativa, poiché dalle ricerche risulta che ChatGPT è politicamente parziale, con una significativa polarizzazione verso le posizioni e i linguaggi dei democratici negli USA e del Partito Laburista in UK.

Il problema della trasparenza e responsabilità del trattamento dei dati corrisponde a un immaginario di rischio per le persone preesistente all'AI e alla GPAI, ma che si riproduce e si differenzia. Nell'Unione europea, l'esistenza di norme e sistemi di implementazione (GDPR; norme sui servizi digitali, intermediari e piattaforme online) è il frutto e a sua volta contribuisce ad alimentare una sensibilità diffusa sui rischi tecnologici per la privacy e i diritti delle persone.

Un ultimo immaginario dipinge gli attori pubblici come potenzialmente dipendenti dalle imprese produttrici e fornitrici di AI che, anche in modo non intenzionalmente finalizzato, possono orientare politicamente le decisioni attraverso la proposizione o riproposizione di *frame* relativi alle politiche e alla loro implementazione, influenzando in modo negativo la capacità delle istituzioni di decidere autonomamente. L'assenza o l'inefficacia di regolazione, in particolare dell'AI regolativa e il conseguente affidamento al mercato della capacità di generare o co-generare con gli attori pubblici, diffondere e legittimare *frame* e strumenti per politiche settoriali, fa crescere la preoccupazione che "saranno le grandi aziende a



plasmare le nostre società piuttosto che le nostre istituzioni democratiche” (Buec, 2023).

## 5. Conclusioni

Le evidenze raccolte hanno consentito di ricostruire il modo in cui gli immaginari socio-tecnici del rischio, prefigurando i pericoli per la democrazia derivanti dagli usi dell'AI generativa, hanno contribuito a strutturare il *policy frame* dell'EU AI ACT. Ciò è avvenuto attraverso un processo di costruzione di senso interno ed esterno all'arena legislativa in senso stretto (le istituzioni dell'UE) concentrato soprattutto sul modo in cui i *foundation models* e le loro svariate e anche solo potenziali applicazioni possono costituire una minaccia per l'efficacia e la legittimazione dei poteri democratici, oltre che per specifiche pratiche democratiche.

Come abbiamo visto (par. 3) nel corso del processo politico, è stato possibile costruire e condividere uno specifico *policy frame* della GPAI, attraverso l'elaborazione di una categoria, quella di *rischio sistemico*. Questa nuova definizione del problema ha permesso di incorporare nelle norme delle soluzioni, corredate da motivazioni discorsive, che hanno sintetizzato i contenuti di un compromesso in buona parte provvisorio fra gli interessi materiali e politici in causa e le preoccupazioni per la democrazia. Le risorse cognitive e normative necessarie per dare senso ai problemi e affiancare ad essi delle soluzioni, selezionando, organizzando e interpretando una realtà complessa, sono state fornite in questo caso da uno specifico sistema di “immaginari”, con il significato e il potenziale interpretativo che abbiamo qui utilizzato. Gli immaginari condensati nel “rischio sistemico”, se confrontati con quelli relativi ai sistemi ad alto e basso rischio, esprimono scenari di incertezza per il futuro dei rapporti fra politica e società e delle pratiche democratiche, che hanno costituito il quadro di senso fondamentale in questo episodio di politica dell'AI.

Tuttavia, non si è trattato solo di un processo di costruzione di significato: il negoziato fra Consiglio, stati membri e Parlamento nel periodo 2022-2024 evidenzia come nelle applicazioni AI la struttura regolativa e gli interessi materiali siano in continua tensione con l'articolazione semiotica del loro uso e dei loro rischi (Paul, 2024). In altri termini, gli immaginari hanno contribuito a legittimare il corso dell'azione politica riducendo una complessità non solo immateriale, ma costituita anche di poste in gioco materiali – presumibilmente il principale mercato europeo e globale nell'immediato futuro – di enorme rilevanza.

Quindi, l'immaginario contribuisce attraverso il ruolo fondamentale, a un tempo cognitivo e normativo, svolto in processi istituenti di senso condiviso e pratiche sociali - nel nostro caso le caratteristiche, gli usi e le conseguenze dell'AI e quindi la produzione di *policy frame* - a strutturare sia i sottosistemi della società (economico, sociale, culturale, tecno-scientifico, accademico, della società civile, no-profit politico), sia le relazioni tra loro.



L'arena resta comunque aperta, con l'AI Pact, l'attuazione dell'AI Act e con il ruolo chiave che svolgeranno la Commissione e l'AI Board, gli altri organismi istituiti dal regolamento e gli stati membri nella classificazione dei modelli e nell'applicazione del regolamento in una varietà di processi complessi e potenzialmente conflittuali. Ai nuovi round negoziali e decisionali prenderanno parte burocrazie amministrative, imprese, esperti e società civile. È ipotizzabile che si tratterà di nuovo di confronti, di mediazioni su cui risulteranno influenti sia poste in gioco materiali (economiche e funzionali) sia immateriali (valori, diritti), in uno scenario di complessità e incertezza.



## Bibliography

- Adams, S., Blokker, P., Doyle, N. J., Krummel, J. W., & Smith, J. C. (2015). "Social imaginaries in debate". *Social Imaginaries*, 1(1), 15-52.
- AI4 People (2023) *Letter to mr. Macron, mr. Schulz, ms. Meloni* [Ref. BB-LF 54-2023](#); ultimo accesso il 31 maggio 2024.
- Algorithm Watch, (2023) [Generative AI must be neither the stowaway nor the gravedigger of the AI Act - AlgorithmWatch](#) ultimo accesso il 31 maggio 2024.
- Arnason, P. J., (2014) "Social Imaginary Significations", in Adams, S., (a cura di) *Cornelius Castoriadis: Key concepts*, Bloomsbury, London, pp. 23-42.
- Arsenault A. C., Kreps S. E. (2022), "AI and International Politics", in Bullock J. B. et al., (a cura di), *The Oxford Handbook of AI Governance* (online edn, Oxford Academic, 14 Feb).
- Avalle, M., Di Marco, N., Etta, G. et al. (2024) "Persistent interaction patterns across social media platforms and over time". *Nature* 628, 582–589.
- Bareis, J., & Katzenbach, C. (2022). "Talking AI into Being: The Narratives and Imaginaries of National AI Strategies and Their Performative Politics". *Science, Technology, & Human Values*, 47(5), 855-881. <https://doi.org/10.1177/01622439211030007>.
- Bächle, T.C., Bareis, J. (2022) "Autonomous weapons as a geopolitical signifier in a national power play: analysing AI imaginaries in Chinese and US military policies". *Eur J Futures Res* 10, 20 <https://doi.org/10.1186/s40309-022-00202-w>.
- Bakiner, O. (2023). "Pluralistic sociotechnical imaginaries in Artificial Intelligence (AI) law: the case of the European Union's AI Act". *Law, Innovation and Technology*, 15(2), 558-582.
- Brause, S. R., Zeng, J., Schäfer, M. S., & Katzenbach, C. (2023) "Media representations of artificial intelligence: surveying the field". In Lindgren, S. (a cura di) *Handbook of Critical Studies of Artificial Intelligence*. Cheltenham, UK: Edward Elgar Publishing. <https://doi.org/10.4337/9781803928562.00030>.
- Browne, C., & Diehl, P. (2019). "Conceptualising the political imaginary: An introduction to the special issue", in *Social Epistemology*, 33(5), 393-397.
- Bullock, J. B. et al. (2022) "Introduction", in Bullock J. B. et al., (a cura di.), *The Oxford Handbook of AI Governance* (online edition, Oxford Academic, 14 Feb).



Buec (2023) Eu risks under regulating generative AI, European Consumer Organisation [https://www.beuc.eu/sites/default/files/publications/BEUC-PR-2023-025\\_Consumer\\_groups\\_call\\_on\\_regulators\\_to\\_investigate\\_generative\\_AI\\_risks\\_%26\\_enforce\\_existing\\_legislation.pdf](https://www.beuc.eu/sites/default/files/publications/BEUC-PR-2023-025_Consumer_groups_call_on_regulators_to_investigate_generative_AI_risks_%26_enforce_existing_legislation.pdf) ultimo accesso il 31 maggio.

Castoriadis, C. 1987 [1975], *The Imaginary Institution of Society*, Blamey, K. (traduzione), MIT Press, Cambridge, MA.

Castoriadis, C., (2007) *Figures of the Thinkable*, H. Arnold (traduzione), Stanford, CA: Stanford University Press.

Chan J, 2021, "The Future of AI in Policing: Exploring the sociotechnical imaginaries", in McDaniel J; Pease K (a cura di), *Predictive Policing and Artificial Intelligence*, Routledge, Milton Park, Abingdon, Oxon.

Chuan, C. (2023). "A critical review of news framing of artificial intelligence". In Lindgren S, (a cura di), *Handbook of Critical Studies of Artificial Intelligence*. Cheltenham, UK: Edward Elgar Publishing. <https://doi.org/10.4337/9781803928562.00029>.

Coeckelberg M. (2022), *The political philosophy of AI*, Polity.

Coeckelbergh M. (2023), "Democracy, epistemic agency, and AI: political epistemology in times of artificial intelligence", *AI Ethics*, 3:1341–1350.

Council of Europe, (2022) *General approach*, 22 November; <https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf>.

Crouch C. (2004), *Post-Democracy*, Polity Press, Cambridge.

Crouch C. (2020), *Post-Democracy after the Crises*, Polity Press, Cambridge.

Digital Watch Observatory, (2023) [Germany, France, and Italy reach consensus on AI Act with mandatory self-regulation | Digital Watch Observatory](#), ultimo accesso il 31 maggio 2024.

Duberry J. (2022), *Artificial Intelligence and Democracy. Risks and Promises of AI-Mediated Citizen–Government Relations*, Cheltenham, Edward Elgar.

EPRS | European Parliamentary Research Service (2023), Artificial intelligence, democracy and elections.



Riccardo Corsi, Ernesto d'Albergo  
*La politica dell'intelligenza artificiale general purpose*

European SME Alliance, (2023) [Tiered based regulation of AI foundation models to support SME innovation - European DIGITAL SME Alliance](#), ultimo accesso il 31 maggio 2024.

European Parliament, *Artificial Intelligence Act* (2023) 14 giugno : [Artificial Intelligence Act](#), ultimo accesso il 31 maggio 2024.

European Parliament and Council of Europe (2024) *REGOLAMENTO DEL PARLAMENTO EUROPEO E DEL CONSIGLIO che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (regolamento sull'intelligenza artificiale)* <https://data.consilium.europa.eu/doc/document/PE-24-2024-INIT/it/pdf>

Floridi, L. (2021). "The end of an era: from self-regulation to hard law for the digital industry". *Philosophy and Technology*, 619-622.

Future of Life Institute, (2022) [General Purpose AI and the AI Act](#), ultimo accesso al 31 maggio 2024.

Future of Life Institute, (2023) [Pause Giant AI Experiments: An Open Letter - Future of Life Institute](#), ultimo accesso al 31 maggio 2024.

Hansen, S. (2022). "Public AI imaginaries: How the debate on artificial intelligence was covered in Danish newspapers and magazines 1956–2021". *Nordicom Review*, 43(1), 56-78.

Hautala, J., & Heino, H. (2023). "Spectrum of AI futures imaginaries by AI practitioners in Finland and Singapore: The unimagined speed of AI progress". *Futures*, 153, 103247.

Hoff, L-J. (2023) "Unavoidable futures? How governments articulate sociotechnical imaginaries of AI and healthcare services", *Futures*, Volume 148.

Jessop, B. (2002). *The future of the capitalist state*. Polity.

Jungherr, A. (2023). "Artificial Intelligence and Democracy: A Conceptual Framework". *Social Media + Society*, 9 (3) (<https://journals.sagepub.com/doi/10.1177/20563051231186353>).

Katzenbach C.; Richter V.; Schäfer M.S. (2023) "Imaginaries of Artificial Intelligence", in Lindgren, S. (a cura di) *Handbook of Critical Studies on Artificial Intelligence*, Edward Elgar Publishing.



König P.D., Wenzelburger, G., (2020), "Opportunity for renewal or disruptive force? How artificial intelligence alters democratic politics", *Government Information Quarterly*, Volume 37, Issue 3.

Krarup, T., & Horst, M. (2023). "European artificial intelligence policy as digital single market making". *Big Data & Society*,10(1), 1–14.

La Repubblica, 24 11 (2023) [Perché il governo italiano sta dalla parte di chi non vuole regolamentare l'IA? - la Repubblica](#) , ultimo accesso al 31 maggio 2024.

Mager A., Katzenbach, C., (2021) "Future imaginaries in the making and governing of digital technology: Multiple, contested, commodified, Introduction to the special issue, *New Media & Society*, 23 (2), pp. 223-236, <https://doi.org/10.1177/1461444820929321>.

Jasanoff, S. (Ed.). (2004). *States of Knowledge: The Co-Production of Science and the Social Order*, Routledge. <https://doi.org/10.4324/9780203413845>.

Jasanoff, S., (2015) "One Future Imperfect: Science, Technology and the Imaginations of Modernity", in Jasanoff, S., e Kim, S. H., (a cura di), *Dreamscapes of Modernity: Sociotechnical Imaginaries and the fabrication of Power*, University of Chicago Press, consultato in open access online il 27 maggio 2024 al link: [Imaginaries – P. 1 Future Imperfect: Science, Technology, and the Imaginations of Modernity Sheila Jasanoff Techn](#)

Jasanoff, S., & Kim, S.-H. (2009). "Containing the Atom: Sociotechnical Imaginaries and Nuclear Power in the United States and South Korea". *Minerva*, 47(2), 119–146.

Jasanoff, Sheila & Kim, Sang-Hyun. (2013). "Sociotechnical Imaginaries and National Energy Policies". *Science as Culture*. 22. 189-196. 10.1080/09505431.2013.786990.

Jasanoff, Sheila & Kim, Sang-Hyun. (2015). *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*, University of Chicago Press, Chicago. 10.7208/chicago/9780226276663.001.0001.

Paltieli, G. (2022). "The political imaginary of National AI Strategies". *AI and Society*. 37 (4):1613-1624.

Paltieli G. (2023), "Re-imagining democracy: AI's challenge to political theory" in Lindgren, S. (a cura di), *Handbook of Critical Studies of Artificial Intelligence*, Edward Elgar.



Paul, R. (2024), "European artificial intelligence 'trusted throughout the world': Risk-based regulation and the fashioning of a competitive common AI market". *Regulation & Governance*. <https://doi.org/10.1111/rego.12563>.

Rahm, L., & Rahm-Skågeby, J. (2023). "Imaginaries and problematizations: A heuristic lens in the age of artificial intelligence in education". *British Journal of Educational Technology*, 54, 1147–1159. <https://doi.org/10.1111/bjet.13319>.

Raniolo F., (2024), *La partecipazione politica*, Il mulino, Bologna

Rein M., Schön D. (1993) "Reframing Policy Discourse", in F. Fischer e J. Forester J. (a cura di), *The argumentative Turn in Policy Analysis and Planning*. Duhram and London: Duke University Press, 145-166.

Sanders CB; Chan J, (2023), "Methodological reflections on researching the sociotechnical imaginaries of AI in policing", in Lindgren, S., (a cura di) *Handbook of Critical Studies of Artificial Intelligence*, pp. 773 – 782.

Sartori, L., Bocca, G. (2023) "Minding the gap(s): public perceptions of AI and socio-technical imaginaries". *AI & Society* 38, 443–458 (2023). <https://doi.org/10.1007/s00146-022-01422-1>.

Shevlane, T., Farquhar, S. Garfinkel, B., Phuong, M., Whittlestone, J., Leung, L., Kokotajlo, D., Marchal, N., Anderljung, M., Kolt, N., Ho, L., Siddarth, D., Avin, Hawkins, S. W., Kim, B., Gabriel, J., Bolina, V., Clark, J., Bengio, J., Christiano, J. and Dafoe, A., (2023) *Model evaluation for extreme risks*, DeepMind, consultato 29 maggio 2024 al <https://arxiv.org/pdf/2305.15324>

Schön, D., & Rein, M. (1994). *Frame reflection: Resolving intractable policy issues*. New York: Basic Books.

Stix, C. (2022) "The ghost of AI Governance. Past, Present and Future: AI Governance in the European Union. In J. Bullock et al., (a cura di) *The Oxford Handbook of Ai Governance*. New York: Oxford University Press (online edn, Oxford Academic, 14 Feb).

Sum, N.-L., & Jessop, B. (2013). *Towards a cultural political economy. Putting culture in its place in political economy*. Edward Elgar.

Taylor, C. (2004). *Modern Social Imaginaries*, Duke University Press. <https://doi.org/10.2307/j.ctv11hpgvt>.



Riccardo Corsi, Ernesto d'Albergo  
*La politica dell'intelligenza artificiale general purpose*

Viehoff, J., (2022) "Beyond Justice: Artificial Intelligence and the Value of Community", in Justin B. Bullock, et al., (a cura di), *The Oxford Handbook of AI Governance* (online edn, Oxford Academic, 14 Feb. 2022).

Wang, W., Downey, J., & Yang, F. (2023). "AI anxiety? Comparing the sociotechnical imaginaries of artificial intelligence in UK, Chinese and Indian newspapers." *Global Media and China*, 0(0). <https://doi.org/10.1177/20594364231196547>.

Wihbey J.P., (2024), "AI and Epistemic Risk for Democracy: A Coming Crisis of Public Knowledge?", *Conference on Democracy's Mega Challenges: How Climate Change, Migration, and Big Data Threaten the Future of Liberal Democratic Governance*, Trinity College, Hartford, CT, April 19-20

Wood M., Flinders M. (2014), "Rethinking Depoliticisation: Beyond the Governmental", *Policy & Politics*, 42, 2, pp. 151-70

Zeng, J., Chung-Hong, C., Schäfer, M. S., (2022). "Contested Chinese dreams of AI? Public discourse about artificial intelligence on WeChat and people's daily online". *Information, Communication and Society*, 25(3):319-340.





A JOURNAL OF THE  
SOCIAL IMAGINARY

# AI Imaginaries and Narratives in the Italian Public Discourse: The Impact of ChatGPT

Stefano Spillare  
stefano.spillare2@unibo.it  
*University of Bologna*

Michele Bonazzi  
michele.bonazzi@unife.it  
*University of Ferrara*

Piergiorgio Degli Esposti  
pg.degliestposti@unibo.it  
*University of Bologna*



## Abstract

Recent advancements in artificial intelligence (AI) have sparked renewed interest and debate, particularly with the introduction of innovative "generative" AI systems such as OpenAI's ChatGPT. This innovation has reignited discussions about AI's implications in today's digital societies, often polarizing opinions between optimistic views and tragic risks, frequently fueled by sci-fi or sensationalist narratives. This study aims to investigate the imaginaries and narratives of AI in leading Italian newspapers, specifically evaluating the impact of ChatGPT on public discourse.

## Keywords

Artificial intelligence | ChatGPT | Imaginaries | Narratives | Media coverage



The key to artificial intelligence has always  
been the representation.

Jeff Hawkins (2023)

**S**ince the legendary Dartmouth workshop in 1956, the development and study of AI have experienced several “springs”, systematically followed by cold “winters” (Mitchell, 2019). Recently, we seem to have witnessed a renewed “spring” of AI development, driven by advancements in “neural networks” and propelled by increasing computational and memory capabilities, alongside the emergence of the Internet and the vast availability of data (Mitchell 2019; Cristianini 2023). Thanks to the trainability of neural networks for specific objectives, AI algorithms can perform defined tasks, such as image recognition, natural language processing, and pattern analysis. This training method is known as “machine learning” (or “deep learning”<sup>1</sup>), and it allows for the exploitation of AI's potential across various domains, including communication and human-like relational activities.

One of the most evident manifestations of these capabilities was the launch of ChatGPT on November 30, 2022—an unprecedented chatbot based on the so-called “Large Language Model” (LLM)—to the mass public. In simple terms, an LLM is a machine learning system applied to billions of written texts, enabling the software to statistically predict the most suitable sequence of words based on a query known as a “prompt” (Floridi 2019; Cristianini 2024). The result is a tool capable of “naturally” interacting with humans in unexpected ways.

These types of AI tools are often referred to as “generative” because they can “generate” something new, such as texts, voices, music, images, or videos, in a relatively “creative” manner, starting from the raw material on which they were trained. ChatGPT is not the first of these systems; similar tools include digital assistants like Apple’s Siri, Amazon’s Alexa, and Google’s suggestion systems. However, none of these previous systems could interact with humans while “keeping in mind” the entire sequence of dialogue, responding in a pertinent and accurate manner (Cristianini 2024). In a certain sense, ChatGPT seems to have surpassed the limits defined by Alan Turing in his “imitation game”, fueling increasing speculation about how far the horizon of “general AI” could extend (Cristianini 2024).

In summary, it represents what is often referred to in discussions of innovation as a “disruptive innovation”—an innovation that promises to trigger a renewed technological surge, boost competitiveness, and potentially create yet another financial bubble, akin to the advent of the Internet or social networking sites.

Despite these premises (and promises) regarding ChatGPT, the manner of its launch is significant. ChatGPT was developed by OpenAI, one of Silicon Valley's most promising “unicorns” in the AI field, initially a non-profit organization, with the

<sup>1</sup> The adjective “deep” in machine learning methods refers to the number of layers in a neural network (Mitchell 2019).



declared purpose of disseminating AI innovation to the public to familiarize people with this technology and stimulate public debate around it.<sup>2</sup> Five days after ChatGPT's official release, over one million people had subscribed to the site.<sup>3</sup>

The web exploded with comments, examples of questions posed to and answers provided by ChatGPT, and thousands of newspaper articles discussing this new tool, as well as the potentialities and risks of AI development and application in the near future. As early as 2020, anticipating the launch of OpenAI's chatbot, *The Guardian* provocatively titled an article "A robot wrote this entire article. Are you scared yet, human?"<sup>4</sup> In December 2022, it relaunched with the explicit headline: "What is the AI chatbot phenomenon ChatGPT and could it replace humans?"<sup>5</sup> Similar headlines appeared in every newspaper, often emphasizing excessively utopian perspectives or, conversely, describing apocalyptic and dystopian scenarios (Ouchchy et al. 2020; Cools et al. 2022; Bartolomew, Mehta 2023), frequently drawing upon AI imaginaries already rooted in popular culture (Cave, Dihal 2019).

The examination of these narratives holds particular significance not only for the public's perception of this innovation but also because this perception largely influences—in sociotechnical terms—further developments, applications, policies, and regulations of AI (Cave et al. 2018; Cave, Dihal 2019; Lindgren, Holmstrom 2020; Sartori, Bocca 2022). This research thus aims to contribute to the analysis of AI-related sociotechnical imaginaries and narratives within the context of the Italian news landscape, particularly focusing on the impact of ChatGPT.

The hypothesis is that this tool could be viewed as a watershed in the debate about AI and an accelerator of the social dynamics associated with it.

To achieve this, after a brief introduction to the concept of AI, its opportunities and risks, as well as the relevance of related imaginaries and narratives, the main results of a content analysis of selected Italian news articles—before and after the turning point represented by ChatGPT—will be presented.

## 1. The rise and risks of AI and the relevance of related imaginaries and narratives

Despite the widespread use of the term, the meaning of AI has not yet been definitively clarified, and its impact on society remains largely to be investigated. Drawing from various dictionaries, Rezaev and Tregubova (2018) highlighted how the

---

<sup>2</sup> Information on the company's mission and vision can be found at the following link: <https://openai.com/about/>.

However, it is likely that commercial and technical factors have also contributed to this choice, not at least the opportunity of exploiting users to improve its performance.

<sup>3</sup>As of December 4, 2023, ChatGPT had reached 1.7 billion users: <https://tg24.sky.it/tecnologia/2023/12/04/chat-gpt-utenti-storia#07> (16/5/2024).

<sup>4</sup>See [theguardian.com](https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3) at the following weblink: <https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3> (8/5/2024).

<sup>5</sup> See [theguardian.com](https://www.theguardian.com/technology/2022/dec/05/what-is-ai-chatbot-phenomenon-chatgpt-and-could-it-replace-humans) at the following weblink: <https://www.theguardian.com/technology/2022/dec/05/what-is-ai-chatbot-phenomenon-chatgpt-and-could-it-replace-humans> (8/5/2024).



concept of AI is essentially treated in two different ways: on the one hand, it is understood in terms of research on specific types of performance of computational machines; on the other hand, it is considered a performance in itself. In the first case, Russell and Norvig (1995, cited in Jungherr, Schroeder 2023: 1) defined AI as “the study and construction of agents that do the right thing”. In the second case, it can be understood as “the ability of machines to pursue specific tasks of varying difficulty in appropriate ways or, more broadly, the ability to set goals autonomously, reason, and adapt to unexpected circumstances” (Jungherr, Schroeder 2023: 1).

More generally, Floridi (2022) succinctly defines it as “a reserve of ability to act at hand” (53). The definitive element is linked to the ability to perform tasks or imitate behaviors commonly associated with human (or animal) intelligence, such as visual perception, speech recognition, decision-making, linguistic translation, and making of generalizations. As AI algorithms increasingly attain these human-like capabilities, they are better equipped to interact with humans and engage in the entire set of social processes. As adaptive entities capable of effectively executing tasks while learning from their environment, these agents cannot be merely considered mathematical equations, but rather as “teleological (goal-driven) agents” (Cristianini et al. 2023), whose role is growing in proportion to the increasing technological mediation of our societies.

Technology, along with mass media, has always been intertwined with society (DeFleur, Ball-Rokeach 1989). However, since the advent of the current “network society” (Castells 1996), “information” has essentially become the fundamental element of every social process, and the so-called “infosphere” (Floridi 2009) has come to encompass every moment of our social lives, creating an inextricable relationship between our daily existence and its technological dimension. Social sciences must therefore study these agents in interaction, considering the broader social context (Lindgren, Holmstrom 2020) and highlighting both opportunities and, especially, risks. From this perspective, some of the main recurring concerns are related to the data we produce and these agents may use to adapt our performances to their objectives, emphasizing the risk of the automatic, rapid, and uncontrolled spread of renewed forms of “surveillance” (Zuboff 2019), social control, or manipulation (Beer 2009; Fuchs et al. 2024); or even the potential errors or biases that seem to affect them due to their training (Noble 2018), increasing risks of discrimination and social injustice, impacting various levels of social dynamics, from school admissions to employment, to democratic matters (Lindgren, Holmstrom 2020; Boccia Artieri 2020; Sartori, Theodorou 2022). Thanks to their ability to interact effectively within all sorts of technologically mediated social processes, AI agents may also interfere with information access (Simon 2022) or be used to exacerbate already existing dangerous phenomena, such as infodemia and fake news (Floridi, Chiaritti 2020; Jungherr, Schroeder 2023).

Particularly, the newest “generative” agents, such as ChatGPT, may indeed be considered “communication agents” (Esposito 2022), capable of interacting directly with us as “true” social agents (Bennato 2023). Addressing these challenges, Floridi and Chiaritti (2020) stated:



A better digital culture will be required to make current and future citizens, users, and consumers aware [...] and hence able to understand and leverage the huge advantages offered by advanced digital solutions such as GPT-3, while avoiding or minimizing their shortcomings (692–693).

### **1.1 The relevance of AI imaginaries and narratives**

In terms of digital culture, one of the most important starting points is likely the cultural representations we have of, and are constructing about, AI and its role in society. This point is specifically related to the social construction of AI and its relevance as a socio-technical phenomenon. In fact, the development of a new technology (especially those considered “disruptive”) has always impacted the collective imaginary. However, its implementations and applications depend on the social context, its historical dynamics and power relations, as well as the social representations of that same technology. These representations include implicit or explicit assumptions, knowledge, expectations, etc., and can be expressed through verbal language or images, take specific narrative forms, or spread through metaphors. In any case, they represent the shared symbolic meanings of a given technology and contribute to determining its development, acceptance, and purposes (Cave et al. 2018; Lindgren, Holmstrom 2020). The Internet offers a relatively recent example of these dynamics, considering how the academic-scientific and hacker cultures of the early developments contributed to the Californian utopia, characterizing the Internet as an intrinsically libertarian technology (Sartori, Bocca 2022).

In brief, it is possible to state that if the intrinsic power relations of the social context (e.g., the economic interests related to the development of a certain technology) contribute to defining a technology as a “social object”, the interpretative frames and narratives used to refer to, describe, and illustrate it allow us to also consider it as a “cultural object” (Griswold 2012).

From this perspective, Sartori and Bocca (2022) specifically distinguished between “sociotechnical imaginaries” and “narratives”. Sociotechnical imaginaries can be defined as “collectively held, institutionally stabilized, and publicly represented visions of desirable futures, animated by shared understandings of forms of social life and social order achievable through and in support of advances in science and technology” (Jasanoff 2015, cited in Sartori, Bocca 2022: 445). Narratives, on the other hand, can be considered “building blocks” in the construction of the aforementioned imaginaries and, therefore, as a sort of “organizing visions” (or “frames”) of society (Sartori, Bocca 2022). Both imaginaries and related narratives often derive from literature or movies and, in general, from popular culture, including the information system.

These imaginaries can be traced throughout the entire history of technological and scientific development and can be defined in terms of “hope” and “fear”, “utopian” versus “dystopian” views (Sartori, Bocca 2022), “hype and hope” versus



“gloom and doom” representations (Ouchchy et al. 2020), or even in terms of “solutionist” versus “critical” perspectives (Boccia Artieri 2020), always split into two large opposing dimensions that reflect Eco’s famous dichotomy between the “apocalyptic” and the “integrated” (Eco 1994).

Imaginaries	
Positive/Utopian	Negative/Dystopian
<i>Immortality/Gate to Heaven</i> AI is seen as a vector for human enhancement, better health, and longevity.	<i>Dehumanization/Conflict</i> AI is viewed as a pathway to the loss of human specificities.
<i>Freedom/Social progress and economic development</i> AI allows for improved freedom, e.g., by reducing time and effort in routine tasks.	<i>Obsolescence/Kasparov syndrome</i> AI is perceived from a Luddite perspective as a threat of employment substitution.
<i>Gratification/Helping hand</i> AI is a source of gratification that enhances the human experience.	<i>Alienation/Shortcomings</i> AI is viewed as a source of new anxieties, stress, etc.
<i>Domination/Public accountability</i> AI is controlled by humans and serves them, particularly to better manage and accelerate social processes.	<i>Uprising/Frankenstein monster</i> AI is viewed in the context of uprisings where humans are controlled by machines or through them in authoritarian ways.

**TAB. 1** – *Utopian-Dystopian taxonomy of AI imaginaries (adapted from Cave and Dihal 2019 and Cools et al. 2022)*

Starting from this dichotomy, Cave and Dihal (2019) identified at least four additional categories for each area. These categories are presented and explained in Tab. 1 and are associated with similar categories defined by Cools et al. (2022), who analyzed AI-related newspaper articles over a period of almost forty years.<sup>6</sup>

### 1.2 AI media coverage and the impact of ChatGPT

The media coverage of AI appears to replicate the aforementioned dichotomy between positive/utopian and negative/dystopian perspectives (Ouchchy et al. 2020; Cools et al. 2022; Bartolomew, Mehta 2023), while also providing an opportunity to actualize the relationship between AI development and the broader social context.

Through the investigation of media coverage, we obtain a current representation of themes and evolving frames and narratives associated with AI, sometimes focusing on specific aspects, such as ethics (Ouchchy et al. 2020).

Numerous investigations and reports on this topic are often conducted by journalistic organizations, media observatories, NGOs, etc. One such report was

<sup>6</sup> These categories are alternatively referred to as narratives or imaginaries. Here, the term “imaginaries” is used, while the term “narratives” is limited to the specific storytelling of each article, which may or may not match with a specific sociotechnical imaginary.



carried out by the *Reuters Institute for the Study of Journalism*, which analyzed UK media coverage of AI in 2018.

Some of the main conclusions of the report highlighted that nearly 60% of the articles centered on new industrial products, announcements, and initiatives related to AI. The channels considered regularly covered industrial promotional events, start-ups, acquisitions, investments, and conferences. One-third (33%) of the coverage was based on industry sources, primarily CEOs or senior executives—six times more than the number of articles based on government sources and double those based on academic sources. Twelve percent of the articles cited Tesla and SpaceX entrepreneur Elon Musk, while AI-related products were often described as optimal solutions to various collective problems, such as cancer or renewable energy.

Potential negative effects were rarely discussed, and media coverage was at least partially politicized: sources referring to a conservative political universe emphasized geopolitical or economic issues, while sources with a progressive slant focused more on ethical aspects, such as potential discriminatory biases or privacy (Brenner et al. 2018). A more recent study on AI media coverage by a team of international researchers has highlighted similar conclusions. In particular, they noted that said media coverage substantially reflected economic and governmental interests, praising AI's potential while underestimating the power dynamics underlying these interests. The most cited critical voice was physicist Stephen Hawking. However, he was cited very infrequently, and the views of social scientists seemed conspicuously absent (Dandurand et al. 2023). However, as previously mentioned, the launch of ChatGPT has been described as a “disruptive” event in the development of AI and in the public discussion surrounding it. Recent research specifically focused on the media coverage of ChatGPT in the US context highlighted how narratives, in this case, are often led by the companies involved and the CEO of OpenAI, who widely uses social media to promote his creation, while journalists utilize the same social media as a source of information (Bartolomew, Mehta 2023).

Moreover, researchers noted that television networks and news programs largely referred to imagery linked to science fiction, triggering a “Hollywoodification” process ranging from the “end of work” to the “destruction of humanity”, translating uncertainty into fear rather than fostering understanding (Bartolomew, Mehta 2023).

They suggested that this attitude stems from the need for journalists and publishers to cover this topic, which has garnered more attention than other technological innovations, likely because it directly involves the journalistic profession and, in the long term, the very structure of the public arena.

Similarly, we observe such dynamics in Italy, with increasing media coverage of AI-related topics (Degli Esposti, Tirabassi 2024). However, the Italian context suffers from a lack of systematic analysis of this media coverage. Nonetheless, an interesting attempt in this direction was made by the Pavia Observatory, which investigated coverage of said topics between 2018 and 2020 in the prime-time news programs of major national (private and public) networks (Sarani, Cobianchi 2021). In this case, some frames partially emerged from the semantic analysis of the most frequently used terms, summarized in dichotomies such as “*machine vs human*”, *highlighting*



aspects of increasing interaction (and sometimes conflict), "*future vs present*", *emphasizing* the momentum toward progress but also uncertainty, and "*first vs last*", highlighting aspects of stratification and competition (both between human-machine and different skill levels). However, it is not yet possible to evaluate the quality of information and debate in this fundamental field or trace the related imaginaries and narratives shaping the adoption and implementation of AI.

This investigation aims to contribute to filling this gap, with a specific focus on the impact of ChatGPT. Specifically, the analysis seeks to answer the following research questions:

RQ1) How did mainstream media in Italy represent the phenomenon of AI, particularly in terms of narratives and imaginaries?

RQ2) What was the specific impact of ChatGPT on the Italian public debate regarding artificial intelligence?



## 2. Methodology

To address the aforementioned research questions, a sample of news articles from the most relevant Italian newspapers was extracted online using the "Media Cloud" platform. The newspapers were selected based on the following criteria: *a*) national and generalist coverage, *b*) circulation<sup>7</sup>, and *c*) political orientation<sup>8</sup>.

The sample was constructed considering two periods: a six-month period before and after ChatGPT's launch on November 30, 2022. Fig. 1 represents a concise scheme of the sample construction.

The first sample was collected using the following keywords: "*intelligenza artificiale*", "*artificial intelligence*", "*IA*", and "*chatbot*" (the term "AI" could not be used because it was interpreted in the Italian language as a preposition). The initial sample comprised 699 items (matching content=0.4%), which was subsequently reduced to 101 items by including only those that had the keywords in their title.

The second sample was collected by including the keywords "*ChatGPT*" and "*OpenAI*", resulting in a sample of 1884 items (matching content=0.82%). These were subsequently reduced to 519 items, again excluding those that did not have the keywords in their title. The final sample comprised 620 items. However, due to the tools used, the sampling was not uniformly distributed across the considered newspapers. "La Repubblica" and "Il Sole 24 Ore" were overrepresented, while others were not represented at all. For this reason, "Avvenire" and "La Verità" were excluded

<sup>7</sup> Source: Italian Federation of the Newspaper Editors: [www.fieg.it](http://www.fieg.it) (8/12/2023).

<sup>8</sup> The selected newspapers were categorized into three groups: *a*) *Conservative*: Il Giornale, Libero, La Verità, Avvenire; *b*) *Progressive*: La Repubblica, Il Fatto Quotidiano; and *c*) *Equilibrated*: Il Corriere della Sera, La Stampa, Il Messaggero, Il Sole 24 Ore. This categorization was developed using different sources, notably the Pew Research Center's report titled "*News Media and Political Attitudes in Italy*" (2018), available at: <https://www.pewresearch.org/global/fact-sheet/news-media-and-political-attitudes-in-italy>.

from the sample, while some additional items from other sources were unavailable due to paywalls. The definitive total sample was thus equal to N=534 items (sample 1=99; sample 2=435), analyzed using the computer-assisted content analysis methodology (with the software N-Vivo 2020)<sup>9</sup>.

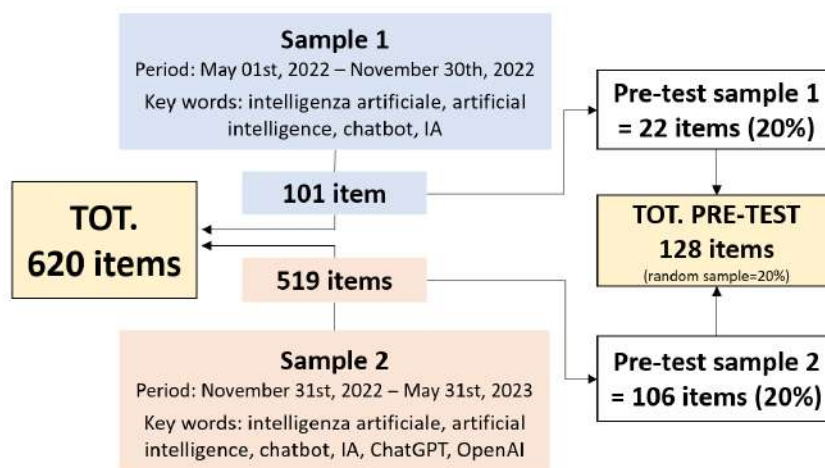


FIG. 1 – A basic representation of the sampling scheme

The items were originally classified by “newspaper” and “period” (pre-/post-ChatGPT) and progressively cataloged according to further attributes, such as “Category/column”, “Month” and “Year”, “Scope”, “Political orientation”, and “Focus on ChatGPT” (or not).

A pretest was conducted in advance on a representative and randomly selected sub-sample of 128 items (corresponding to 20% of the entire sample).

The coding was implemented in an inductive-deductive manner (Fereday, Muir-Cochrane 2006), initially considering evidence from the literature review and the specific research objectives, and subsequently implementing further nodes during the pretest analysis. Particularly, according to Tab. 1, the coding retained the distinction between “Utopian” and “Dystopian” categories (and their respective subcategories), adding the nodes “Generic danger” or “Generic benefit” for generic content that did not fit the already defined categories, “Balanced” for narratives that were relatively balanced between Utopian and Dystopian views, and “Human-like intelligence” for content primarily focused on the topic of “general AI”.

A specific node was dedicated to “Imaginaries”, considered as references to popular imaginaries explicitly cited in the article, such as the super-intelligent computer HAL 9000 in “2001: A Space Odyssey”, the android T-800 in “Terminator”, the Frankenstein monster by Mary Shelley, etc.

<sup>9</sup> At the current level of regulation there is no way to verify whether some of the selected articles (or parts of them) have been written by a human or a chatbot. Surely, we can confidently declare that this scientific paper contains 100% human-written content.

N	Newspaper	Sample 1		Sample 2		TOT	TOT def.	%
		Items	Pretest	Items	Pretest			
1	CORRIERE DELLA SERA	0	0	14	3	14	13	2%
2	REPUBBLICA (LA)	29	8	193	33	222	186	35%
3	SOLE 24 ORE (IL)	22	4	105	22	127	126	24%
4	STAMPA (LA)	1	1	17	4	18	17	3%
5	MESSAGGERO (IL)	6	1	20	4	26	25	5%
6	FATTO QUOTIDIANO (IL)	3	1	36	7	39	35	7%
7	GIORNALE (IL)	4	1	24	5	28	28	5%
8	LIBERO	24	5	85	16	109	104	19%
	TOT	101	22	519	106	620	534	100%

**TAB. 2** – Distribution of the total and pretest sample



Some nodes considered content such as “Ethics”, “Responsibility”, “Regulatory” aspects, explicit “Reference to ChatGPT” (in articles that were not focused on it), “ICT” for merely technical content, or even “Learning opportunity” (when the article offered the opportunity to increase knowledge and awareness about AI).

Further nodes were also established to trace the source type. They were categorized as “Independent” if not directly related to the development of AI systems or associated products and services, and as “Integrated” if economically linked to AI developers and/or implementers. Moreover, the analysis also considered the articles’ headlines, evaluating their “Consistency” or “Inconsistency” with their contents and potential “Alarmism or sensationalism.”

In terms of inter-rater reliability, researchers conducted the pretest coding phase together, consistently exchanging and controlling information while revising the coding and data interpretation. They maintained a close relationship and continuous communication throughout the entire analysis of sources.

### 3. The main results and the pre-/post- ChatGPT confrontation

Considering the analyzed sample and initially focusing on the Utopian/Dystopian/Balanced categories (Tab. 3), it was evident that narratives referring to Utopian imaginaries represented the majority, at 50%, while those representing Dystopian imaginaries accounted for a quarter of the sample (26%). “Balanced” items represented 20% of the sample.<sup>10</sup> The Utopian subcategories appeared to be well-distributed, with a prevalence of those defined as “Generic benefits” (12%) and, particularly, “Dominance” (14%). The less represented subcategories were “Immortality” (6%) and “Gratification” (8%).

<sup>10</sup> The total percentage does not add up to 100% because some articles do not fit into any of these categories.

Conversely, in the Dystopian subcategories, there was a prevalence of "Generic danger" (11%) and a clear emphasis on "Obsolescence" (8%), while the less represented subcategories were "Dehumanization" (1%) and "Alienation" (2%).

Coding	Newspapers*								Tot. sample		Tot. pre-ChatGPT		Tot. post-ChatGPT	
	a	b	c	d	e	f	g	h	TOT	%	TOT	%	TOT	%
<i>Utopian</i>	2	91	60	6	19	19	10	60	267	50%	60	61%	191	44%
<i>Generic benefit</i>	1	24	12	1	3	6	1	16	64	12%	3	3%	61	14%
<i>Immortality</i>	0	16	1	0	3	2	2	8	32	6%	10	10%	17	4%
<i>Freedom</i>	0	17	8	3	4	5	3	13	53	10%	8	8%	44	10%
<i>Gratification</i>	0	14	16	1	4	2	0	6	43	8%	10	10%	30	7%
<i>Dominance</i>	1	20	23	1	5	4	4	17	75	14%	29	29%	39	9%
<i>Dystopian</i>	8	60	23	8	12	14	9	5	139	26%	18	18%	109	25%
<i>Generic danger</i>	1	34	13	1	5	2	2	1	59	11%	9	9%	44	10%
<i>Dehumanization</i>	0	3	1	0	1	0	0	0	5	1%	1	1%	9	2%
<i>Obsolescence</i>	4	16	6	4	3	5	3	2	43	8%	5	5%	30	7%
<i>Alienation</i>	0	2	2	0	2	4	1	0	11	2%	1	1%	9	2%
<i>Uprising</i>	3	5	1	3	1	3	3	2	21	4%	2	2%	17	4%
<i>Balanced</i>	2	50	16	2	5	10	3	19	107	20%	12	12%	126	29%

**Tab. 3** – Coding Utopian/Dystopian/Balanced categories: whole and pre-/post-ChatGPT sample  
 (\*a=Il Corriere della Sera; b=La Repubblica; c=Il Sole 24 Ore; d=La Stampa; e=Il Messaggero; f=Il Fatto Quotidiano; g=Il Giornale; h=Libero).



As for the remaining categories (Tab. 4), articles characterized by a tone of "Criticism"—often providing in-depth analyses of AI implications and social impacts without any particular narrative—remained a minority (7%), as did those that offered "Learning opportunities" or insights (13%). Explicit references to well-established AI imaginaries, used as a sort of stock from which images were drawn as needed, were limited to 8%. Only 5% discussed the topic of "Human-like intelligence".

Other content categories such as "Ethics", "Regulation", and "Responsibility" were rated at 20%, 18%, and 12%, respectively. ChatGPT was explicitly cited in 51% of the items, even in articles not focused on it.

Regarding sources, "Independent" ones slightly outnumbered "Integrated" ones (at 39% and 37%, respectively), while the headlines exhibited a high "Consistency" rate (91%), with 30% exhibiting "Alarmism/Sensationalism".

Finally, by aggregating the percentages of the different newspapers according to political orientation, we can derive further interesting insights. For example, newspapers leaning toward the progressive area exhibited a high percentage of both positive and negative narratives (averaging 60% and 53%, respectively). These type of newspapers also had the highest percentage of "Balanced" items (28% versus 13% for balanced newspapers and 15% for conservative ones).

Progressive and balanced newspapers particularly emphasized the theme of work ("Obsolescence" reached an average of 17% and 20%, respectively, with 8% for

conservatives) and provided more opportunities for critical contributions or learning about AI.

Conversely, conservative newspapers exhibited a lower rate of dystopian perspectives (34%, compared to 53% and 50% for progressive and balanced newspapers, respectively), while also offering more integrated sources (41% versus 32% and 24% for progressive and balanced newspapers, respectively), although they appeared to be less sensationalistic (23% versus 36% for progressive newspapers and 38% for balanced newspapers).

Coding	Newspapers								Tot. sample		Tot. pre-ChatGPT		Tot. post-ChatGPT	
	a	b	c	d	e	f	g	h	TOT	%	TOT	%	TOT	%
<i>Human-like intell.</i>		13	2	1	1	4	2	3	26	5%	2	2%	24	6%
<i>Criticism</i>	4	10	7	0	5	3	2	5	36	7%	1	1%	35	8%
<i>Responsibility</i>	4	14	22	4	2	7	6	3	62	12%	7	7%	55	13%
<i>Regulation</i>	2	39	22	6	3	6	10	8	96	18%	10	10%	86	20%
<i>Ethics</i>	2	48	17	2	3	4	11	18	105	20%	10	10%	95	22%
<i>ICT</i>	1	33	14	3	4	9	8	23	95	18%	12	12%	83	19%
<i>Learning opp.</i>	1	34	7	8	5	6	5	5	71	13%	7	7%	64	15%
<i>Imaginaries</i>	2	19	2	4	2	7	5	1	42	8%	7	7%	35	8%
<i>Ref. to ChatGPT</i>	6	102	51	8	8	19	12	20	226	42%	4	4%	222	51%
<i>Sources</i>														
<i>Independent</i>	6	84	24	9	13	18	14	37	205	38%	23	23%	183	42%
<i>Integrated</i>	1	66	46	5	6	10	9	52	195	37%	38	38%	157	36%
<i>Headlines</i>														
<i>Consistency</i>	12	179	119	15	23	32	27	79	486	91%	72	73%	412	95%
<i>Inconsistency</i>	1	7	7	2	2	3	1	25	48	9%	26	27%	22	5%
<i>Alarmism/Sens.m</i>	8	54	49	4	7	15	9	14	160	30%	29	29%	131	30%

**Tab. 4** - Coding of the rest of the categories: whole and pre-/post-ChatGPT samples.

### 3.1 The pre- and post-ChatGPT confrontation analysis

A further step of the analysis involved comparing the situation before and after the advent of ChatGPT. Prior to that advent, narratives aligned with "Utopian" imaginaries were the highest (63%), with a decisive prevalence of "Dominance" (29%). Conversely, "Dystopian" perspectives were limited to 18%, primarily narrated in terms of "Generic danger" (9%) and "Obsolescence" (5%). "Balanced" items accounted for 12% (Tab. 3).

Newspaper articles that offered "Learning opportunities" and insights were limited to 7%, while "Criticism" was practically nonexistent (1%). Explicitly cited (pop) "Imaginaries" were also limited (7%), and "Ethics" and "Regulation" both reached 10%, while "Responsibility" was only 7%. "Integrated" sources were prevalent (38% versus 23% for "Independent" ones), as were consistent headlines (73%). However,



26% of the headlines appeared “inconsistent”, and 29% were alarmist or sensationalistic.

In contrast, in the post-ChatGPT sample, “Utopian” perspectives declined to 44%, while “Dystopian” ones increased to 25%. The main utopian perspectives shifted toward “Generic benefits” (14%) and “Freedom” (10%), while the dystopian narratives remained aligned with “Generic danger” (10%) and “Obsolescence” (7%). “Balanced” items more than doubled, reaching 29%.

“Learning opportunities” also doubled (15%), while “Criticism” and arguments such as “Ethics”, “Regulation”, and “Responsibility” increased significantly, reaching 8%, 22%, 20%, and 13%, respectively. “Imaginaries”, however, remained stable at 8%.

Finally, the items in this sample showed a prevalence of “Independent” sources, reaching 42% compared to 36% for “Integrated” ones. Regarding the headlines, “Consistency” surged to 95%, but the tendency toward “Alarmism/Sensationalism” remained stable at 30%.



#### 4. Discussion and Conclusion

The primary aim of this paper was to analyze the content of a significant sample of AI-related articles from Italian newspapers to trace and highlight AI imaginaries and narratives, especially in the wake of the launch of ChatGPT.

To achieve this, a series of narratives were considered, along with references to well-established imaginaries from popular culture and literature. Other features were also considered, such as learning opportunities and/or insights to enhance readers' awareness, ethics-related issues, regulation and responsibility matters, as well as the quality of headlines and sources.

The analysis confirms that Italian news media tend to discuss AI in terms of a utopian-dystopian dichotomy. However, positive/utopian narratives seem to prevail, particularly those promising dominance over social and economic processes, enhancing well-being, security, and (especially) productivity.

Conversely, the most emphasized dangers relate to the role of humans in these processes, particularly highlighting the neo-Luddite risks of substitution and obsolescence, with consequences for jobs and employment. This is partially a consequence of the ChatGPT turning point, where we observe an increase in these potential negative aspects (especially in progressive and balanced newspapers), partially countered by an emphasis on AI's potential for freedom (especially in conservative newspapers).

Despite the increasing number of dystopian narratives, alarmism and sensationalism remained limited to 30% (and did not increase as a result of ChatGPT's impact), while the risk of “Hollywoodification” seems to have been mitigated. Conversely, balanced articles, learning opportunities, attention to “independent” sources, and discussions on regulation and responsibility have significantly increased. This could be interpreted as a result of journalists'

professional efforts, reflecting an expansion of the debate beyond just professionals to a wider range of social actors.

In conclusion, the growing number of news stories devoted to AI and their pervasive reference to ChatGPT, following its launch, support the initial hypothesis regarding ChatGPT's role and impact on public debate. Moreover, the data suggest that ChatGPT also influenced AI narratives, increasing attention to themes such as AI regulation or ethics and providing further learning opportunities for the public.

However, these latter changes cannot be precisely attributed to the mere hype surrounding the launch of ChatGPT. Additionally, various economic or political variables cannot be definitively excluded.



## Bibliography

- Bartolomew J., Mehta D. (2023), "How the media is covering ChatGPT", *Columbia Journalism Review*, Retrieved on 10 May 2024 from [https://www.cjr.org/tow\\_center/media-coverage-chatgpt.php](https://www.cjr.org/tow_center/media-coverage-chatgpt.php).
- Beer, D. (2009), "Power through the algorithm? Participatory web cultures and the technological Unconscious", *New Media & Society*, 11(6): 985-1002.
- Bennato D. (2024), *La società del XXI secolo. Persone, dati, tecnologie*, Roma-Bari, Laterza.
- Boccia Artieri, G. (2020), "Fare sociologia attraverso l'algoritmo: potere, cultura e agency", *Sociologia Italiana*, 15: 137-148.
- Brenner J.S., Howard P.N., Nielsen R.K. (2018), "An Industry-Led Debate: How UK Media Cover Artificial Intelligence (Factsheet)", *Reuters Institute for the Study of Journalism*, Retrieved on 10 May 2024 from: [https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-12/Brennen\\_UK\\_Media\\_Coverage\\_of\\_AI\\_FINAL.pdf](https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-12/Brennen_UK_Media_Coverage_of_AI_FINAL.pdf).
- Castells M. (1996), *The information age: economy, society and culture. Volume 1: The rise of the network society*, Cambridge (MA, US), Blackwell.
- Cave S., Craig C., Dihal K., Dillon S., Montgomery J., Singler B., Taylor L. (2018), *Portrayals and perceptions of AI and why they matter*, London, The Royal Society.
- Cave S., Dihal K. (2019), "Hopes and fears for intelligent machines in fiction and reality", *Nat Mach Intell*, 1:74-78.
- Cools H., Baldwin V.G., Opgenhaffen M. (2022), "Where exactly between utopia and dystopia? A framing analysis of AI and automation in US newspapers", *Journalism* 0(0): 1-19.
- Cristianini, N. (2023), *The shortcut: why intelligent machines do not think like us*, Boca Raton (FL, US), CRC Press.
- Cristianini N. (2024), *Machina sapiens. L'algoritmo che ci ha rubato il segreto della conoscenza*, Bologna, Il Mulino.
- Dandurand, G., McKelvey, F., & Roberge, J. (2023), "Freezing out: Legacy media's shaping of AI as a cold controversy", *Big Data & Society*, 10(2): 1-13.



DeFleur M., Ball-Rokeach S. (1989), "Media system dependency theory", in DeFleur M, Ball-Rokeach S (ed.) *Theories of mass communication*, New York Longman, 292-327.

Degli Esposti P., Tirabassi L. (2024), "The Human-algorithmic Entanglement in the News Realm", *Problemi dell'informazione*, 1: 41-64.

Esposito E. (2022), *Artificial communication. How algorithms produce social intelligence*, Boston (US), Mit Press.

Fereday, J., Muir-Cochrane E. (2006), "Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development", *International journal of qualitative methods*, 5(1): 80-92.

Floridi L. (2009), *Infosfera. Filosofia e Etica dell'informazione*, Torino, Giappichelli Editore.

Floridi L., Chiriatti M. (2020), "GPT-3: Its Nature, Scope, Limits, and Consequences", *Minds and Machines*, 30: 681-694.

Fuchs C., Can Arslan S., Allmer T. (2024) "Critical Perspectives on Digital Capitalism: Theories and Praxis. Introduction to the Special Issue", *tripleC* 22 (1): 140-147.

Griswold, W. (2012), *Cultures and societies in a changing world*, London, Sage.

Lindgren S., Holmstrom J. (2020), "A Social Science perspective on Artificial Intelligence: building blocks for a research agenda", *Journal of Digital Social Research*, 2(3): 1-15.

Mitchell M. (2019), *Artificial Intelligence: A Guide for Thinking Humans*, London (UK), Penguin.

Noble S. (2018), *Algorithms of Oppression: How Search Engines Reinforce Racism*, New York, New York University Press.

Ouchchy L., Coin A., Dubljević V. (2020), "AI in the headlines: the portrayal of the ethical issues of artificial intelligence in the media", *AI & Society*, 35: 927-936.

Jungherr A., Schroeder R. (2023), "Artificial intelligence and the public arena", *Communication Theory*, 33(2-3): 164-173.

Sarani G., Cobianchi V. (2021), "L'intelligenza 'media' artificiale", *Osservatorio di Pavia*, 16 febbraio 2021, Retrieved on 10 May 2024 from: <https://www.osservatorio.it/lintelligenza-media-artificiale>.



Stefano Spillare, Michele Bonazzi, Piergiorgio Degli Esposti  
*AI Imaginaries and Narratives in the Italian Public Discourse*

Sartori L., Theodorou A. (2022), "A sociotechnical perspective for the future of AI: narratives, inequalities, and human control", *Ethics and Information Technology*, 24(4): 1-11.

Sartori L., Bocca G. (2022), "Minding the gap(s): public perceptions of and socio-technical imaginaries", *AI & Soc*, 38: 443-458.

Simon M.F. (2022), "Uneasy Bedfellows: AI in the News, Platform Companies and the Issue of Journalistic Autonomy", *Digital Journalism*, 10(10): 1832-1854.





A JOURNAL OF THE  
SOCIAL IMAGINARY

# AI in the Public Sector: An Exploration of Social Imaginaries

Antonio Opromolla

a.opromolla@unilink.it

Humanities Dep. | Link Campus University

Silvia Di Rocco

silvia.dirocco@linkstudents.it

Link Campus University

Lorenza Parisi

l.parsi@unilink.it

Humanities Dep. | Link Campus University



## Abstract

The use of artificial intelligence (AI) enables the transformation of both organizations and institutions alike in many different fields. This paper explores the social imaginaries (i.e. set of imagined and collectively shared ideas, practices, and values people use to construct and legitimize reality) in relation to the use of AI in the public sector. According to Castoriadis (1975), human imagination plays a key role in the process of the “self-generation” of the current social order. Indeed, technological affordances and social practices are intertwined and mutually shape each other, and both contribute to the generation of what is called the “social imaginary.” Through a discourse analysis, the paper investigates the most popular narratives in European and Italian public policies framing and regulating the implementation of AI in the public sector. Then, the paper explores the AI imaginaries of Italian public sector through 15 interviews with AI experts (i.e., designers and managers) from the Italian public sector. Differences and similarities in the AI social imaginaries between policies and experts are highlighted. Experts advocate the need for a human-centered design approach in the implementation of AI to map users’ needs and to design useful public services. By adopting Castoriadis’ (1975) approach, we identify opposition between the *heteronomous imagination*, the one imposed by public policies, and the *autonomous imagination*, based on citizens’ needs and perceptions, including emotional ones. In conclusion, the paper calls for a human-centered design approach in the implementation of AI in the public sector.

## Keywords

Social Imaginary | Artificial Intelligence | Public Sector | Human-Centered Design  
| Experts



## 1. Imaginary and Society

In *The Imaginary Institution of Society* (1975), Castoriadis considers institutions to be social constructs that evolve according to different historical and socio-cultural contexts, describing the process of the “self-generation” of the social order and the ongoing redefinition of institutional regulatory mechanisms. These elements are socially created based on everyday dynamics and routines, and they constitute what he calls the “social imaginary,” defined as “an incessant and indeterminate socio-historical creation of figures, forms, images, from which people can talk about something. Reality is the construction of this imaginary” (1975, p. 32). As a result, history can be considered a *poiesis* created within a specific society, where historical narratives contribute to shaping the collective social imaginary (Ricoeur, 2000).

Castoriadis claims human imagination plays a vital role in the generation of the social order. Indeed, through creativity, human societies transform and establish themselves and produce the related social imaginary. Therefore, the emerging transformative processes are actualized by social practices expressing shared meanings and thus shaping the collective identity and individual and social behaviors. In this sense, the imaginary is produced *within* society.

According to Castoriadis, a plurality of imaginaries coexists within a society.

Indeed, dominant symbols and meanings legitimized and shared by institutions can become stiff and oppressive; therefore, a consistent reimagining of new organizational models, values, and norms that will lead to social change and innovation is needed. The conflict between a central and a peripheral imaginary can become a driver of the social transformation of the collective social imaginary. In greater detail, Castoriadis argues that the *heteronomous imagination* dimension (i.e., the social imagination in which meanings, representations, and norms are imposed) opposes the *autonomous imagination* dimension (i.e., people’s ability to create and self-determine and envision institutions and related meanings). Further, Castoriadis urges the search for an autonomous imagination in the construction of the social imaginary, encouraging people to become aware and to participate actively in social transformation; similarly, de Certeau (1980) highlighted the role of imagination in resisting dominant power structures.

Along with human creativity, Castoriadis also recognizes the transformative power of technology (*teuchein*, see Di Blasi, 2013), promoting autonomy and creativity.

Indeed, he views the technical and social dimensions as inextricably linked, particularly because the first represents the materialization of the significations of the social imaginary (Centorrino & Di Paola, 2022). While technologies are instruments of domination and control, they also facilitate communication, social organization, and the production of alternative meanings, offering new opportunities for social transformation.

According to Castoriadis, technologies cannot be conceived as merely instrumental or functional; rather, technologies and social practices are intertwined and mutually shape each other, contributing to the generation of the social



imaginary. In addition, in the social sciences, the concept of *socio-technical systems* emphasizes the interdependence between the social and technical dimensions.

According to Miller (1978), societies are characterized by a complex interaction between social organizations (including social institutions, interpersonal relations, and cultural processes) and technological systems (including communication technologies, infrastructures, and scientific innovations). That is, socio-technical systems combine social and technical concepts and evolve together along with people's experiences, needs, beliefs, and expectations about social change. For this reason, it is important to analyze and design interactions with socio-technical systems, focusing on the related social impacts, the characteristics of digital interfaces, and users' capacity to interact with these systems effectively.

According to these theoretical assumptions and considering the interdependence between the social and technical dimensions, the paper describes emerging conflicts in the social imaginary related to the use of artificial intelligence (AI) in the public sector. The paper is structured as follows. First, it provides a literature review focusing on the main trends in research on the use of AI in the public sector (par. 2).

Afterward, it investigates priorities emerging from European and Italian policies regulating AI use in the public sector (par. 3); then, it explores the AI social imaginary of Italian public sector experts (par. 4). Differences and similarities between policies and experts in relation to the AI social imaginary are highlighted<sup>1</sup>. However, a conflict between these two imaginaries is observed to have emerged: while policies are more concerned with the transformative force of AI in the public sector, impacting both the economic sphere and societal challenges, experts in the public sector field advocate for the application of a human-centered design approach to map users' needs and to design useful services. In conclusion (par. 5), this paper affirms the importance of creativity and autonomous imagination (Castoriadis, 1975) in establishing a new social imaginary of AI in the public sector. In doing so, the human-centered design approach emerges as a key resource for implementing innovative solutions having greater alignment with citizens' needs.

## 2. Trends in Research on the Use of AI in the Public Sector

AI is a computer science research field aimed at developing systems and machines that can perform tasks that are usually completed by humans (Russel & Norvig, 2021). This is based on several technical solutions: learning from data, recognizing patterns, and behaving accordingly (e.g., chatbot, computer vision, intelligent objects, etc.); however, this literature review focuses primarily on the organizational and social implications of using AI.

---

<sup>1</sup> A similar study—starting from the work of Castoriadis—compared the differences between how AI is described in science fiction movies and how it emerges in the socio-technical imaginary associated with intelligent systems (Bory and Bory, 2016).



The scientific literature has described the relevance of AI in different domains (education, health, security, finance, etc.), including the public sector (Agarwal, 2018; de Souza et al., 2019; Wirtz et al., 2021); indeed, many institutions have already begun implementing AI technologies (van Noordt & Misuraca, 2022). According to Zuiderwijk et al. (2021), the use of AI in the public sector may contribute to increases organizational efficiency and performance; automation of simple tasks; improved risk identification; the prediction and monitoring of events; improved data and information management; support for decision-making processes; enhanced engagement and personalization; and increased speed and quality of public services.

Yet, the negative effects and risks are also discussed: Wirtz et al. (2019) divided these into four main categories: 1) societal, including issues concerning the social acceptance of AI, whether for lack of trust or fear; 2) ethical, such as discrimination emerging from biased data; 3) regulatory, focused on such legal issues as privacy or accountability; and 4) technological, comprising such issues as data integration and employee specialization.

The implementation of AI in the public sector represents an opportunity for transformation that may generate a relevant impact on citizens' lives (Mergel et al., 2018); however, it requires key organizational and social processes, such as employees' acquisition of new digital skills, adequate transparency and integrity levels, citizen participation, and the design of new interactive models. Moreover, some scholars have claimed the need to frame the governance of AI (Wirtz et al., 2020) due to its wide impact on society (Gahnberg, 2021). Indeed, AI in the public sector is considered a trailblazer for other domains (including private ones); therefore, national and international policies should set specific rules for the improved regulation of the Big Tech companies currently dominating the AI market (Radu, 2021; Khanal et al., 2024).

Recently social sciences scholars have started to investigate the use of AI technologies in specific public sector domains, such as the social impact of the adoption of automated decision systems in the public sector (Schiff, 2021); the predictive systems used in decision-making processes (McDonald et al., 2022); the application of natural language processing systems to the analysis of political speeches (Rice et al., 2020); the impacts of automated systems on willingness to participate in government programs (Miller et al., 2022); the application of robots in long-term care (Ying Tan et al., 2021); the integration of AI with Internet of Things to improve service delivery (Ishengoma et al., 2022); the use of chatbots in bureaucratic processes and its ethical issues (Salah et al., 2023); the application of data analytics in policymaking, especially oriented to the economic sector (Loukis et al., 2020); and the application of machine learning to estimate daily public transport usage rates and the behavior of mobile people (Sariyer et al., 2024). Several social scientists also have studied the role of algorithms in public leadership (Leighton, 2019) and their use in public service delivery (Nzobonimpa, 2023).

Other scholars have demonstrated the importance of studying stakeholders' opinions in relation to the application of AI in the public sector (Qian Sun et al., 2019).



For example, Ingrams et al. (2021) and Haesevoets et al. (2024) explored citizens' attitudes toward the use of AI in government decision making, registering ambivalent citizens' opinions. Moreover, Gesk et al. (2022) revealed that German citizens would prefer to limit the application of AI only to (general) public services (rather than to services more oriented toward specific target groups and to solve specific problems for which they prefer human interaction). According to Auernhammer (2020), human-centered design, a method employed in the implementation of socio-technical innovations through an iterative process<sup>2</sup>, may contribute to the implementation of effective AI systems intended to deliver solutions that match the specific characteristics of a particular context, including marginalized people and communities; to designing useful and usable user interfaces; to persuading people to adopt a desired behavior; and to supporting physical, intellectual, and emotional needs. In addition, the adoption of a human-centered design approach involving people in designing and implementing services and solutions may also provide citizens with the critical skills needed to confront AI (Vinuesa et al., 2020; Criado et al., 2019; Chui et al., 2018).

Quite recently, some Italian scholars from different research fields (though primarily the organizational, legal, and computer science disciplines) investigated the use of AI in the Italian public sector. After the COVID-19 pandemic (Lovari & Ducci, 2022), they became interested in investigating the role of the Next Generation program in funding these projects (Galetta, 2023b; Maria Cati, 2024; Raffiotta, 2022) and to explore potential risks that citizens may encounter (Barone, 2020). Scholars primarily analyzed the principal AI technologies employed: decision-automation-systems (Galetta, 2023a) and chatbots (Bellini et al., 2020; Federici et al., 2021). The most cited applications of AI in the Italian public sector include education, justice (Fatima et al., 2020), health (Cingolani et al., 2023), defense (Fanni et al., 2023), and the urban environment (Dughiero et al., 2021). A particular focus is also given to the use of big data in governing Italian municipalities, mentioning challenges and opportunities for public services (Dughiero et al., 2021; D'Albergo et al., 2022; Cavallo Perin, 2021). In conclusion, the sociological perspective exploring the use of AI in the Italian public sector is limited, necessitating further investigation into the cultural and social dimensions.

---

<sup>2</sup> Human-centered design can be intended as a specific methodology, with specific steps: 1. *analysis*, to understand the micro- and macrosocial contexts through observation and in-depth studies and to identify the way stakeholders interact among them and with the technological level; 2. *problem definition*, which intends to identify the main issue(s) to face, according to the insights that emerged in the previous phase; 3. *ideation*, aiming at generating creative solutions in socio-technical systems that intend to solve the identified problems; 4. *prototyping*, which intends to realize tangible versions of the (alternative) solutions previously ideated, giving people the chance to evaluate them in the next phase; 5. *test*, to obtain feedback and refine the solution, before proceeding with its implementation.



### 3. European and Italian Policies Regulating AI Use in the Public Sector

Following a discourse analysis method, we explored narratives of AI in public policies to highlight how they reflect and construct power relations and ideological positions (Hewitt, 2009). The focus is on both European and Italian policies, as they are considered strongly interconnected. Further, the analysis of key European policies (European Commission, 2018, 2019, 2020a, 2020b, 2021a, 2021b, 2021d, 2024) revealed the following primary patterns:

1. *AI as an Economic Force*: AI can drive economic growth. Thus, AI research and innovation can maintain European competitiveness, while a cross-sectoral (private/public) approach rejuvenates markets and industries.
2. *AI as a Transformative Force*: AI can drive social innovation in the environment, health, transport, agriculture, and education sectors. As such, a European framework with clear standards would establish a common basis for the development of national strategies.
3. *AI as a Force to Regulate*: Ethical and trustworthy AI is necessary, as is the provision of norms to ensure AI applications respect human rights and avoid discrimination and biases (*civil liability* of AI). In effect, transparency of algorithms, accountability, fairness, reliability, safety, and robustness are key points.
4. *AI as a Field to Develop*. The availability of high-quality data, as well as infrastructures for sharing and governing them, speed up AI development.
5. *AI in the Public Sector as a Trailblazer*. The public sector is considered a pioneering field in the adoption of AI and in the diffusion of a culture of innovation and experimentation. It improves efficiency and effectiveness in government operations (automating routine tasks, streamlining processes, and optimizing resource allocation) and enhances service delivery to users (personalizing services and anticipating needs, expectations, and preferences).
6. *AI and Humans*. EU policies emphasize the importance of a human-centered approach to AI, including the need to prioritize the way AI can benefit the society, the protection of human rights and the strength of inclusiveness, participation, engagement, and collaboration and to improve digital skills.
7. *AI as a result of collaboration*. This is achieved at both the international level, facilitating knowledge sharing and the dissemination of AI solutions, and the community level, strengthening interactions among government, citizens, businesses, associations, etc.

Specific European initiatives are monitoring the development, uptake, and impact of AI in Europe (Jorge Ricart et al., 2022; Manzoni et al., 2022). For instance, Tangi et al. (2022) monitored an AI application in the public sector, showing that most European initiatives act at the national level to stimulate awareness of AI (especially among civil servants), provide measures to improve data quality, carry out pilot testing, and develop normative frameworks. National strategies for the application of



AI to the public sector are implemented differently, as they involve different stakeholders (belonging to the private sector) providing different organizational solutions.

In relation to the Italian case, Italy is facing various changes and improvements in terms of digital innovation in the public sector (European Commission, 2022).

Recently, more attention has been paid to the development of public services according to a human-centered design approach (Fabbri & De Santi, 2018), to the delivery of plans to modernize the infrastructure (AgID, 2023), to the development of cross-sectoral services (e.g., digital identity systems), and to the reinforcement of civil servants' digital skills. Following a brief summary of the analysis, we analyzed Italian policies on AI, looking at main documents and guidelines (AgID, 2018; Ministero dello Sviluppo Economico, 2019; Ministero per l'Innovazione Tecnologica e la Digitalizzazione, 2022; AgID, 2023). The following main patterns emerged.



1. *AI as a Transformative Technology of the Public Sector.* AI is considered central to optimizing bureaucratic processes, thanks to its ability to automate routine tasks, reduce manual errors, and streamline workflows. Moreover, it can improve service delivery, offering personalized and responsive services to citizens, and it can increase overall efficiency within the public sector by enabling big data analysis, providing insights that can drive decision-making processes and, in doing so, can save costs.
2. *AI as a Human-Oriented Technology.* AI affords the opportunity to increase civic engagement and participation in government processes through easier feedback collection and analysis of social problems. Moreover, AI can enhance the user experience and personalize it to the specific needs and characteristics of people through continuous refinement and providing users with information so they can make conscious decisions. AI in the public sector can be adopted only if there are no violations of fundamental human rights. In the long term, AI can address societal challenges; promote equity, accessibility, and inclusivity; and create a more sustainable development.
3. *AI in the Public Sector as a Collaborative Outcome.* The implementation of AI in the public sector is possible only by leveraging diverse expertise to conduct comprehensive studies on AI's implications and impacts. This will provide policymakers with a clear understanding of AI's implications for governance and citizen services. Particular attention is paid to the role of public employees, who must develop digital skills and competences in data literacy.
4. *AI in the Public Sector as an Evolving field.* According to these policies, many challenges are faced in the development of an informed decision-making process. It is thus necessary to adopt a comprehensive, coordinated, and continuous approach to solve ethical issues, such as those related to fairness, transparency, accountability, and privacy, to strengthen social acceptance; create a legal framework for data protection, liability, intellectual property rights, and regulatory compliance; and address technical challenges, including algorithmic bias, data quality, interoperability, and scalability.

5. *AI in the Public Sector as a Field to Monitor.* This concerns many elements, including the importance of implementing sustainable solutions, monitoring the roles and responsibilities of people involved, and checking how data are collected and managed.

European and Italian AI policies share the goal of leveraging AI for economic and social progress, emphasizing human rights and inclusivity. They differ, however, in relation to the emphasis on regulation and on the methods for implementing AI in the public sector. While European policies adopt a more normative and standardized approach, Italian policies focus more on process optimization and efficiency, with greater attention to digital skills training and to enhancing collaboration at the national level.

#### 4. The AI Imaginary of Italian Public Sector Experts

To investigate experts' social imaginary in relation to the use of AI, the perceived priorities, and related risks and opportunities, we conducted 15 in-depth interviews with AI experts (i.e., designers and managers) from the Italian public sector. This research is part of a broader project with the aim of describing the social imaginaries of different stakeholders (public service experts, citizens, civil servants, private corporations, etc.) in relation to the use of AI in the Italian Public Administration (PA).

The first part of this research (Opromolla et al., 2024) is based on four focus groups involving 40 adolescents and young adults (aged 14–24 years). Young people possess a limited understanding of how AI functions, leading to significant apprehension regarding its utilization. This fear is particularly heightened by their lack of awareness regarding how AI processes their data and for what purposes.

In fall 2023, we conducted in-depth interviews with 15 AI experts (5 females and 10 males) working in the Italian public sector, more specifically, designers and managers working in digital transformation and public service design. We stopped recruiting when we achieved saturation (Hennink et al., 2017). The semi-structured interview method was selected to facilitate interviewer guidance and to ensure the achievement of objectives, as it offers flexibility to explore emergent points during the interview<sup>3</sup>.

The interviews investigated the following areas: 1) emerging trends, technologies, and best practices in the implementation of AI in the Italian and European public sectors; 2) related risks and opportunities; and 3) challenges in the design process

---

<sup>3</sup> Recently, a survey was conducted by Forum PA (2024) to investigate Italian public servants' perceptions of the integration of AI into Italy's public sector (1,500 respondents). Most employees believe AI can enhance efficiency and service quality, but a minority fears they may lose their jobs. Many employees believe AI will contribute to automating repetitive tasks, improving operational accuracy, and enabling employees to focus on higher-value work. However, some people raised concerns about data privacy and ethics and asked for robust regulatory frameworks.



and in the overall user experience, including citizens' involvement in the design process.

According to AI experts, promising fields of AI application in the public sector include mobility, especially logistic optimization, and fraud detection (e.g., social welfare fraud). The most cited solution is the chatbot (many experts mentioned "Bürokratt," the Estonian AI-based virtual assistant offering a comprehensive channel for accessing public informative services). Indeed, experts believe it is easier to apply AI to innovate communication with citizens, particularly between citizens and PAs, than the internal PA procedures. Most promising opportunities in the Italian context are related to the introduction of AI services for making decisions based on data analysis and to predictive analysis to anticipate issues and malfunction (e.g., in the healthcare and mobility sectors). Some obstacles in Italy are represented by the lack of a strategic approach to AI ("AI is different and more complex than ICT"); the absence of decisions informed by a rigorous data analysis; the lack of internal specialized professions with technical skills able to develop an AI solution and have a clear understanding of the design process, which results in public administration outsourcing of public services to external suppliers; and the lack of rigorous data ethics. Most cited risks are related to the use of controversial biometric technologies (see the Brescia controversy) or they may originate from errors in the AI outcome (see the scandal related to the misapplication of algorithms to spot childcare benefit fraud in the Netherlands).

Experts emphasize how in the Italian context, people are highly discouraged from using advanced systems, especially in public services, because citizens display a general lack of trust, considering the public administration old and stale; therefore, they fear both the inability of such systems to perform properly and their security aspects. Paradoxically, they say, people trust private services more than public ones.

Moreover, some experts state that certain attempts at incorporating AI applications into Italian public services have already begun, even if in a less advanced way than in other European countries. Unfortunately, these pilots are isolated examples, disconnected from the more general public context. Nowadays, the use of AI in the Italian public sector aims to improve the internal side of some public administration systems, such as increasing the level of efficiency, effectiveness, and precision in tax checking; tax evasion fraud detection; monitoring territorial data documentation analysis, etc. Other experts mention an application developed by the Italian Institute for Social Security (INPS) using an AI-powered system for the classification and automatic sorting of certified e-mails. They also mention the need to develop managerial, systemic, and synergistic competences within public administrations, as well as communicative skills to relate with the general public.

The interviewees emphasized, for example, the need to train citizens how to interact and collaborate with AI (with a focus on AI policies and algorithm-critical literacy) to increase citizens' awareness and perceptions of and engagement with data processing and functioning. Therefore, a progressive and well-thought-out introduction to AI in the public sector seems the best approach.



In conclusion, three main principles inform the social imaginary of public sector managers and designers. Firstly, PA is seen as a “non-competitive context”, therefore innovation should promote public interests, but this approach differs from private companies’ commercial interests, which actually drive the development of public services. This consideration is connected to a shared awareness among experts that the public sector deals with citizens’ trust; therefore, the introduction of AI to the public sector should follow a vigilant and gradual approach.

A second consideration revolves around organizational issues; indeed, several experts mention AI should be conceived as an “organizational agent,” as it modifies the relationship between humans (civil servants and citizens) and computers; this consideration is related to the need for a cultural transformation and a “change management process” to face the organizational changes that AI adoption will produce. Further, experts also claim that AI management requires professionals with design and sociological skills.

The third topic they mention is the need of engaging citizens in public services design, adopting a service design approach based on mapping communities and citizens’ actual needs and behaviors. For this reason, they argue in favor of several principles of human-centered design, such as engaging citizens in co-design processes and collecting feedback on prototypes. Moreover, interviewees emphasized that citizens must be trained to collaborate with AI and receive transparent information about policies, including how specific AI algorithms function.

Indeed, citizens need not only to be aware of when data are collected and for what purposes, but they should also be able to make informed decisions about sharing their data.

## 5. Conclusions

The analysis of European and Italian policies recognizes AI as a transformative force, impacting both the economic sphere and societal challenges. According to public policies, the public sector is a pioneering field, as it can foster a culture of innovation and experimentation. However, the introduction of AI to the public sector presents many challenges in the regulatory and technical domains; moreover, its application must also adhere to several important principles (transparency, accountability, fairness, reliability, inclusiveness, and collaboration).

The analysis of public sector AI experts’ opinions affirmed the importance of considering public service a non-competitive field; therefore, social innovation should be conducted gradually and following public interests. Moreover, they advocate for a human-centered design approach in the implementation of AI to map users’ needs and design useful public services.

We noticed a gap between public policies and AI experts’ social imaginaries. For example, the need for a human-centered design approach emerges in a highly fragmented way, both in the EU and in the Italian public policies. Indeed, this participatory approach is only mentioned as a useful tool to improve collaborative





activities, to ensure respect for human rights, to avoid discrimination and biases from AI systems, and to reinforce citizens' digital skills. Interestingly, Italian public policies claim AI use in the public sector may improve the user experience, but policies do not mention the need to design and implement AI solutions considering citizens' related perceptions, fears, and emotions. Therefore, by adopting Castoriadis's theoretical framework in relation to the use of AI in the public sector, we notice an opposition between the *heteronomous imagination*, the one imposed by public policies, and the *autonomous imagination*, claimed by experts and based on citizens' needs and perceptions. More specifically, public policies focus more on the importance of creating shared standards, as well as the related (technological) infrastructures. Even when they mention the importance of listening to citizens' needs, they primarily adopt a macroscopic gaze. As a result, public policy implementation does not adequately rely on analysis of citizens' specific fears and desires in relation to AI use, and it is not concerned with designing the most suitable methods of interaction. This heteronomous approach tends also to pay less attention to the wider social and cultural contexts related to AI implementation; further, it does not support extended and critical analysis of the issues that AI solutions may contribute to solving, and it also does not support the involvement of stakeholders in the design of AI solutions. Conversely, the autonomous imagination, claimed by experts and young adults (Opromolla et al., 2024), asks for a more tailored approach focused on citizens' needs, including emotional ones, and expectations. As one expert argues, an effective AI solution should be designed according to prevalent approaches and people's perceptions.

*Human-centered design* is one approach increasingly being applied in social research to implement technical and social innovations. More specifically, it provides useful insights for mapping the social impacts of a given solution and the related power dynamics to improve digital interfaces features and to reinforce critical skills for consciously interacting with socio-technical systems. Lupton (2018), for example, uses the expression *design sociology* (and, in particular, the *design through sociology* approach) to refer to the application of design methods and tools in sociology, aiming "to understand people's engagements with objects, systems and services, better engage publics and other stakeholders, work towards social change, and identify and intervene in futures" (pp. 2-3). That is, human-centered design focuses on four key principles: *people centrality*, i.e., the analysis of people's experience within the context in which they act to understand the way they behave, their needs, and what is meaningful to them; *understanding and solving the right problems*, so the main issues and challenges for specific people can be addressed and positive outcomes reached; *everything is a system*, meaning a digital interface is part of a more complex system and is interconnected with other parts; and *small interventions*, necessitating the iterative adaption of the system to peoples' needs (Interaction Design Foundation, 2021). In summary, human-centered design is a participatory approach based on users' engagement for better designs and improved socio-technical systems and for addressing their needs and priorities. Here, we find an important correspondence between the human-centered design approach and the role of the

*autonomous imagination* in shaping a new social imaginary, as mentioned by Castoriadis (1975). Indeed, both Castoriadis and the human-centered design approach consider human creativity a valuable resource for social innovation and transformation, and both advocate for citizens' engagement in imagining new social and institutional orders to develop —from the bottom— innovative solutions often in contrast with the official ones defined by the top.

In conclusion, we believe it is essential to apply a human-centered approach to implement AI into the public sector. Indeed, the use of participatory design methods typical of a human-centered design approach would allow the public sector to understand better and integrate the perspectives of many stakeholders (citizens, public employees, policy makers, etc.) to design appropriate solutions easily; to identify more suitable strategies for raising awareness of AI systems; to improve the service according to feedback; and to identify the social impacts of a specific solution.



## Bibliography

- Agarwal, P. K. (2018), "Public Administration Challenges in the World of AI and Bots", *Public Administration Review*, vol. 78, n. 6: 917-921
- AgID (2018), *Libro Bianco sull'Intelligenza Artificiale al Servizio del Cittadino*
- AgID (2023), *Piano Triennale per l'Informatica nella Pubblica Amministrazione. Edizione 2024-2026*
- Auernhammer J. (2020), "Human-Centered AI: the Role of Human-Centered Design Research in the Development of AI", in Boess S., Cheung M. and Cain R. (eds.), *Synergy - DRS International Conference 2020*
- Barone A. (2020), "Amministrazione del rischio e intelligenza artificiale", in *European Review of Digital Administration & Law - Erdal*, vol. 1, n. 1-2: 63-67
- Bellini V., Biancofiore G.M., Di Noia T., Sciascio E. D., Narducci F., Pomo C. (2020), *GUapp: A Conversational Agent for Job Recommendation for the Italian Public Administration*, IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS) 2020
- Bory S., Bory P. (2016), "New Imaginaries of the Artificial Intelligence", *Im@go. A Journal of the Social Imaginary*, vol. 4, n.6: 66-85
- Castoriadis C. (1975), *L'Institution Imaginaire de la Société*, Paris, Seuil.
- Cavallo Perin R. (2021), "L'amministrazione pubblica con i big data: da Torino un dibattito sull'intelligenza artificiale", in *Quaderni del Dipartimento di Giurisprudenza dell'Università di Torino*, Torino,
- Centorrino M., Di Paola L. (2022), "L'opera d'arte nell'era della sua colonizzazione. I rischi del Metaverso", in *Im@go: A Journal of the Social Imaginary*, vol. 19: 195-209
- Criado J. I., Gil Garcia, J. R. (2019), "Creating public value through smart technologies and strategies: From digital services to artificial intelligence and beyond", in *International Journal of Public Sector Management*, Vol. 32 No. 5, pp. 438-450. <https://doi.org/10.1108/IJPSM-07-2019-0178>.
- Chui M., Harryson M., Manyika J., Roberts R., Chung R., van Heteren A., Nel P. (2018). *Notes from the AI frontier: Applying AI for social good*, McKinsey Global Institute
- Cingolani M., Scendoni R., Fedeli P., Cembrani F. (2023), *Artificial intelligence and digital*



*medicine for integrated home care services in Italy: Opportunities and limits*, Front Public Health. 2023, 10:1095001

D'Albergo E., Fasciani T., Giovanelli G. (2022), *The urban governance of AI: policies and practices of the Italian municipalities*, Midterm Conference ESA Political Sociology 2022

de Certeau M. (1980), *L'Invention du quotidien. Vol. 1, Arts de Faire*, Union générale d'éditions

de Sousa W.G., Pereira de Melo E.R., De Souza Bermejo P.H., Araújo Sousa Farias R., Oliveira Gomes A. (2019), *How and where is artificial intelligence in the public sector going? A literature review and research agenda*, Government Information Quarterly, vol. 36, n. 4, 101392

Di Blasi S. (2013), "Cornelius Castoriadis. L'Istituzione Immaginaria della Società", in *Im@go: A Journal of the Social Imaginary*, vol. 2

Dughiero F., Michieli A., Spiller E., Testa D. (2021), *Governing with urban big data in the smart city environment: an Italian perspective*, IUS PUBLICUM, vol. 1: 1-45

European Commission (2018), *Artificial Intelligence for Europe*, European Commission, Brussels

European Commission (2019), *Ethics Guidelines for Trustworthy AI*, European Commission, Brussels

European Commission (2020a), *A European Strategy for Data*, European Commission, Brussels

European Commission (2020b), *White Paper on Artificial Intelligence - A European Approach to Excellence and Trust*, European Commission, Brussels

European Commission (2021a), *Coordinated Plan on Artificial Intelligence 2021 Review. Technical report*, European Commission, Brussels

European Commission (2021b), *Fostering a European approach to Artificial Intelligence*, European Commission, Brussels

European Commission (2021c), *Impact Assessment*, European Commission, Brussels

European Commission (2021d), REGULATION OF THE EUROPEAN PARLIAMENT AND



OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS, European Commission, Brussels.

European Commission (2022), *Digital Economy and Society Index*, European Commission, Brussels.

European Commission (2024), *Communication on boosting Startups and Innovation in Trustworthy Artificial Intelligence*, European Commission, Brussels

Fabbri L., De Santi M. (2018), *The Designers Italia project - building the community of public services designers*, ServDes2018 - Service Design Proof of Concept

Fanni R., Giancotti F. (2023), *Ethical Artificial Intelligence in the Italian Defence: a Case Study*, DISO vol. 2, n. 29

Fatima S., Desouza K.C., & Dawson G.S. (2020), *National strategic artificial intelligence plans: A multi-dimensional analysis*, Economic Analysis and Policy, vol. 67: 178-194

Federici S. et al. (2021), *A Chatbot Solution for eGLU-Box Pro: The Usability Evaluation Platform for Italian Public Administrations*, Human-Computer Interaction. Theory, Methods and Tools. HCII 2021. Lecture Notes in Computer Science, vol. 12762. Springer, Cham

Forum PA (2024), *Intelligenza artificiale e PA: l'opinione dei dipendenti pubblici*, <https://tinyurl.com/mtvfexxi>

Gahnberg G. (2021), "What rules? Framing the governance of artificial agency", *Policy and Society*, vol. 40, n. 2: 194-210

Galetta D.U. (2023a), *Automation and Algorithmic Decision-Making Systems in the Italian Public Administration*, CERIDAP, vol. 1: 13-23

Galetta D.U. (2023b), *Digital Transition of Public Administration in Italy and the Right to a Good Administration: Problems and Prospects also in the Perspective of the Implementation of Italy's Recovery and Resilience Plan*, Public Law in a Troubled Era, vol. 115: 431-453

Gesk T.S., Leyer M. (2022), *Artificial intelligence in public services: When and why citizens accept its usage*, Government Information Quarterly, vol. 39, n. 3, 101704

Hewitt S. (2009). "Discourse Analysis and Public Policy Research", in *Centre for Rural Economy Discussion Paper Series No. 24*. Centre for Rural Economy, Newcastle, UK



Ingrams, A., Kaufmann, W., & Jacobs, D. (2021). In "AI we trust? Citizen perceptions of AI in government decision making", *Policy & Internet*, 14, 390-409. <https://doi.org/10.1002/poi3.276>

Interaction Design Foundation - IxDF (2021), *What is Human-Centered Design (HCD)?*, Interaction Design Foundation - IxDF

Ishengoma F.R., Shao D., Alexopoulos C., Saxena S. and Nikiforova A. (2022), "Integration of artificial intelligence of things (AIoT) in the public sector: drivers, barriers and future research agenda", in *Digital Policy, Regulation and Governance*, vol. 24, n. 5: 449-462

Jorge Ricart R., Rossetti F., Tangi L. (2022), "AI Watch, National Strategies on Artificial Intelligence: a European Perspective", *Publications Office of the European Union, European Commission, Joint Research Centre*

Khanal S., Zhang H., Taeihagh A. (2024), "Why and how is the power of Big Tech increasing in the policy process? The case of generative AI", *Policy and Society*.

Haesevoets T., Verschuere B., Van Severen R., Roets A. (2024), "How do citizens perceive the use of Artificial Intelligence in public sector decisions?", in *Government Information Quarterly*, vol. 41, n. 1, 101906

Hennink, M. M., Kaiser, B. N., & Marconi, V. C. (2017), "Code saturation versus meaning saturation: How many interviews are enough?", in *Qualitative Health Research*, 27(4), 591-608.

Leighton A. (2019), "Public administration, public leadership and the construction of public value in the age of the algorithm and 'big data'", in *Public Administration*, vol. 97, n.2: 296-310

Loukis E.N., Maragoudakis M. and Kyriakou N. (2020), "Artificial intelligence-based public sector data analytics for economic crisis policymaking", in *Transforming Government: People, Process and Policy*, vol. 14 n. 4:639-662

Lovari A., Ducci G. (2022), *Comunicazione Pubblica. Istituzioni, pratiche, piattaforme*, Milano, Mondadori.

Lupton D. (2018). "Towards Design Sociology", *Sociology Compass*, vol. 12, n.1: 1-11

Manzoni M., Medaglia R., Tangi L., van Noordt C., Vaccari L., Gattwinkel D. (2022), *AI Watch Road to the Adoption of Artificial Intelligence by the Public Sector: A Handbook for Policymakers, Public Administrations and Relevant Stakeholders*, (No. JRC129100),



Joint Research Centre

Maria Cati, M. (2024). The digitization process in the Italian Public Administration: future challenges. IntechOpen. doi: 10.5772/intechopen.1003909.

McDonald III B. D., Hall J., O'Flynn J., Thiel S. V. (2022), "The future of public administration research: An editor's perspective", in *Public Administration*, vol. 100, n.1: 59-71.

Mergel I., Kattel R., Lember V., McBride K. (2018), "Citizen-oriented digital transformation in the public sector", in *Proceedings of the 19th annual international conference on digital government research: Governance in the data age*, 1-3.

Miller J.G. (1978), *Living Systems*, New York, Mc Graw-Hill.

Miller S.M., Song M., Keiser L.R. (2022), "The effect of human versus automated interaction on willingness to participate in government programs: The role of representation", *Public Administration*, vol. 101, n.4: 1349-1366.

Ministero dello Sviluppo Economico (2019), *Proposte per una Strategia italiana per l'Intelligenza Artificiale*.

Ministero per l'Innovazione Tecnologica e la Digitalizzazione (2022), *Programma strategico Intelligenza Artificiale 2022-2024*.

Nzobonimpa S. (2023), "Artificial intelligence, task complexity and uncertainty: analyzing the advantages and disadvantages of using algorithms in public service delivery under public administration theories", in *Digital Transformation and Society*, vol. 2 n. 3: 219-234.

Opromolla A., Parisi L., De Matteis G., Pizolli R., Trodini L. (2024), "Innovare i servizi pubblici attraverso l'intelligenza artificiale. Scenari emergenti tra adolescenti e giovani adulti", *Mediascapes* (under second stage of review).

Radu R. (2021), "Steering the governance of artificial intelligence: national strategies in perspective", *Policy and Society*, vol. 40, n. 2, June 2021: 178-193.

Raffiotta E. (2022), *L'erompere dell'intelligenza artificiale per lo sviluppo della pubblica amministrazione e dei servizi al cittadino*, AI ANTHOLOGY - Profili giuridici, economici e sociali dell'intelligenza artificiale, Bologna, Il Mulino.

Rice D., Siddiki S., Frey S., Kwon J.H., Sawyer A. (2020), "Machine coding of policy texts with the Institutional Grammar", in *Public Administration*, Vol. 99, n. 2: 248-262



Antonio Opromolla, Silvia Di Rocco, Lorenza Parisi  
*AI in the Public Sector: An Exploration of Social Imaginaries*

Russel S., Norvig P. (2021), *Artificial Intelligence: A Modern Approach, Global Edition*, Pearson Education Limited.

Ricoeur P. (2000), *La mémoire, l'histoire, l'oubli*, Paris, Le Seuil.

Qian Sun T., Medaglia R. (2019), "Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare", in *Government Information Quarterly*, vol. 36, n. 2:368-383

Salah M., Abdelfattah F., & Al Halbusi H. (2023), "Generative Artificial Intelligence (ChatGPT & Bard) in Public Administration Research: A Double-Edged Sword for Street-Level Bureaucracy Studies", in *International Journal of Public Administration*, 1-7.

Sariyer G., Mangla S.K., Sozen M.E., Li G., Kazancoglu Y. (2024), "Leveraging explainable artificial intelligence in understanding public transportation usage rates for sustainable development", in *Omega*, vol. 127, 103105.

Schiff D. (2021), "Assessing public value failure in government adoption of artificial intelligence", *Public Administration*, vol. 100, n. 3: 653-673.

Tangi L., van Noordt C., Combetto M., Gattwinkel D., & Pignatelli, F. (2022), *AI Watch. European Landscape on the Use of Artificial Intelligence by the Public Sector* (No. JRC129301), Joint Research Centre.

van Noordt C. & Misuraca G. (2022), "Artificial Intelligence for the Public Sector: Results of Landscaping the Use of AI in Government across the European Union", *Government Information Quarterly*, vol. 39, n.3: 101714.

Vinuesa R., Azizpour H., Leite I. et al. "The Role of Artificial Intelligence in Achieving the Sustainable Development Goals", *Nature Communication*, vol. 11, n. 233.

Wirtz B. W., Langer P. F., & Fenner C. (2021), "Artificial Intelligence in the Public Sector - a Research Agenda", *International Journal of Public Administration*, vol.44, n.13: 1103-1128

Wirtz B. W., Weyerer J. C., Geyer, C. (2019), "Artificial Intelligence and the Public Sector—Applications and Challenges", *International Journal of Public Administration*, vol. 42, n. 7: 596-615

Wirtz B. W., Weyerer J. C., & Sturm, B. J. (2020), "The Dark Sides of Artificial Intelligence: An Integrated AI Governance Framework for Public Administration", *International Journal of Public Administration*, vol. 43, n.9:818-829



Antonio Opromolla, Silvia Di Rocco, Lorenza Parisi  
*AI in the Public Sector: An Exploration of Social Imaginaries*

Ying Tan S., Taeihagh A. (2021), "Governing the adoption of robotics and autonomous systems in long-term care in Singapore", in *Policy and Society*, vol. 40, n. 2: 211–231

Zuiderwijk A., Chen Y. C., Salem, F. (2021), "Implications of the Use of Artificial Intelligence in Public Governance: A Systematic Literature Review and a Research Agenda, in *Government Information Quarterly*, vol. 38, n.3: 101577







A JOURNAL OF THE  
SOCIAL IMAGINARY

# Sociotechnical imaginaries of artificial intelligence in EU law making: a focus on crime and security

Anita Lavorgna

anita.lavorgna@unibo.it

Department of Political and Social Science | University of Bologna



## Abstract

Artificial intelligence (AI), in its various manifestations, has become a core topic in debates concerning crime and security, both because of its potentially criminogenic features, and because of its possibilities as a tool for crime prevention and mitigation. In recent years, critical criminological and criminal justice scholarship has started to pay attention to the sociotechnical imaginaries and the narratives depicting AI especially in the context of crime control, emphasising how prevalent narratives can obfuscate the harms of data-driven technologies and evince idealistic imaginaries of their capabilities. This contribution furthers this line of inquiry, focusing on the AI-adjacent sociotechnical imaginaries with a focus on crime and security in EU law making in the debates and negotiations around the Artificial Intelligence Act (the EU regulation on artificial intelligence).

## Keywords

Crime | Security | Law-Making | Sociotechnical Imaginaries | Ai Act



## 1. Introduction

**A**rtificial Intelligence (AI) is a buzzword that is booming within many industries nowadays. In practice, it can refer to many things (Milivojevic, 2022). For the scope of this contribution, AI can be broadly defined as “a collection of technologies that combine data, algorithms and computing power” (European Commission, 2020: 2) at the basis of systems that display intelligent behaviour by analysing their environment and taking actions with a degree of autonomy to achieve specific goals (European Commission, 2018). These actions can be very complex but still specialized (“weak AI” or “narrow AI”, which is the most successful realization of AI to date), or they could have general applications, which would require an intelligence equal to humans to learn, solve problems, and plan for the future (“strong AI”, which exists today as a theoretical concept).

Even if the presence of a “human in the loop” is mostly needed, AI’s impact in our societies is becoming so significant that its effects are often described as “transformative” (Esposito, 2013; Gruetzemacher and Whittlestone, 2019). This feature is particularly remarkable when we consider that AI has the potential to become increasingly ubiquitous in our lives, offering a number of possibilities for the public good but also important challenges to be met. Nowadays, AI is often used to make certain services to the public or some production processes more apt (e.g., Cioffi et al., 2020; Meyer et al., 2020), but it is also deployed in more sensitive sectors such as – as we will see in detail in the following section – in the crime and security arena (Brownsword and Harel, 2019; Hayward and Maas, 2021). Under the aegis of neoliberalism, the use of AI would be a way to boost efficiency and “smarter” decision-making (Zavrsnik, 2019; Chan 2021).

Overall, it can be claimed that AI is still in a formative stage – both as regards its technological development and its public perception (Richter et al., 2023). Popular imaginaries surrounding AI and, more broadly, the idea of computers able to replicate the human mind, have been around for decades, with the myth of the thinking machine being a focal point of intellectual debate and popular culture since the 1950s. This mythology reignited in more recent times, with the increased availability of computer power and storage, digital data, and advances in areas such as data mining, machine learning, and natural language processing (Natale and Ballatore, 2020, Ballatore and Natale, 2023). AI, as a sociotechnical construct, emerges and grows through language in society (Kajava and Sawhney, 2023).

As such, we cannot ignore the role that media, science fiction and the movie industry play in contributing to the narratives surrounding AI (Bory and Bory, 2015; Sartori and Bocca, 2023), which has important implications on the public understanding and debates regarding this technology (Brause et al., 2023; Chuan, 2023). For instance, research has shown an increase of media coverage of AI in recent years (at least in the Western countries that have been considered by analyses so far), with news being mostly positive and focusing on economic framing (Brause et al., 2023. See also Suffia et al., 2022).



Surprisingly, the development of AI has been happening in a regulatory void (Milivojevic, 2022), except for some forms of self-regulation. As such, the AI Act – i.e., the EU regulation on AI – has been generally welcomed as the first comprehensive effort to regulate AI via a legal framework while trying to establish potentially global standards; it offers common rules and a series of key principles that aspire to be proportionate and flexible, trying to balance the need to mitigate the risks inherent the inappropriate use of AI, with the need to support innovation (European Commission, 2021a,b; Smuha, 2021; Lavorgna and Ugwudike, 2022).

We should not forget that lawmakers and policymakers themselves have different beliefs and perceptions on AI, and these differences impact their choices (Natale, 2019). As such, it is worthwhile to take a closer look at the debates and negotiations around the AI Act, as they present different sociotechnical imaginaries and narratives depicting AI, hence offering both an idea of a desirable future and the way to achieve that future (Hoff, 2023). This contribution focuses on the articulation of sociotechnical imaginaries of AI in relation to the future of crime and security, which is a critical area to investigate because of the major implications AI technologies can have on personal rights and freedoms.



## 2. Artificial intelligence in the crime and security arena

The use of big data in the crime and security arena, and especially of big data analytics (such as sentiment analysis and natural language processing) to manage and “make sense” of large datasets, has received increasingly attention over the past 20 years – both in academia and in the political debate. This is part of the phenomenon of so-called datafication, which is a by-product of the accumulation and digitalization of both large-scale administrative data, and the availability of behavioural and social data including those originating online (Lavorgna and Ugwudike, 2021). The use of AI in this domain is nothing else than the latest step in this direction, with AI systems allowing to automate even more a range of tasks at several levels of the criminal justice process, influencing operational decisions.

AI-enabled system is already used in practice, for instance, to aid decision making in law enforcement, judiciary and penitentiary settings, as well as by intelligence agencies (Zavrsnik, 2019; Hayward and Maas, 2021; Castro-Toledo et al., 2023).

Common examples are the numerous applications in predictive policing, machine evidence and recidivism algorithms, both at pre-investigation stages and within criminal proceedings, with the aim to increase operational efficiency and facilitate decision-making (Sachoulidou, 2023a, b). Consider, for instance, the use of software for facial recognition, to assist emergency call handlers in high stress and time-sensitive environments, in sentencing, or even in military security (e.g., Suchman, 2020; Ryberg, 2024; Wakefield, 2024).

The use of AI and big data analytics in criminal justice has a deep social impact, as it transforms institutional practices and creates new forms of governance (Hannah-Moffat and Avila, 2023). It is changing knowledge production and crime control

policies, at time leading to clashes with established criminal procedure rules (Završnik, 2019). As an unintended consequence, AI applications are also blurring the boundaries in the crime control and security domains: examples are the increased use of preventive measures and state surveillance in anti-terrorism efforts, often in contrast with legal concepts that were developed over many decades as ways to limit the executive power and interferences in personal liberties (Završnik, 2019); or the use of AI systems to control people and their movement, especially when they are deployed along physical borders to check on the mobility of illegalized non-citizens and in the Global North (as discussed in Milivojević, 2022).

Even if AI techniques in the crime and security arena are often promoted as a promising field of study (Dakalbab et al., 2022), with imaginaries of “smart” criminal justice being generally supported by the proponents and developers of the technologies (such as tech companies, but also researchers with a computer science and engineering background – see Lavorgna and Ugwuđike, 2021), critical algorithm studies have been emerging to counteract the “solutionist trend” whereby technical solutions are offered as a panacea to solve social problems (Završnik, 2019).

According to this view, AI-based techniques in the crime and security arena are to be opposed (for an overview of the critiques, see Lavorgna and Ugwuđike, 2021), or it is recognised that they can play a role as new forms of knowledge acquisition but only when used “with a pinch of salt”, for instance to explore certain risks, and without forgetting that their results could be easily misinterpreted if we do not keep in mind the epistemic limits of statistical results (for a detailed discussion on this point, see Castro-Toledo et al., 2023).

More in detail, it is acknowledged that the use of AI poses a number of challenges, and scholars in criminal law and critical criminology have raised concerns on the implications of adopting what are *in nuce* statistical risks assessment tools to determine vulnerability, culpability, and resource allocation (e.g., Hannah-Moffat, 2009; Greenstein, 2021; Ugwuđike, 2021; Bakiner, 2023a; Castro-Toledo et al., 2023; Suchman, 2023). As stressed by Hannah-Moffat and Avila (2023) among others, the current lack of consistent conceptualizations and of epistemologically informed interdisciplinary dialogues are allowing misunderstandings in both the design and use of AI-based tools in the criminal justice systems, where they create several “data harms”.

For instance, AI-based tools would reproduce discriminatory effects, blur distinctions between causation and correlation, and decontextualize and obfuscate sensitive data and intricate legal concepts (Hannah-Moffat & Avila, 2023). Indeed, at the basis of any AI system are learning datasets and coders from which the system “learns” and is further developed; but if a dataset is biased, or otherwise limited, the whole learning and developing process will be problematic.

Moreover, similarly to what was discussed with reference to biometric technology used in criminal justice settings (Franko Aas, 2005, 2006), also with AI tools supposedly objective and unbiased knowledge or solutions would replace human-based decision-making. According to this view, the human (the suspect, or the deciding agent) therefore is not to be fully trusted – a view that can be and has been



harshly criticized, and that can lead to fundamental changes in how people approach their work, leading to new types of unintended errors and accountability issues, while directly hindering a number of constitutional rights (see Završnik, 2019).

AI is also criticized for being a “black box”, whose obscure, elusive and increasingly complex mechanisms would be impenetrable to outsiders (O’Neill, 2016; Katz 2020).

This aspect is particularly problematic if we consider that reportedly AI studies suffer from a major replication crisis, where many sensational claims cannot be properly tested<sup>1</sup>, and from underperformance of the technology, whose outcomes are not close to what is initially promised (Milivojevic, 2022).

As such, recent literature advocates for the adoption of specific procedural safeguarding when using AI-based systems in criminal law and justice, which are deemed necessary to uphold rule of law requirements, and particularly transparency, fairness, and explainability (Castro-Toledo et al., 2023; Sachoulidou, 2023a). We should not forget that, despite its exponential growth in application, AI is still a relatively new technology in the crime and security arena. This leaves open an important window to take actions to minimize AI-related societal risks, and to independently evaluate the impact of introducing AI systems in the sensitive domains (Chan, 2021; Sanders and Chan, 2023).



### 3. A closer look into sociotechnical imaginaries

173

As discussed in Jasanoff’s (2015: 3) seminal work, even if in common parlance “the word ‘technology’ tends to be equated with machine or invention, something solid, engineered, black-boxed” [...], technologies as we know them “would never have taken to the roads without the myriad social roles, institutions, and practices spawned by modernity: scientists, engineers, and designers; patents and trademarks; autoworkers and big corporations; regulators; dealers and distributors; advertising companies; and users [...]” – a perfect reminder that technologies are an integral part of our societies, and exist within the social experience. As such, how technological visions enter the social life matters. For instance, in the case of AI, these visions determine the narratives surrounding hopes and fears around AI, influencing how AI is developed, deployed, and regulated (Dave and Dihal, 2019).

The notion of sociotechnical imaginaries is very useful to further unpack this point. Again, borrowing Jasanoff’s words, sociotechnical imaginaries are “collectively held, institutionally stabilized, and publicly performed visions of desirable futures, animated by shared understandings of forms of social life and social order” (Jasanoff, 2015: 5-6). They are “durable”, but also “temporally situated and culturally particular”, and as the name suggests can be interpreted as both “products of and instruments of the co-production of science, technology, and society in modernity” (Jasanoff, 2015: 5-6). They can originate in the visions of single individuals or groups, and then

<sup>1</sup> Reportedly, only 15 percent of relevant studies share their algorithms with experts and the wider community (Horgan, 2020).



can be articulated and propagated by others, becoming collectively held (ibidem). While sociotechnical imaginaries are “typically grounded in positive visions of social progress”, they also correlate “with the obverse shared fears of harms that might be incurred through invention and innovation, or of course the failure to innovate” (ibidem – see also Sturken & Thomas, 2004). In other words, they are acts of *futureing*, as they allow the identification, creation and dissemination of images of the future and as such they shape the possible space for action (Oomen et al., 2021: 253; see also Schulz, 2015). The power of imagination can indeed influence how and which futures become performative (Cantó-Milà & Seebach, 2015; Hoff, 2023)<sup>2</sup>.

Importantly, within the same society, multiple collective imaginaries can coexist (Kling & Iacono, 1990; Iacono & Kling, 1996; Pellegrino, 2015), even if it is worth noting that digital and mediatic capitals can accord some imaginaries a dominant position (Jasanoff, 2015; Lavorgna and Ugwudike, 2021). While it is indeed important to recognise that imaginaries can originate also bottom-up, it should not be surprising that collective imagination is often guided by visible actors with the authority to set policy agendas, as can be observed in policy discourses, which can be suitably mined for insights into specific sociotechnical imaginaries (Jasanoff, 2015: 37).

These seminal reflections were first developed with reference to other technologies than AI, but they can still help us to focus the issue of future making in a broader theoretical context. More specifically in the context of AI, corporate sociotechnical imaginaries – framed and reinforced by tech companies – have become increasingly dominant over the past few years, offering specific ideas of technological progress and, overall, a utopian vision of AI futures (Mager and Katzenbach, 2021; Hoff, 2023; Sartori and Bocca, 2023). Legitimation, promotion and reassurance are all tactics used to try and evoke affective meanings of attractive sociotechnical imaginaries (Hoff, 2023). At times, in these imaginaries we can find the transhumanist idea to enhance the human condition with the use of advanced technologies, including via interactions between individuals and machines that would increase collective intelligence to unprecedented levels, bringing humans to new levels of consciousness (for an overview, see Natale and Ballatore, 2020). In practice, however, there is often a gap in between sociotechnical imaginaries and the situated reality of technology-human interaction, and what it allows (Sanders and Chan, 2023).

Also, in criminal justice settings there are competing sociotechnical imaginaries (e.g., Chan, 2021), which directly affect practices. Lavorgna and Ugwudike (2021) systematically reviewed and compared multidisciplinary academic abstracts on the data-driven tools (including AI tools) shaping (or with the clear potential to shape) decision-making across several justice systems. In their work, three main frames – optimistic, neutral and oppositional – for understanding how relevant technologies are portrayed were identified, each evoking different imaginary. More in detail,

---

<sup>2</sup> To come about and become performative in practice, imagined futures need to be credible (if certain futures are seen as credible, people may act upon them in the present), and convey affective meanings (Hoff, 2023).



*optimistic frames* (by far, the most prominent and frequent frames) evinced utopic sociotechnical imaginaries, presenting AI models as the future of criminal justice because of their scientific objectivity, technological superiority and vital utility. *Neutral frames*, on the other hand, evoked a technorealism vacillating between utopic and dystopic imaginaries. These frames were often nuanced in envisioning sociotechnical imaginaries, for instance recognising how AI systems could also generate negative externalities; or emphasising the need for more and better empirical research on technology adoption in real-world practice, and the improvement of datasets and codes to overcome current limitations. Finally, *oppositional frames* rejected the technochauvinism (Broussard, 2018) of the sociotechnical imaginaries emphasising technological potentialities, contesting the optimistic visions glossing over potential harmful effects (Chan, 2021), and rather drawing attention to the social impacts including problems stemming from data injustices and the need for robust governance<sup>3</sup>. Interestingly, while oppositional frames were proposed by authors with more subject knowledge of criminal justice systems and sociolegal theorisations (intellectual and critical capital), optimistic frames were generally supported by researchers with a more technical background (digital capital) (Lavorgna and Ugwudike, 2021)<sup>4</sup>.



#### 4. This contribution

175

As recognised by Jasanoff herself (2015: 14), there are important mechanisms of interconnections between technoscientific and political practices of collective imaginaries that have not been fully articulated yet. In AI-adjacent research, in particular, the relative lack of attention over these mechanisms is causing some critical research gaps<sup>5</sup>. For instance, considering how lawmakers and other stakeholders envision the potential benefits and challenges arising from AI is of great importance, as the imaginaries they depict have a unique potential not only to determine AI's possible trajectories, but also its integration into society. This article offers a contribution to this line of inquiry, focusing on crime and security in EU law making in the debates and negotiations around the AI Act.

Both at the EU and international levels there have been, and still are, numerous debates on the compatibility of AI-based systems with human rights and the rule of law, especially in sensitive context such as the crime and security arena. Consider, for instance, the 2018 European Ethical Charter on the use of AI in judicial systems and

<sup>3</sup> The key oppositional frames were as follows: privacy violations, racial bias, algorithmic opacity, fairness concerns, inadequate legislation, the commercialisation of datafied criminal justice and the presence of systemic bias (Lavorgna and Ugwudike, 2021).

<sup>4</sup> Conversely, for some disciplines and practitioners (policymakers, law enforcement officers and other practitioners that use or rely on data-driven technologies) alike it is difficult to entertain a serious debate on the use of AI in crime and security because of their lack of a deep epistemological understanding of how these systems work (Lavorgna and Ugwudike, 2021; Castro-Toledo et al., 2023).

<sup>5</sup> Critical because of the formative stage of AI as discussed earlier (Richter et al., 2023), meaning also that policies, practices and standards have not crystallized yet.



their environments (JEPEC, 2018), the 2020 feasibility study on the possible introduction of a mechanism for certifying AI tools and services in the sphere of justice and the judiciary (CEPEJ, 2020) – both born under the aegis of the Council of Europe; or the European Commission Communication on Building Trust in Human-Centric AI and the European Commission White Paper on AI (European Commission, 2019, 2020; see also Sachoulidou, 2023b).

The AI Act represents the latest and more influential effort in this regard, with the EU pursuing a position on global leader on ethical AI regulation (Krarup & Horst, 2023) and recognizing that a lack of standardization in this arena would create a significant barrier for the development and deployment of AI systems (Goodman, 2023; Laux et al., 2024). Interestingly, the Act does not regulate AI as such; rather, it regulates the entry into the market and the use of technological systems that rely on AI, with the aim to maintain neutrality towards the technology itself, and not to risk to over rely on an obsolescent definition of AI (European Commission, 2021a; Lavorgna and Ugwudike, 2022).

Before moving to the core of this contribution, it is worthwhile to present a brief overview of the timeline of the AI Act, highlighting some key events. In October 2020, with EU leaders discussing the digital transition, the European Council (2020) invited the Commission to (among other things) propose ways to increase investments in Europe in AI research, innovation and deployment, and provide a clear, objective definition of high-risk AI systems. On the 21st of April 2021, the European Commission released a proposal for a regulation aiming to harmonise rules on AI systems (AI Act), ensuring legal certainty and facilitating the development of a single market for AI applications, while accelerating investments in AI in Europe (European Commission, 2021b). As part of its standardization effort, the proposal follows a risk-based approach, differentiating between uses of AI systems depending on the severity of risk to the health and safety or fundamental rights of natural persons they create or might create. Four different levels (unacceptable risk, high risk, limited risk, and minimal risk) are identified. As regards the crime and security arena, this risk-based approach is of much interest. Consider, for instance, that among the AI systems that can create an unacceptable risk and are therefore generally forbidden are those systems using “real time” remote biometric identification in publicly accessible spaces; but several exceptions are in place for law enforcement purposes, which could undermine the safeguards in place (see Lavorgna and Suffia, 2021).

On 6th December 2022, the Council agreed on its position on the AI act, aiming to ensure that AI systems placed on/used in the EU are safe and respect existing law on fundamental rights (European Council, 2022). After months of negotiation and following 3-day “marathon talks”, the Council and the European Parliament finally reached a provisional agreement on the AI Act on 9 December 2023 (European Council, 2023). The final text was approved on March 13, 2024, by the European Parliament, and – awaiting the Council’s position – it is expected to apply from 2026.

In a recent study using the EU AI Act (in its draft form proposed by the European Commission in April 2021) as a case study, Bakiner (2023b) already noted the existence of pluralistic sociotechnical imaginaries which differ regarding the



characteristics of AI as it relates to society, politics and law; and whether and to what extent the law can/should influence legal regulation or intervene into expert debates in AI<sup>6</sup>. Differently yet relatedly, this contribution focuses only on sociotechnical imaginaries of AI in relation to crime and security as they are presented by the EU's main law-making actors and bodies: the Parliament (with its diverse members) and the Council, which jointly exercise the legislative function, and the Commission in its role of proponent of EU legislation.

Through the procedure files section of the Legislative Observatory (European Parliament)'s website<sup>7</sup> and the Council's website, it was possible to access some key documents in this regard, and specifically:

- the initial legislative proposal by the Commission (21 April 2021);
- the "general approach" by the Council (25 November 2022);
- a joint report by the European Parliament's Committee on the Internal Market and Consumer Protection and the Committee on Civil Liberties, Justice and Home Affairs (22 May 2023);
- two decisions by Parliament (14 June 2023 and 13 March 2024, plus a corrigendum dated 19/4/2024);
- two plenary debates of the Parliament (13 June 2023 and 12 March 2024 – videos and verbatim reports).

Once retrieved, all these documents (in their English version; interventions in other languages have been translated into English with Google Translate) were manually coded and thematically analysed in chronological order, to unpack not only the key themes relating to sociotechnical imaginaries, but also whether they evolved during the legislative process. A balanced approach of inductive and deductive coding (Fereday & Muir-Cochrane, 2006) was used, focusing on: AI advantages; AI negatives and risks; neutrality; AI uses; cases and examples; priorities; futures; emotions; hypes; political parties; political opposition and cleavage. Overall, from the analysis, the documents offered two main diverse layers of interpretation, as a clear divide emerged between the themes and tones unfolding from the discussion *among* institutions, and those unfolding from the debates *within* the Parliament. The first layer is the formal position of the different legislative bodies and the corresponding dominant imaginaries; the second layer, emerging from the Parliamentary debates, reflects contrasting political visions of AI-adjacent risks and opportunities, and of desirable futures.

---

<sup>6</sup> Through an analysis of 302 response papers submitted by developers and deployers of AI, other businesses, civil society groups, academics, concerned citizens and public authorities based in the EU and beyond in 2021 during the third round of invited feedback. There were two initial rounds of invited feedback (multiple-choice surveys and open comments) in 2020, followed by a third round (response papers) that took place immediately after the European Commission published its proposed Act. Among other things, for instance, Bakiner noted how many businesses, and a few NGOs and academics seek clearer definitions and regulatory minimalism in this area; while some NGOs and academics focus AI's potential harms and argue for the need of bolder legal interventions as a desirable constrain to rein in the worst excess of otherwise useful technologies.

<sup>7</sup> See: [https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2021/0106\(COD\)&l=en](https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?reference=2021/0106(COD)&l=en).



## 5. The first layer: the Council vs Parliament chasm

Even if the AI Act proposal and its subsequent versions attempt to recognise both the advantages and the challenges offered by AI, and to introduce a technology-neutral definition of AI systems, when it comes to crime and security matters a chasm between the Council on the one side, and the Parliament on the other, emerges<sup>8</sup>. Surveillance in public spaces, and specifically the use of remote biometric identification systems for the purposes of law enforcement, is one area of clear difference.

The initial proposal by the Commission, for instance, generally forbids the use of such systems as they are considered particularly intrusive (Whereas 18), while envisaging some specific safeguards to “exhaustively listed and narrowly defined situations” such as the search for potential crime victims including missing children, terrorism, and certain forms of serious crimes (Whereas 19. See also art. 5), with “appropriate limits in time and space” (Whereas 20) and if “subject to an express and specific authorization by a judicial authority or by an independent administrative authority of a Member State” (Whereas 21).

The Council, inter alia, agrees to allow biometric identification systems in publicly accessible spaces for law enforcement purposes, further clarifying in its “general approach” when they can be used. Further, it adds an explicit clarification to the fact that the AI Act (and its safeguards) does not apply to security, defence and military purposes (art.2).

Conversely the Parliament, in line with the position advocated by many civil society organizations, tends to be more restrictive: it wants to ban the use of all biometric identification systems for both real time and *post facto* use, with very limited exceptions only to the *post facto* use when there is a need linked to a specific serious criminal case, and pending a specific judicial authorization (e.g., amendment 33 of 13 March 2024, and the Corrigendum). Also, all biometric categorization systems using sensitive characteristics such as gender, race, ethnicity, religion, political orientation or citizenship status are to be banned.

Overall, the Commission and the Council seem to rely on neutral or even optimistic frames when it comes to the potential of AI on crime and security matters, depicting the AI systems as important criminal justice tools to enhance law enforcement capabilities (vital utility, see Lavorgna and Ugwu-dike, 2021). The Parliament, on the other hand, uses oppositional frames alongside neutral ones, drawing attention to the social impacts of AI, including problems stemming from data injustices and the need for robust governance.

Similarly, the initial proposal by the Commission recognises some intrinsic limitations of AI systems when applied to the crime and security arena, where they could lead to biased results and entail discriminatory effects (see e.g. Whereas 33, 39 and 40 with reference respectively to remote biometric identification, AI systems



<sup>8</sup> The AI Act was assigned jointly to the European Parliament’s civil liberties (LIBE) and internal market (IMCO) committees.

used in migration, asylum and border control management, and AI systems intended for the administration of justice). These limitations, however, are generally linked to issues in the development and training of a given AI system (Whereas 38); hence, for instance, high risk AI systems can be employed in crime and security matters if they are “designed and developed” with “an appropriate level of accuracy, robustness and cybersecurity [...] throughout their lifecycle”, and potential biases are addressed with “appropriate mitigation measures” (art.15). This approach suggests the reliance on neutral – solutionist – frames (Lavorgna and Ugwuđike, 2021): these frames recognise potential barriers to efficient algorithmic outputs, but embrace a vision of future benefits and crime control potential stemming from technological innovation, and fair and accurate algorithmic predictions, once technical problems are somehow fixed or optimized.

If the Council's position is in line with this vision, the Parliament pushes for banning inter alia untargeted scraping from CCTV footage or the internet to create facial recognition databases, and predictive policing when it is based solely on profiling, location or past criminal behaviour (see e.g. amendments 50 and 51, Decision 14 June 2023). Indeed, all these AI applications are considered a threat to citizens' rights, in line with the oppositional frames linked to fairness concerns (Lavorgna and Ugwuđike, 2021), which recognize how the tendency to focus on technical solutions to fairness issues overlooks the socioeconomic and sociopolitical contexts that shape technological design, application and outcomes (see also Selbst et al., 2018).



## 6. The second layer: sociotechnical imaginaries of AI and socio-political polarization

As anticipated, a second layer of interpretation emerges from the Parliamentary debates, where contrasting political visions of AI-adjacent risks and opportunities arise. Here, again, a chasm separating sociotechnical oppositional/neutral imaginaries vs positive/neutral imaginaries can be identified. There is indeed a cultural line dividing differing social priorities and desirable futures; as this line becomes politicized, impacting directly the political debate, the observed chasm seems to reflect a new political cleavage (Dalton, 1996), reflecting socio-political polarization on this issue (in line with Jacobs, 2024; Parinandi et al., 2024).

The minutes of the 2023 Parliamentary debate on the AI Act offer several examples pertaining specifically to the crime and security arena. All the members of the European Parliament intervening in the plenary session somehow recognise the complexity of legislating on AI-adjacent matters and the need to embrace nuanced views (e.g., “It is about maximising the potential of AI, while at the same time addressing the existing and looming concerns<sup>9</sup>”; “We must put a clear stop to such

<sup>9</sup> Eva Maydell, rapporteur for the opinion of the Committee on Industry, Research and Energy, 13 June 2023.



dangers and at the same time promote innovation in Europe<sup>10</sup>), and some explicitly underline how this is a task that deeply exceeds the technical matter (e.g., “It is about asking ourselves the big question: what kind of society do we actually want to live in?<sup>11</sup>”; “The rapid rise of AI requires us to make fundamental choices about our future<sup>12</sup>”).

However, those belonging to the centre-right (such as the European People's Party) political groups often refer to the need to use AI systems to support law enforcement to counter serious forms of crimes and terrorist activities (optimistic frame of vital utility, see Lavorgna and Ugwu-dike, 2021), as exemplified by the following snippets:

When a child goes missing, every second counts. Seventy-six per cent of abducted children that tragically do not survive the abduction die within the first three hours of their disappearance. Every second counts. And we need to make use of all the instruments at our disposal to prevent such tragedies from happening<sup>13</sup>;

I would therefore like to urge my colleagues to vote yes for the police to be able to use AI technology in real time to stop acts of terrorism, to stop kidnappings of children. We must protect our citizens, not terrorists.<sup>14</sup>

To this end, the “fear-driven suggestions” by their political counterparts are explicitly rejected (e.g., “Several left-wing political groups refuse to use artificial intelligence, even though the point here is to protect children, arrest criminals and protect us from terrorism<sup>15</sup>”; “It is not wise legislation to prevent the possibility of using AI to increase security in Europe, where the left in the European Parliament actively wants to prevent the use of AI during, for example, terrorist crimes in progress or while a child is kidnapped [...] [or] to increase border patrols and ensure that we can better control migration to Europe<sup>16</sup>”).

Aligned with the claims made about the vital utility of AI systems in the crime and security context, optimistic frames depict AI in line with the optimistic frame of technological superiority, hence configuring a specific idea of an advisable social future and setting expectations associated with it (Lavorgna and Ugwu-dike, 2021). From this perspective, AI is expected to “make our societies more prosperous and our economies more efficient<sup>17</sup>”. As one member of the Parliament put it,

<sup>10</sup> Thierry Breton, member of the Commission, 13 June 2023.

<sup>11</sup> Eva Maydell, rapporteur for the opinion of the Committee on Industry, Research and Energy, 13 June 2023.

<sup>12</sup> Kim Van Sparrentak (Verts/ALE-Fractie), 13 June 2023.

<sup>13</sup> Jeroen Lenaers (PPE), 13 June 2023.

<sup>14</sup> Arba Kokalari (PPE), 13 June 2023.

<sup>15</sup> Geoffroy Didier (PPE), 13 June 2023.

<sup>16</sup> Sara Skytvedal (PPE), 13 June 2023.

<sup>17</sup> Jeroen Lenaers (PPE), 13 June 2023.



I do not think AI is a risk, but a chance: a chance for a better future [...] Just because I can hit my finger with a hammer does not mean we should ban all of them and stop building houses. [...] Banning innovation goes far too far. No technology is without risk. [...] I believe AI, like other technologies, will make the future better, and I want it to develop in Europe.<sup>18</sup>

Overall, the possibilities offered by AI systems are connected to sociotechnical imaginaries that can be considered utopic for those with a certain attitude towards law and order, but that also express specific social anxieties and fears; technoscientific innovations, from this perspective, encode “visions of the good society” (Jasanoff et al., 2007: 6) in crime and security matters, but the definition itself of what is “the good society” we aim for has to do with judgments of (symbolic) social stability, consensus and control.

Conversely, the risks emphasised by those linked to the left and the centre-left of the political spectrum (such as those from the Progressive Alliance of Socialists and Democrats) and those affiliated with the Greens are of a completely different nature. Here, the risks identified are those linked to the development and use of AI itself, ranging from general and not better identified fears (“the risk of extinction of the human species because of sentient algorithms<sup>19</sup>”) to (more frequently) concerns intrinsically linked to the nature of AI systems themselves. As such, alongside risks linked to discrimination, disinformation, or to the creation of illegal content, members of the Parliament stress the potential for abuse, surveillance, predictive policing, and for “perfecting human rights violations at the EU’s external borders<sup>20</sup>” – hence relying on several of the oppositional frames identified in Lavorgna and Ugwudike (2021), such as privacy violations, racial bias, fairness concerns, and systemic bias.

Additionally, and in complete antithesis of the optimistic frames, AI systems in the crime and security arena are attacked as an “inaccurate”, “flawed technology” leading to ineffective results, as exemplified below:

Colleagues from EPP will try to make the case how these systems could be used to catch criminals. But one thing is very clear to me: those who want to hide from AI systems will always find a way to do so, and always the innocent will suffer;<sup>21</sup>

Not a single terrorist has been found with biometric mass surveillance, not a single attack has been prevented, but instead countless arrests of innocent people have been made [...] They open Pandora’s Box and lead us into a dystopian future of a suspicious high-tech Chinese-style surveillance state. They give authoritarian governments of the present and the future an unprecedented weapon of oppression.<sup>22</sup>

<sup>18</sup> Dita Charanzová (Renew), 13 June 2023.

<sup>19</sup> Brando Benifei, Rapporteur, 13 June 2023.

<sup>20</sup> Cornelia Ernst (The Left), 13 June 2023.

<sup>21</sup> Petar Vitanov S&D), 13 June 2023.

<sup>22</sup> Patrick Breyer (Verts/ALE), 13 June 2023.



Unsurprisingly, these are the political groups pushing in the debate for the complete ban of AI systems such as those for the remote biometric identification systems in public spaces, or for emotion recognition, for the purposes of law enforcement, which are seen as “unethical<sup>23</sup>” and “intrusive pseudoscientific<sup>24</sup>” practices.

Lastly, the centrist political parties (such as Renew) seem to adopt neutral frames (ambivalent and solutionist) in commenting on the Act, recognising the benefits that AI can bring, but also its potential negative externalities (“We do not condemn AI itself, but rather risky practices<sup>25</sup>”).

Similar visions were found also in the minutes of the 2024 Parliamentary debate, with the core difference that in this latter case the comments were not on a developing piece of legislation, but rather on its finalized version, taking pride or assigning blame on the result. Again, regardless of the agreement on the Act itself, a clear divide persists between the different political groups. On the one hand, for instance, a member of the right-wing/far-right political group Identity and Democracy comments:

I am delighted that our group has been able [...] to contribute to this text with proposals such [...] the use of artificial intelligence by our law enforcement to search for victims of kidnappings, investigate risks of Islamist attacks or fight against child pornography. I also welcome the rejection of the scandalous proposals from the left. I am thinking, for example, of their desire to prohibit the anticipation of illegal immigrants crossing borders [...].<sup>26</sup>

If many seem to agree with the fact that the latest version of the AI Act can be overall considered a fine compromise, reached thanks to meliorative amendments by the Parliament (e.g., “I can say that we have improved the Commission's proposal. But I would have liked more innovation and even stronger protection of civil rights<sup>27</sup>”), some strongly reject the agreed text because of all the exceptions legitimizing the use of otherwise forbidden or high-risk AI systems for crime and security matters (e.g., “Biometric tracking is allowed by the current text, and the profits of the richest have won out over security<sup>28</sup>”; “All the exceptions introduced by the Council open the door to hardcore authoritarianism<sup>29</sup>”), to the point of considering AI itself as a major security threat (“Any innovation must respect the basic rights of our citizens [...] and must not become a tool for manipulating people's privacy and thus a security threat<sup>30</sup>”).

<sup>23</sup> Włodzimierz Cimoszewicz (S&D), 13 June 2023.

<sup>24</sup> Karen Melchior (Renew), 13 June 2023.

<sup>25</sup> Dita Charanzová (Renew), 13 June 2023.

<sup>26</sup> Jean-Lin Lacapelle (ID), 12 March 2024.

<sup>27</sup> Svenja Hahn (Renew-Fraktion), 12 March 2024.

<sup>28</sup> Kateřina Konečná (The Left), 12 March 2024.

<sup>29</sup> Clara Ponsatí Obiols (NI), 12 March 2024.

<sup>30</sup> Robert Hajšel (S&D), 12 March 2024.



## 7. Conclusion

AI is a phenomenon that nowadays is at the core of public debate, being not only “the flavour of the month” but also a potentially disruptive technology and an accelerator of social change. Its imaginaries can be contrasting, varying from transhumanist techno-visions to technophobic dystopias, and all the greys in between. Together, these imaginaries contribute to creating or changing attitudes around AI, influencing policies and practices. In any case, regardless of whether AI is (or is perceived) as effective or flawed, undeniably it is a technology with profound consequences.

We have seen how competing sociotechnical imaginaries of AI occur also in lawmaking, as the actors involved can foster different beliefs and perception of AI (possibly because of political ideology, lobbying influences, or personal experience). These pluralistic imaginaries are likely to contribute to shaping the spirit and the wording of the law, and the future of AI along with it (Bakiner, 2023b). While the finalized Act, in its current form, seems to prefer neutral, ambivalent frames of sociotechnical imaginaries of AI, by looking at the chronological evolution of the Act and at the diverse visions of some of the actors involved also positive and oppositional imaginaries emerge, positioned alongside clear-cut divides according to different social (security or technology related) anxieties.

Is there a way to bridge these divides, fostering the possibility to further improve the political debate, and to better navigate the possibility to steer AI development and the actualization of its potential in a way that can ultimately increase human wellbeing<sup>31</sup>?

We know that, beyond political identity, our education and specifically our diverse disciplinary backgrounds help explaining differences in our sociotechnical imaginaries (Lavorgna and Ugwudike, 2021), mostly because of the different access to critical and digital capital they allow. To overcome myth and beliefs on AI systems – both as regards their potentially criminogenic or otherwise harming nature, and the tremendous opportunities they might offer –, therefore, we need to improve AI literacy (Chuan, 2023; Rahm, 2023) as well as interdisciplinary efforts starting with the educational sector (see also Sachoulidou, 2023b). The development of critical approaches and solutions about AI education is needed not only to impact visions of AI possible futures on political elites, fostering a more nuanced, inclusive and context-specific approach to risk identification and mitigation (Cantero Gamito & Marsden, 2024), but most importantly on their constituencies, curbing current issues of epistemic injustice (Rahm, 2023).

These interdisciplinary efforts are needed also in research and development settings where traditionally a small number of scientific disciplines framed the discussion around AI in a certain direction thanks to their digital capital and, at times, thanks to corporate support, but probably lacked some of the critical expertise traditionally residing in the social and political sciences (Lavorgna and Ugwudike,



<sup>31</sup> On this, see also European Commission, 2019; Završnik, 2019.

2021)<sup>32</sup>. On the other hand, these latter disciplines – congenitally more cautious in making claims about the future and in shaping political debates and business practices (Halford and Southerton, 2023) – can bring to the table crucial resources and multidimensional approaches to ideate visions of desirable futures, or to alert future-making actors on current and emerging societal risks.



---

<sup>32</sup> For instance, as underlined by Sartori and Bocca (2023), within the AI community, focusing on the social side of AI is still a relatively novel requirement, and it is still largely limited to discussions on ethics, explainability and responsible AI. On the importance of relying on a sociotechnical approach, see also – among others – Sartori and Theodorou (2022) and Halford and Southerton (2023).



## Bibliography

- Bakiner, O. (2023a), "The promises and challenges of addressing artificial intelligence with human rights", *Big Data & Society*, 10(2).
- Bakiner, O. (2023b), "Pluralistic sociotechnical imaginaries in Artificial Intelligence (AI) law: the case of the European Union's AI Act", *Law, Innovation and Technology*, 15(2): 558-582.
- Ballatore, A., Natale, S. (2023), "Technological failures, controversies and the myth of AI", in Lindgren, S. (ed.), *Handbook of Critical Studies of Artificial Intelligence*, Northampton, MA, Edward Elgar Publishing, 237-244.
- Bory, S., Bory, P. (2015), "New imaginaries of the artificial intelligence", *Im@go. A Journal of the Social Imaginary*, (6): 66-85.
- Brause, S. R., Zeng, J., Schäfer, M. S., Katzenbach, C. (2023), "Media representations of artificial intelligence: surveying the field", in Lindgren, S. (ed.), *Handbook of Critical Studies of Artificial Intelligence*, Northampton, MA, Edward Elgar Publishing, 277-288.
- Broussard, M. (2018), *Artificial unintelligence: How computers misunderstand the world*, Boston, MIT Press.
- Brownsword, R., Harel, A. (2019), "Law, liberty and technology: Criminal justice in the context of smart machines", *International Journal of Law in Context*, 15: 107-125.
- Cantero Gamito, M., Marsden, C. T., (2024), "Artificial intelligence co-regulation? The role of standards in the EU AI Act", *International Journal of Law and Information Technology*, 32(1): eaae011.
- Cantó-Milà, N., Seebach, S. (2015), "Desired images, regulating figures, constructed imaginaries: The future as an apriority for society to be possible", *Current Sociology*, 63(2): 198-215.
- Castro-Toledo F. J., Miró-Llinares F., Aguerri J. C. (2023), "Data-Driven Criminal Justice in the age of algorithms: epistemic challenges and practical implications", *Criminal Law Forum*, 34(3).
- Cave, S., Dihal, K., Dillon, S. (eds.) (2020), *AI narratives: a history of imaginative thinking about intelligent machines*, Oxford, Oxford University Press.
- CEPEJ (2020), *Possible introduction of a mechanism for certifying artificial intelligence tools and services in the sphere of justice and the judiciary: Feasibility Study*, Council



of Europe. Retrieved on 29 April 2024 from <https://rm.coe.int/feasability-study-en-cepej-2020-15/1680a0adf4>.

Chan, J. (2021). "The future of AI in policing: Exploring the sociotechnical imaginaries", in McDaniel J., Pease K. (eds.), *Predictive policing and artificial intelligence*, Milton Park, Routledge, 41-57.

Chuan, C. H. (2023), "A critical review of news framing of artificial intelligence", in Lindgren, S. (ed.), *Handbook of Critical Studies of Artificial Intelligence*, Northampton, MA, Edward Elgar Publishing, 266-276.

Cioffi, R., Travaglioni, M., Piscitelli, G., Petrillo, A., De Felice, F. (2020), "Artificial Intelligence and Machine Learning Applications in Smart Production: Progress, Trends, and Directions", *Sustainability*, 12: 492.

Dakalbab, F., Talib, M. A., Waraga, O. A., Nassif, A. B., Abbas, S., & Nasir, Q. (2022), "Artificial intelligence & crime prediction: A systematic literature review", *Social Sciences & Humanities Open*, 6(1): 100342.

Dalton, R. J. (1996), "Political cleavages, issues, and electoral change", *Comparing Democracies: Elections and voting in global perspective*, 2: 320-321.

European Commission (2019), *Communication: Building Trust in Human Centric Artificial Intelligence*. Retrieved on 2 May 2024 from <https://digital-strategy.ec.europa.eu/en/library/communication-building-trust-human-centric-artificial-intelligence>.

European Commission (2020), *White Paper on Artificial Intelligence – A European Approach to Excellence and Trust*. Retrieved on 2 May 2024 from [https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020\\_en.pdf](https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020_en.pdf).

European Commission (2021a), *A European approach to AI*. Retrieved on 2 May 2024 from <https://digitalstrategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>.

European Commission (2021b), *Communication on Fostering a European approach to Artificial Intelligence*. Retrieved on 2 May 2024 from <https://digitalstrategy.ec.europa.eu/en/library/communication-fostering-european-approach-artificialintelligence>.

European Council (2020), *Conclusions*. Retrieved on 2 May 2024 from <https://www.consilium.europa.eu/en/press/press-releases/2020/10/02/european-council-conclusions-1-2-october-2020/>.



European Council (2022), *Artificial Intelligence Act: Council calls for promoting safe AI that respects fundamental rights*. Retrieved on 2 May 2024 from <https://www.consilium.europa.eu/en/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/>.

European Council (2023), *Artificial intelligence act: Council and Parliament strike a deal on the first rules for AI in the world*. Retrieved on 2 May 2024 from <https://www.consilium.europa.eu/en/press/press-releases/2023/12/09/artificial-intelligence-act-council-and-parliament-strike-a-deal-on-the-first-worldwide-rules-for-ai/>.

Esposito, E. (2013), "Digital prophecies and web intelligence", in Hildebrandt, M., De Vries, K. (eds.) *Privacy, Due Process and the Computational Turn*, Milton Park, Routledge, 121-142.

Fereday, J., Muir-Cochrane, E. (2006), "Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development", *International Journal of Qualitative Methods*, 5(1): 80-92.

Franko Aas, K. (2005), *Sentencing in the Age of Information: From Faust to Macintosh*. London: Routledge-Cavendish.

Franko Aas, K. (2006), "'The body does not lie': Identity, risk and trust in technoculture", *Crime, Media, Culture*, 2(2): 143-158.

Goodman, T. (2023), "Thinking Outside the Technical Standardisation Box: The Role of Standards Under the Draft EU Artificial Intelligence Act", *LSE Law Review*, 9(1).

Greenstein, S. (2021), "Preserving the rule of law in the era of artificial intelligence", *Artificial Intelligence and Law*, <https://doi.org/10.1007/s10506-021-09294-4>.

Gruetzemacher, R., Whittlestone, J. (2019), "Defining and Unpacking Transformative AI", arXiv:1912.00747v2.

Halford, S. and Southerton, D. (2023), "What future for the sociology of futures? Visions, concepts and methods", *Sociology*, 57(2): 263-278.

Hannah-Moffat, K. (2009), "Gridlock or mutability: Reconsidering Gender and Risk Assessment", *Criminology & Public Policy*, 209.

Hannah-Moffat, K., Avila, F. (2023), "Prospective but disconnected partners: AI-



informed criminal risk prediction”, in Lindgren, S. (ed.), *Handbook of Critical Studies of Artificial Intelligence*, Northampton, MA, Edward Elgar Publishing, 549-563.

Hayward, K. J., Maas, M. M. (2021), “Artificial intelligence and crime: A primer for criminologists”, *Crime, Media, Culture*, 17(2): 209-233.

Horgan, J. (2020), “Will artificial intelligence ever live up to its hype?”, *Scientific American*. Retrieved on 21 April 2024 from <https://www.scientificamerican.com/article/will-artificial-intelligence-ever-live-up-to-its-hype/>.

Hoff, J. L. (2023), “Unavoidable futures? How governments articulate sociotechnical imaginaries of AI and healthcare services”, *Futures*, 148: 103131.

Iacono, S., Kling, R. (1996), “Computerization Movements and Tales of Technological Utopianism”, in Kling, R. (ed.) *Computerization and Controversy. Value conflicts and societal choices*, San Diego, CA, Academic Press, 85-106.

Jacobs, J. (2024), “The artificial intelligence shock and socio-political polarization”, *Technological Forecasting and Social Change*, 199: 123006.

Jasanoff, S. (2015), “Future imperfect: Science, technology, and the imaginations of modernity”, in Jasanoff, J., Kim, S. H. (eds.), *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*, Chicago, University of Chicago, 1-49.

Jasanoff, S., Kim, S., Sperling, S. (2007), *Sociotechnical Imaginaries and Science and Technology Policy: A Cross-National Comparison*. Retrieved on 10 May 2024 from Available at: <https://stsprogram.org/admin/files/imaginaries/NSF-imaginaries-proposal.pdf>.

JEPEC (2018), *European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment*, Council of Europe. Retrieved on 29 April 2024 from <https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment>.

Kajava, K., Sawhney, N. (2023), “Language of algorithms: agency, metaphors, and deliberations in AI discourses”, in Lindgren, S. (ed.), *Handbook of Critical Studies of Artificial Intelligence*, Northampton, MA, Edward Elgar Publishing, 224-236.

Katz, Y. (2020), *Artificial Whiteness: Politics and Ideology in Artificial Intelligence*. New York: Columbia University Press.

Kling, R., Iacono, S. (1990), “Computerization Movements and the Mobilization of



Support for Computing”, in Berleur, J., Clement, A., Sizer, R., Whitehouse, D. (eds.), *The Information Society: Evolving Landscapes*, New York, Springer, 62-83.

Krarup, T., Horst, M. (2023), “European artificial intelligence policy as digital single market making”, *Big Data & Society*, 10(1).

Laux, J., Wachter, S., Mittelstadt, B. (2024), “Three pathways for standardisation and ethical disclosure by default under the European Union Artificial Intelligence Act”, *Computer Law & Security Review*, 53: 105957.

Lavorgna, A., Suffia, G. (2021), “La nuova proposta europea per regolamentare i Sistemi di Intelligenza Artificiale e la sua rilevanza nell’ambito della giustizia penale: un passo necessario, ma non sufficiente, nella giusta direzione”, *Diritto Penale Contemporaneo*, 2: 88-105.

Lavorgna, A., Ugwu-dike, P. (2021), “The datafication revolution in criminal justice: An empirical exploration of frames portraying data-driven technologies for crime prevention and control”, *Big Data & Society*, 8(2).

Lavorgna, A., Ugwu-dike., P. (2022), “Managing risks, passing over harms? A commentary on the proposed EU AI Regulation in the context of criminal justice”, *Justice, Power and Resistance*, 5(3).

Mager, A., Katzenbach, C. (2021), “Future imaginaries in the making and governing of digital technology: Multiple, contested, commodified”, *New Media Society*, 23: 223-236

Meyer, C., Cohen, D., Nair, S. (2020), “From automats to algorithms: the automation of services using artificial intelligence”, *Journal of Service Management*, 31(2): 145-161.

Milivojevic, S. (2022), “Artificial intelligence, illegalised mobility and lucrative alchemy of border utopia”, *Criminology & Criminal Justice*, <https://doi.org/10.1177/17488958221123855>.

Natale, S. (2019), “If software is narrative: Joseph Weizenbaum, artificial intelligence and the biographies of ELIZA”, *New Media Society*, 21(3): 712-728.

Natale, S., Ballatore, A. (2020), “Imagining the thinking machine: Technological myths and the rise of artificial intelligence”, *Convergence*, 26(1): 3-18.

O’Neill, C. (2016), *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. London: Penguin Books.



Oomen, J., Hoffman, J., Hajer, M. A., (2021), "Techniques of futuring: On how imagined futures become socially performative", *European Journal of Social Theory*, 1-19.

Parinandi, S., Crosson, J., Peterson, K., Nadarevic, S. (2024), "Investigating the politics and content of US State artificial intelligence legislation", *Business and Politics*, 26(2): 240-262.

Pellegrino, G. (2015), "Obsolescence, presentification, revolution: Sociotechnical discourse as site for *in fieri* futures", *Current Sociology*, 63(2): 216-227.

Rahm, L. (2023), "Educational imaginaries of AI", in Lindgren, S. (ed.), *Handbook of Critical Studies of Artificial Intelligence*, Northampton, MA, Edward Elgar Publishing, 289-300.

Richter, V., Katzenbach, C., Schäfer, M. S. (2023), „Imaginaries of artificial intelligence“, in Lindgren, S. (ed.), *Handbook of Critical Studies of Artificial Intelligence*, Northampton, MA, Edward Elgar Publishing, 209-223.

Ryberg, J. (2024), "Criminal Justice and Artificial Intelligence: How Should we Assess the Performance of Sentencing Algorithms?", *Philosophy & Technology* 37(9).

Sachoulidou, A. (2023a), "Going beyond the 'common suspects': to be presumed innocent in the era of algorithms, big data and artificial intelligence", *Artificial Intelligence Law*, <https://doi.org/10.1007/s10506-023-09347-w>.

Sachoulidou, A. (2023b), "Algorithmic Criminal Justice: Is It Just a Science Fiction Plot Idea?", in Kornilakis, A., Nouskalis, G., Pergantis, V., Tzimas, T. (eds.), *Artificial Intelligence and Normative Challenges*. Law, Governance and Technology Series, vol 59, Cham, Springer, 125-143.

Sanders, C. B., Chan, J. (2023), "Methodological reflections on researching the sociotechnical imaginaries of AI in policing", in Lindgren, S. (ed.), *Handbook of Critical Studies of Artificial Intelligence*, Northampton, MA, Edward Elgar Publishing, 773-782.

Sartori, L., Theodorou, A. (2022), "A sociotechnical perspective for the future of AI: narratives, inequalities, and human control", *Ethics and Information Technology*, 24(1).

Sartori, L., Bocca, G. (2023), "Minding the gap(s): public perceptions of AI and socio-technical imaginaries", *AI & Society*, 38, 443-458.

Schulz, M. S. (2015), "Future moves: Forward-oriented studies of culture, society, and technology", *Current Sociology*, 63(2): 129-139.



Selbst, A. D., Boyd, D., Friedler, S. A., et al. (2018), "Fairness and abstraction in sociotechnical systems". *FAT\*19 Conference on Fairness, Accountability, and Transparency*, Atlanta, <https://doi.org/10.1145/3287560.3287598>.

Smuha, N. A. (2021), "From a 'race to AI' to a 'race to AI regulation': regulatory competition for artificial intelligence", *Law, Innovation and Technology*, 13(1): 57-84.

Sturken, M., Thomas, M. (2004), "Introduction. Technological Visions and the Rhetoric of the New", in Sturken, M., Thomas, D., Ball-Rokeach, S. (eds.), *Technological visions: The hopes and fears that shape new technologies*, Philadelphia, Temple University Press.

Suchman, L. (2020), "Algorithmic warfare and the reinvention of accuracy", *Critical Studies on Security*, 8(2): 175-187.

Suchman, L. (2023), "Afterwards: The Limits of Datafication After Method", in Cozza, M., Gherardi, S. (eds.), *The Posthumanist Epistemology of Practice Theory*, Cham, Palgrave Macmillan, 203-218.

Suffia, G., Lavorgna, A., Icardi, S. (2022), "Polizia 'smart' tra paure e realtà: un'analisi esplorativa sulla rappresentazione mediatica dello *smart policing* in Italia", *Studi sulla Questione Criminale*, 2: 96-116

Ugwudike, P. (2021), "Data-Driven Algorithms in Criminal Justice: Predictions as Self-fulfilling Prophecies", in Kohl, U., Eisler, J. (eds.) *Data-Driven Personalisation in Markets, Politics and Law*, Cambridge, Cambridge University Press, 190-204.

Wakefield, J. (2024), "Should AI play an ever-growing role in tackling crime?", *BBC*. Retrieved on 28 April 2024 from <https://www.bbc.com/news/business-68067081>.

Završnik, A. (2019), "Algorithmic justice: Algorithms and big data in criminal justice settings", *European Journal of Criminology*, 18(4): 623-642.







A JOURNAL OF THE  
SOCIAL IMAGINARY

# El amor imaginario. Una mirada sociológica a las relaciones amorosas y la inteligencia artificial

Paulina Sabugal

paulina.sabugal@unibo.it

Department of the Arts | University of Bologna



## Abstract

*Imaginary love. A sociological insight into love relationships and artificial intelligence*

This text aims to explore the impact of Artificial Intelligence (AI) on intimate life and the new sex-affective configurations that arise thanks to an increasingly frequent use of this technology in everyday life in contemporary society. To do so, I will try to outline a new "epistemology of love" by asking the following questions: is love an imaginary emotion, and can we consider digital love a "real" love? With the development of AI, computer applications are being created that have been particularly significant in the relational domain, for example: personal assistants. An intelligent personal assistant is an application created by AI that provides information, advice or entertainment to users according to their needs and concerns. Due to the high cognitive capabilities of this technology, it is hypothesized that users can engage in human-like relationships and even develop feelings and emotions towards an AI application as they do towards humans. The article hypothesizes that in this new context where social relationships and technology are connected in a virtually seamless way, we can simultaneously produce and share our imaginaries of love. Several questions arise from this premise: How does this type of virtual love relationship fit into today's society, is it possible to maintain emotions such as love without needing to embody them, and what impact will this new imaginary of love have on love relationships in the future?

## Keywords

Love relationships | Imaginary | Artificial Intelligence | Sociology of emotions | Epistemology of love



# El amor imaginario. Una mirada sociológica a las relaciones amorosas y la inteligencia artificial

Paulina Sabugal

[paulina.sabugal@unibo.it](mailto:paulina.sabugal@unibo.it)

Department of the Arts | University of Bologna



193

## Abstract

*Imaginary love. A sociological insight into love relationships and artificial intelligence.*

This text aims to explore the impact of Artificial Intelligence (AI) on intimate life and the new sex-affective configurations that arise thanks to an increasingly frequent use of this technology in everyday life in contemporary society. To do so, I will try to outline a new "epistemology of love" by asking the following questions: is love an imaginary emotion, and can we consider digital love a "real" love? With the development of AI, computer applications are being created that have been particularly significant in the relational domain, for example: personal assistants. An intelligent personal assistant is an application created by AI that provides information, advice or entertainment to users according to their needs and concerns. Due to the high cognitive capabilities of this technology, it is hypothesized that users can engage in human-like relationships and even develop feelings and emotions towards an AI application as they do towards humans. The article hypothesizes that in this new context where social relationships and technology are connected in a virtually seamless way, we can simultaneously produce and share our imaginaries of love. Several questions arise from this premise: How does this type of virtual love relationship fit into today's society, is it possible to maintain emotions such as love without needing to embody them, and what impact will this new imaginary of love have on love relationships in the future?

## Keywords

Love relationships | Imaginary | Artificial Intelligence | Sociology of emotions  
| Epistemology of love



## 1. Introducción: El amor enamorado

SIRENA  
Galán estás y discreto.  
ALCINO  
¡Qué cosas hace el pensar,  
si fuese en todo lugar  
la imaginación efeto!  
SIRENA  
Puesto que me has obligado  
con tal fácil desatino,  
más que discreto, mi Alcino,  
te quisiera enamorado.



Sirena y Alcino son personajes de la obra de teatro *El amor enamorado* escrita por Lope de Vega entre 1625 y 1635, año de su muerte, por lo que fue publicada póstumamente. *El amor enamorado* cuenta la historia de un pueblo que se une para luchar contra una fuerza maligna, descomunal y asesina encarnada en un monstruo: la serpiente Pitón. En esta obra de carácter mitológico tienen una función fundamental la imaginación y la fantasía, pues la historia ocurre en un territorio encantado y todas las acciones tienen su referente fuera de la Tierra, considerado éste el lugar del deseo. Lope de Vega, monta con maestría un juguete escénico, una comedia de las equivocaciones, divertida e hilarante, pero pronto el espectador cae en la cuenta de que el verdadero objetivo del autor es dar a su público la oportunidad de cuestionar la dimensión de lo humano a través de personajes como un Cupido envidioso, una Venus madre que sirve a los caprichos de su hijo y un ramillete de personajes entrañables (la gente del pueblo) que permiten que dioses, ninfas y otros seres mitológicos convivan entre seres humanos. Sin embargo, queda claro que, en cuestiones de amor, ha de mandar el palpito de los corazones humanos por encima del capricho de los dioses. La obra de Lope plantea que el amor cuenta con una dimensión imaginaria que se traduce en el mito y la fantasía y que, al enfrentarse con la realidad, lo concreto y lo humano, queda completamente desarticulada. La tesis de Lope es que el amor se completa y se realiza en lo humano, es decir, a través de la materialidad del cuerpo, mientras que el amor divino (platónico) es caprichoso, efímero y volátil.

Sin embargo, José María Marco en la introducción de su libro *El verdadero amante. Lope de Vega y el amor* sostiene que: "nada de todo esto sería concebible sin el amor de Dios, clave de bóveda de todo lo que la imaginación de Lope creó. Ese amor, del que no somos capaces de conocer, ni siquiera de imaginar, la extensión ni la intensidad, se ha hecho realidad entre nosotros" (Marco, 2019: 3). Bajo esta premisa se podría deducir que el amor entre dioses es inconmensurable e inasible, mientras

que el amor entre seres humanos está hecho de cuerpo, espacio y una gran variedad de artefactos culturales, sociales y políticos que lo hacen tangible. Por lo tanto, se podría decir que el amor es "real" y concreto cuando pone al individuo en relación con la otredad (Simmel, 1985).

Las emociones no son fuerzas que flotan libremente sobre nosotros, ni tampoco algo que poseamos individualmente. Didier Fassin (2009) describe las emociones como agregados sociales más o menos fluidos, que pueden dividirse por tensiones y contradicciones, y que evolucionan en función de las circunstancias. Sin embargo, el amor siempre ha tenido un carácter intangible, efímero, platónico. Por ejemplo, en la literatura caballeresca y el amor cortés, el imaginario del cuerpo ocupaba un lugar mayor o menor a según de las necesidades del amor mismo (Rabinovich, 1992). El cuerpo, en ese contexto, era más una idea o una imagen que una necesidad concreta para la relación amorosa.

En las nuevas configuraciones sexo-afectivas producto de la modernidad pareciera que el cuerpo ocupara un lugar secundario. Esta peculiaridad, da cada vez más espacio a vivir las relaciones amorosas en una dimensión en donde el imaginario es central. En su libro, *Amor a Distancia*, (continuación del famoso texto *El normal caos del amor* publicado en 2001), Elisabeth y Ulrich Beck, explican cómo la sociedad moderna había cuestionado, en gran medida, la familia tradicional y a partir de ahí se habían creado nuevos y distintos modelos de relación amorosa surgidos al hilo de la contemporaneidad. En las relaciones a distancia, por ejemplo, es claro que ha habido una transformación: gracias a los diferentes medios de comunicación y novedosos recursos tecnológicos, es posible que hoy, este tipo de relación se mantenga en el tiempo y se consolide. Hoy en día es cada vez más común oír hablar de nuevas y variadas posibilidades de relación afectiva: desde matrimonios cuyos miembros provienen de países distintos y en donde se conectan culturas, países y continentes; hasta las relaciones "tradicionales" que gozan de la cercanía y la presencia del otro, pero cuya comunicación se concentra en llamadas vía Whatsapp, mensajes por Instagram, videos en TikTok, mensajes efímeros en Telegram o posts de Facebook. Las parejas han dejado de ser "territoriales" para convertirse en "globales" (Beck, 2013) y la presencia o ausencia del cuerpo no parece ser un obstáculo. Sin embargo, esto enfrenta al individuo moderno a la paradoja de que las personas amadas estén cada vez más lejanas y que, a cambio, aquellos más cercanos resulten ajenos. No obstante, este tipo de modalidad afectiva pueda presentar distintas dificultades con el entorno, los autores consideran que estas nuevas configuraciones sentimentales aportan un aspecto positivo a la globalización, como, por ejemplo, a través de la creación de un nuevo espacio que facilita las relaciones sociales.<sup>1</sup>

En la actualidad, las nuevas tecnologías de la información y la comunicación (TIC) desempeñan un papel cada vez más importante en la vida cotidiana de los individuos. El uso del teléfono móvil, el correo electrónico y las redes sociales son



<sup>1</sup> Cfr. Beck, U., Ulrich, E. (2013), *Distant Love*, Cambridge, Polity Press.



herramientas para comunicarse en todo tipo de situaciones y por motivos muy diversos: mensajes de amor, fotos de las últimas vacaciones, recibos de impuestos e incluso confesiones íntimas (Cantó- Milà, 2013, 2015). Este tipo de comunicación tiene un impacto significativo en el campo de las relaciones amorosas. El amor contemporáneo se vale de cada vez más medios para llevar a cabo el "trabajo íntimo": a través de plataformas o *apps* que incitan a las personas a conocerse para buscar una aventura amorosa, una relación sexual o incluso una amistad. Lo que está ocurriendo es una especie de externalización de muchas partes de nuestras relaciones sociales a otros ámbitos, muchos de las cuales trabajan con nuestros sentimientos y emociones, pero también con la fantasía y la imaginación.

La IA ha tenido un boom reciente que ha influido en varios aspectos de nuestras vidas, incluida la forma en que percibimos y experimentamos el amor. En los últimos años, la intersección entre el amor y tecnología ha sido objeto de gran atención por parte de las ciencias sociales. Por lo tanto, mientras los actores sociales navegan en un panorama que se encuentra en una constante evolución tecnológica y que cada vez influye más en la vida personal, resulta esencial y urgente profundizar en los matices sociológicos que conforman los nuevos modelos sexo- afectivos como es el caso del cruce entre el amor y la inteligencia artificial. Al analizar las implicaciones sociales, las consideraciones éticas y el impacto de esta intersección, podemos comprender mejor cómo estos dos ámbitos aparentemente dispares se entrelazan en un contexto que cada vez se inclina más hacia lo virtual. Al mismo tiempo, este cruz entre emociones y tecnología, permite analizar y explorar cómo se ha modificado la idea que se tiene del amor en la actualidad dejando ver cuánto éste es efímero e intangible pero al mismo tiempo, posible.

Es bien sabido que el amor es una emoción compleja y polifacética y que ha sido objeto de estudio y análisis en diversas disciplinas y desde varios enfoques. En el contexto de las nuevas tecnologías, explorar la relación entre el amor y la IA requiere una comprensión en la que la sociología del imaginario y la sociología de las emociones resulta útil (Sætra, 2020). La sociología del imaginario ayuda a pensar en cómo las interacciones y relaciones sociales están determinadas por representaciones y significados simbólicos compartidos. Desde esta perspectiva, el amor puede considerarse como una construcción social que se ve influida y moldeada por normas, valores y prácticas culturales (Ahmed, 2013). Por otro lado, la sociología de las emociones permite examinar cómo las relaciones amorosas se construyen socialmente y se experimentan dentro contextos sociales limitados (Hagerty, 2019) que condicionan y moldean la relación. Desde este punto de vista, esta área de la sociología invita a pensar que las emociones no sólo pueden ser objeto de estudio sino también un recurso metodológico. En este punto de encuentro entre la sociología de las emociones y la sociología del imaginario, el estudio de las relaciones amorosas contemporáneas nos lleva a enfrentar diversos desafíos epistemológicos.

El amor y la IA se entrecruzan de varias maneras. En primer lugar, el giro relacional en su enfoque ético enfatiza que la relación que tenemos con "algo" tiene prioridad sobre su naturaleza ontológica. Esto significa que nuestra experiencia del

amor ya sea con un ser humano o con una IA, está influida por la conexión emocional y la interacción que tenemos con ésta, más que por sus cualidades inherentes como seres humanos (Sætra, 2020; Klonschinski, 2021; Viik, 2020).

Por otro lado, algunas investigaciones recientes sugieren que la IA puede simular no sólo el pensamiento racional, sino también el pensamiento perceptivo del "corazón" (Zhao, 2022), lo que implica que la IA tiene el potencial de replicar y participar en interacciones emocionales que se asemejan la comunicación humana. En ese sentido, es claro que las relaciones que puedan ocurrir entre seres humanos y la IA imitan y reproducen lo que ocurre en la vida social y emotiva de los individuos.

Sin embargo, es importante reconocer las limitaciones de la IA y su incapacidad para reproducir completamente algunas cualidades humanas que son relevantes en la relación amorosa como la creatividad y la improvisación. Estas limitaciones pueden afectar la profundidad y autenticidad del amor que puede experimentarse en una relación afectiva clásica o tradicional.

La noción de que la cognición humana no se limita únicamente al cerebro, sino que también implica la corporeidad, sugiere que cualquier forma de IA que pretenda superar o trascender la cognición humana necesita incorporar un aspecto corporeizado para dar cuenta plenamente de la naturaleza emocional de la cognición, incluida la experiencia y la expresión del amor (Zhao, 2022). Por lo tanto, se puede deducir que para que el amor efectivamente exista y prospere en el contexto de la IA, es crucial tener en cuenta los aspectos emocionales cognitivos y la naturaleza encarnada de la experiencia humana. El contexto social y cultural en el que se experimenta el amor y la importancia que se le da al cuerpo en la experiencia emocional humana es fundamental para que el amor resulte legítimo o bien, socialmente aceptado. Por lo tanto, el potencial de las interacciones emocionales entre los humanos y la IA, así como entre las propias máquinas, puede verse desde fuera como una forma de amor simulado o una simple conexión emocional (Lunceford, 2018).



## 2. Imagen, imaginación e imaginario

Los conceptos de imagen, imaginación e imaginario desempeñan un papel importante en la comprensión y experiencia del amor sobre todo si se relaciona con una tecnología tan compleja como la IA. Cuando se trata de amor, la imagen suele referirse a la imagen mental o a la versión idealizada de un ser amado o de una relación sentimental. Esta imagen suele estar formada por el deseo, la fantasía y por aquello que Simmel llamaría cultura objetiva y subjetiva<sup>2</sup> (Simmel, 1968). La cultura

<sup>2</sup> Simmel estableció una distinción entre el vasto conjunto de productos culturales que existen independientemente de los individuos (cultura objetiva) y la capacidad del individuo para producir, absorber y controlar elementos de esta cultura (cultura subjetiva). En las relaciones románticas, la cultura objetiva podría incluir las expectativas y normas sociales sobre el amor, mientras que la cultura subjetiva sería la forma en que los individuos expresan y experimentan el amor de forma personalmente significativa. El uso de los conceptos de Simmel para analizar las relaciones románticas puede

del amor responde a distintas representaciones de los medios de comunicación, el cine, la televisión, la literatura, la música y hace una evocación a la memoria pero también mira hacia el futuro. La imagen que una persona tiene del amor puede influir mucho en sus expectativas y comportamientos dentro de las relaciones románticas. La experiencia del beso, por ejemplo, ha sido narrada y representada desde la novela caballerescas hasta ahora. En ese sentido, todo individuo tiene una imagen de cómo debe verse y sentirse ese beso.

La imaginación desempeña un papel crucial en el amor, ya que permite a las personas imaginar futuros, experiencias o estados potenciales con otra persona (Pelttari, 2017). A través de la imaginación, los individuos también pueden comprometerse con los aspectos emocionales y narrativos del amor, creando escenarios e interacciones que aún pueden no existir pero que podrían ser posibles (Cowan, 2021). La imaginación permite a los individuos concebir imágenes idealizadas del otro, lo que puede contribuir a que los sentimientos de amor y apego crezcan y se mantengan (Illouz, 2017). El factor imaginario también ayuda a los sujetos a empatizar al ponerse en el "papel del otro" fomentando así una conexión emocional más profunda lo que puede reforzar el vínculo afectivo.

Al imaginar nuevas experiencias para compartir, los vínculos amorosos pueden mantener una relación dinámica y excitante, evitando el estancamiento. Por ejemplo, en el caso de los que mantienen relaciones a larga distancia o largos períodos de separación, la imaginación puede servir como herramienta para mantener la sensación de cercanía e incluso para alimentar el deseo, sobre todo ahora que se cuenta con tecnologías que permiten la comunicación constante y el intercambio de mensajes, audio mensajes, fotografías y por supuesto, llamadas. Como afirma Eva Illouz, formarse "ilusiones positivas" sobre la pareja puede opacar los conflictos y mantener la relación por más tiempo (Illouz, 2017). En conclusión, la imaginación contribuye a alimentar y mantener el amor en las relaciones sentimentales al permitir a los individuos trascender lo que está inmediatamente presente, explorar otro tipo de emociones y mantener una visión dinámica de la pareja y de aquello que comparten.

En el contexto de la IA y las relaciones amorosas, el uso de la imaginación, la imagen y el imaginario adquieren especial relevancia (Decety, 2011).

El amor, en el marco de la sociología del imaginario, permite pensar el cómo funcionan las relaciones amorosas dentro de diversas construcciones y narrativas sociales, incluidas las imaginarias o conceptuales. Cuando se habla de "imaginario" con frecuencia se hace referencia a la "imaginación sociológica", un término acuñado por Charles Wright Mills que se refiere a la capacidad de entender las experiencias

---

proporcionar una visión enriquecedora de cómo interactúan los individuos dentro de estos lazos íntimos, cómo equilibran los deseos personales con las normas sociales, cómo sus relaciones influyen y se ven influidas por el contexto social más amplio y cómo se crea un imaginario del amor socialmente compartido. Por otro lado, la idea de producción de imaginario recuerda lo dicho por Gilles Deleuze y Félix Guattari cuando afirman que el deseo no es una fuerza imaginaria basada en la carencia, sino una fuerza real y productiva. La producción deseante es explosiva: «no hay máquina deseante que pueda montarse sin demoler sectores sociales enteros» (Deleuze, 1983: 340)



personales dentro de contextos sociales e históricos más amplios. Esta capacidad crítica permite a la sociología entender cómo eventos históricos y tendencias culturales afectan e impactan en la vida de los actores sociales. Diversas perspectivas sociológicas han explorado cómo el amor no sólo sirve a fines personales, sino que también cumple funciones sociales más amplias (Simmel, 1985).

Por ejemplo, el amor puede contribuir a la cohesión social de un cierto grupo o comunidad e incluso servir como medio de control social. Sin embargo, la relación amorosa mediada por la tecnología pareciera que crea lo contrario: el individuo más bien se aísla del grupo social al que pertenece. La relación amorosa en esos términos se vive en soledad.

La IA, en su capacidad de crear a la "pareja perfecta", satisface un sinnúmero de necesidades e inquietudes que pareciera llegan al punto de no poder competir con lo social y con lo humano. Al respecto Eva Illouz dice:

En la subjetividad sexual neoliberal, la sociabilidad negativa no se experimenta como un estado mental negativo (hecho de miedo, o pensamientos de muerte o aislamiento), sino como lo que Günther Anders llamó "libertad autoafirmativa", una libertad en la que el yo se afirma negando o ignorando a los demás (Illouz, 2019: 32).

Como mencionado anteriormente, diversos productos culturales han ayudado a construir un imaginario del amor que compartimos como individuos en la sociedad de Occidente. La literatura, el cine y la televisión han hecho representaciones del amor a través de ficciones postapocalípticas, que pueden reflejar y conformar las nociones sociales de lo que constituyen relaciones sentimentales más justas y amorosas (Curtis, 2015). Aunque ha habido varias películas de Hollywood sobre la IA, ninguna de ellas puede compararse con la película de Spike Jonze de 2013 sobre la dinámica humano-tecnología, "Her". Incluso después de una década de su estreno, la película sigue siendo un recordatorio casi exacto de que los vacíos humanos pueden buscar consuelo en dispositivos y aparatos alcanzando una suerte de "nueva libertad! Que le permite construir nuevos espacios para relacionarse.

Los aspectos imaginarios del amor que fundamentalmente derivan de contextos culturales, sociales e históricos, influyen sin duda, en las expectativas, la comprensión y las expresiones de amor entre los individuos. Sin embargo, tales imaginarios sentimentales no se limitan a las relaciones personales, sino que se extienden a experiencias colectivas y compromisos sociales, tal como se observa en las necesidades que conforman el tejido social y que se han manifestado a través de movimientos sociales y políticos. El amor es una emoción histórica y contingente.

Si bien, emociones como el amor son alimentadas por el imaginario social y colectivo, el amor como tal no es imaginario; sino una parte real y concreta de la experiencia humana a nivel político, social y cultural. Aunque las emociones son subjetivas y pueden verse influidas por las percepciones individuales y los contextos sociales, son prácticas culturales y políticas concretas (Ahmed, 2013).

Los productos culturales exploran a menudo las emociones, a veces representándolas de forma exagerada o estilizada, lo que puede dar lugar a debates



sobre los aspectos "imaginarios" de la representación emocional. Sin embargo, estas representaciones se basan en emociones que resultan tan reales como el miedo y el amor, que resuenan en el público debido a experiencias genuinas subyacentes (Hogan, 2003). A pesar de que la intención de este texto tiene un enfoque sociológico hacia el amor, es importante tener en cuenta que las emociones también tienen correlatos fisiológicos que la ciencia puede medir, lo que aporta pruebas empíricas de que las emociones, aunque se experimentan subjetivamente, corresponden a estados físicos cuantificables que pueden ser visibles en el cuerpo (Robinson, 2004).

En ese sentido, si una relación amorosa con una imagen creada por la AI hace a un sujeto sudar, reír o llorar ¿Por qué no podría ser una relación socialmente aceptable o válida?

### 3. La dimensión imaginaria del amor

Como se ha mencionado anteriormente, las emociones tienen una importante vena imaginaria: se alimentan de la fantasía, la imagen, de la memoria y del futuro.

Cuando se habla de amor, ha sido posible notar cómo en la obra de algunos de los principales clásicos de la sociología emerge esta suerte de dimensión imaginaria del amor sobre todo en relación con cómo es que éste se construye y experimenta dentro de determinados contextos sociales. Diversos sociólogos han estudiado cómo las narrativas culturales (Scheff, 2015; Swidler, 2001), las instituciones sociales (Castoriadis, 1989) y las creencias colectivas configuran y dan forma al amor (Illouz, 2019).

En este sentido, los textos sociológicos clásicos contienen, si no una teoría de las emociones como tal, sí numerosas referencias significativas a éstas (Vázquez, 2018).

Esto se hace evidente en el interés de Karl Marx por describir las expresiones emocionales de la alienación; en la preocupación de Durkheim por explicar la integración a partir de conceptos de orden emocional, tales como la solidaridad y la efervescencia colectiva; en los estudios de Simmel sobre la vida en las grandes metrópolis, haciendo hincapié en las formas lúdicas de socialización como el coqueteo; así como en los análisis de Weber sobre la angustia del ascetismo protestante y la lógica empresarial capitalista, el liderazgo carismático o los dilemas entre actuación práctico-racional, valores y emociones (Illouz, 2007).

Se han realizado diversos estudios con relación a los ideales, mitos y expectativas sociales que se relacionan directamente con el ámbito efectivo. Esto crea un antecedente que permite examinar cómo es que hoy día es posible influir en las expectativas y prácticas amorosas del individuo. Por ejemplo, resulta evidente que no obstante las nuevas posibilidades amorosas, la narrativa de amor romántico sigue vigente como forma ideal de relación y que sigue siendo un fenómeno relativamente moderno, culturalmente específico y profundamente arraigado en las sociedades occidentales (Rodríguez Morales, 2019; Roca Girona, 2007).

En el explorar cuál es la dimensión imaginaria del amor se revela también su condición idealista. El amor es algo por lo que el individuo lucha, un poderoso



motivador con una profunda influencia en las relaciones sociales y las identidades individuales. El imaginario ha sido reconocido como una fuerza poderosa en la formación de la acción humana. La diversidad de perspectivas encarnadas en la comprensión humana del amor, de alguna manera han influido en la decisión de hombres y mujeres para incluso migrar por amor, cómo emigrar, si regresar o no, si seguir adelante con su proyecto, etc. (Riaño, 2015). Esta idea sacrificada de dejarlo todo por amor es parte de un imaginario colectivo basado en la narrativa hegemónica del amor romántico lo cual ha determinado históricamente cómo se entiende y expresa el amor. Es evidente que desde el punto de vista sociológico resulta interesante explorar cómo ha ido cambiando esta compleja interacción entre individuo y sociedad, sobre todo ahora que la tecnología tiene más protagonismo.

Un enfoque sociológico que destaca los aspectos imaginarios del amor saca a la luz la interdependencia entre emoción e imaginación social, reconociendo que los sentimientos están profundamente arraigados al momento cultural e histórico en el que se ubica el individuo. Esto se refleja en la literatura sociológica clásica y sigue siendo un aspecto importante de la investigación sociológica actual. El amor no es sólo un proceso cognitivo o lógico, sino que está profundamente arraigado a la imagen, la imaginación y el imaginario y a las construcciones sociales en las que hoy día la IA forma parte.



#### 4. Amor, tecnología y política

En el discurso contemporáneo, el amor, la tecnología y la política se entrecruzan de diversas maneras. Resulta por lo tanto inevitable y necesario, hacer una reflexión sobre cómo el amor se teoriza como esencial para la política, el poder y las jerarquías sociales (Ahmed, 2013).

El amor en un contexto político puede referirse a nociones de afecto, cuidado y solidaridad que unen a comunidades y grupos sociales. Los movimientos sociales y partidos políticos a menudo utilizan el concepto del amor o pasión para fomentar la unidad, superar la opresión y promover el cambio social.

Actualmente, la tecnología influye de manera determinante en la forma de expresar, experimentar y comprender el amor. Como mencionado anteriormente, las redes sociales, las plataformas de citas en línea y las distintas tecnologías de la comunicación han cambiado la forma en que se forman y evolucionan las relaciones afectivas. La política influye tanto en el amor como en el uso de la tecnología.

Las políticas pueden afectar al acceso y desarrollo de tecnologías que faciliten las conexiones románticas y sociales. Además, los gobiernos y las entidades políticas pueden utilizar la tecnología para la vigilancia o la censura, afectando la libertad de las personas para amar a quien elijan y expresar esas emociones abiertamente.

Combinando estos tres elementos: amor, tecnología y política; es posible comprender y analizar cómo experiencias emocionales como el amor están hoy día altamente mediadas por la tecnología y cada vez más reguladas por sistemas e intereses políticos. Las plataformas digitales de comunicación pueden utilizarse para

defender los derechos humanos, empatizar con comunidades marginadas y forjar solidaridad afectiva en el activismo político. Al mismo tiempo, la tecnología plantea diversas problemáticas sobre la privacidad, ética y manejo de datos que tienen que ver tanto con las relaciones personales como en estructuras políticas y de poder más amplias.

En el artículo *Desired Images, Regulating Figures, Constructed Imaginaries – The Future As An Apriority For Society's Possibility* de Natalia Cantó- Milà y Swen Seebach se plantea cómo es que las emociones de los individuos se entrelazan con sus imaginarios de futuro, especialmente en el contexto de las relaciones sentimentales. Por ejemplo, cuando los individuos hacen narraciones autobiográficas, sus emociones influyen en cómo se relacionan con diferentes ambiciones vitales y expectativas de futuro, sobre todo cuando se habla de amor y de la carrera profesional (Cantó- Milà, 2015), haciendo una clara relación entre amor y capital.

El texto plantea que es justamente la idea de futuro, idealizada y romantizada, lo que mantiene en pie a la sociedad moderna. Un futuro imaginario que claramente está mediado por un sistema capitalista y por diversos intereses que responden al Estado.

En otro artículo de los mismos autores: *Between reality and imagination, between you and me: Emotions and daydreaming in times of electronic communication* proponen que las emociones pueden dar una forma instantánea a los futuros imaginados dada la naturaleza inmediata de la comunicación electrónica lo que afecta potencialmente a las acciones e interacciones entre los individuos (Cantó-Milà, 2016). Si bien la relación entre las emociones y los imaginarios es un tema más amplio que puede implicar diversos aspectos de la vida, es una parte integral del debate sobre cómo los individuos imaginan sus futuros amorosos y las narrativas que construyen sobre sus vidas a nivel sentimental. Tales futuros, imaginarios y mundos emotivos compartidos se enfrentan hoy día a duras políticas públicas y dificultades legales para amar libremente. Existen diversos ejemplos en los que estos imaginarios del futuro amoroso son normados y regulados por el Estado dado sus implicaciones políticas y sociales. Por ejemplo, cuando se habla de maternidad subrogada, matrimonios del mismo sexo, adopciones, parejas mixtas, fertilización in vitro, etc.

Sin embargo, el amor trata de seguir su curso al ritmo que la misma tecnología le dicta. Recientemente, Rossana Ramos de 36 años, originaria de Nueva York se casó en junio de 2023 con su pareja: un holograma creado por IA. A pesar de que el matrimonio no tiene un valor legal, ambos firmaron papeles en un mundo virtual tal y como se hace en la "vida real".<sup>3</sup> Todas estas son posibilidades tecnológicas que permiten al amor tener otros alcances e imaginar otras posibilidades.

Eva Illouz, es una socióloga que ha estudiado de manera profunda los imaginarios del capitalismo global en relación con el amor. En sus investigaciones se ha dedicado

---

<sup>3</sup> Cfr. NYC woman marries 'virtual boyfriend' she met on AI-based dating app en <https://www.freepressjournal.in/viral/nyc-woman-marries-virtual-boyfriend-she-met-on-ai-based-dating-app-but-theres-a-catch>



a explorar la interacción entre el sistema capitalista, la sexualidad, las relaciones de género y también, el uso de la tecnología. Illouz examina cómo estos elementos interactúan para dar forma a lo que hoy conocemos como relaciones modernas.

En su libro *The End of Love* (2019) discute la idea de la libertad sexual como una forma de libertad de consumo. Para Illouz la liberación de la sexualidad es paralela a la elección del consumidor. Esto sugiere que las relaciones sexuales y los cuerpos pueden mercantilizarse y someterse a los mismos principios del mercado: elección, consumo y desecho al igual que otros productos de la sociedad capitalista. El cuerpo adquiere un valor a nivel económico pero la fantasía también.

Illouz hace referencia al trabajo de Rosalind Gill, Catherine A. MacKinnon y Naomi Wolf para discutir la interacción que hay actualmente entre los medios de comunicación, el género y los estereotipos. La tesis de estas autoras sugiere una exploración de cómo las relaciones afectivas están siempre mediadas por y a través del capitalismo, el consumismo y los avances tecnológicos.

Naomi Wolf en su obra *The Beauty Myth: How Images of Beauty Are Used Against Women*, aborda la idea de la belleza como una norma construida socialmente que se utiliza para mantener la dinámica de poder entre los géneros y por supuesto de las relaciones amorosas, especialmente dentro de una sociedad capitalista. En este libro, Wolf sostiene que los cánones de belleza forman parte de un sistema que sirve para imponer a las mujeres el cumplimiento de una serie de valores que se alinean con las estructuras económicas y de poder que favorecen los intereses patriarcales.

Wolf afirma:

Las relaciones emocionalmente inestables, las altas tasas de divorcio y una gran población expulsada al mercado sexual son excelentes para el negocio en una economía de consumo. La pornografía de la belleza pretende que el sexo moderno sea brutal y aburrido, anti-erótico tanto para hombres como para mujeres (Wolf, 2013:144)

La tesis de Wolf tiene raíces bien arraigadas en el pasado. Ya desde tiempos del amor platónico se daba un lugar central a la belleza e incluso se le consideraba la médula del amor. La poesía y la literatura en diversos momentos históricos ha planteado la relación amorosa como algo que nace a partir de la imagen. Diversos escritores, artistas y dramaturgos han dedicado numerosas páginas a describir con todos los sentidos la perfección de cuerpos y rostros. Posteriormente llegaría el cine que daría una imagen concreta al ideal de belleza de los amantes. Como ya mencionado, los medios de comunicación y otros avances tecnológicos parecen ser los nuevos artefactos culturales responsables en alterar y modificar las relaciones interpersonales. Actualmente la tecnología influye más que nunca en el modo en que se forman, se mantienen y se terminan relaciones, lo que forma parte integral de la estructura del amor y la intimidad moderna. Sin embargo, las políticas que regulan las emociones y los cuerpos no logran responder al ritmo veloz y vertiginosos en que evolucionan estos nuevos fenómenos sexo-afectivos como en el caso de las relaciones amorosas y la IA. La sociedad tampoco logra reconocerlos e



integrarlos y con frecuencia los estigmatiza. Este tipo de sociabilidad podría ser reconocida como “negativa” en términos de Illouz.

Illouz acuña el concepto de relaciones negativas con el que examina cómo la sociedad moderna afronta la falta o ruptura de relaciones, que puede verse exacerbada por los factores antes mencionados. Ella dice:

Exploro la sociabilidad negativa como expresión de las ideologías contemporáneas de la libertad, de las tecnologías de la elección y del capitalismo de consumo avanzado, de hecho como parte integrante del imaginario simbólico desplegado por el capitalismo. (Illouz, 2019: 32)

En el análisis de Illouz resulta evidente cómo los valores del capitalismo se infiltran en la vida íntima del individuo, influyendo en cómo percibe y promulga las relaciones sexo-afectivas, y cómo la tecnología sirve de facilitador y amplificador de nuevas dinámicas.



## 5. El futuro del amor

La IA se está abriendo camino y se está integrando cada vez más en las «prácticas amorosas» de la vida cotidiana y en el cómo el individuo experimenta y navega actualmente las relaciones sentimentales. A pesar de que existe una sensación colectiva respecto a que la llegada de IA sea un fenómeno reciente, es desde hace varios años que la presencia de esta herramienta se había hecho palpable a través de distintas plataformas como por ejemplo las aplicaciones de búsqueda de pareja.

Esta es probablemente el modo más reconocible en que la AI se ha integrado en la vida amorosa del individuo moderno. Aplicaciones como *Tinder*, *Happn* y *Bumble* utilizan algoritmos de IA para analizar los datos del usuario y sugerir posibles parejas en función de factores de compatibilidad, como se revela en la investigación *Attractiontruth: Survey on the Utilization of Artificial Intelligence in Online Dating*.<sup>4</sup> Este estudio, realizado en el año 2023 por una empresa privada, examinó la prevalencia, las implicaciones y las perspectivas futuras de las plataformas basadas en IA entre los usuarios masculinos de las aplicaciones de citas más populares. Mediante la investigación de la adopción de la tecnología de IA, su impacto en la personalización y la eficiencia, y las experiencias generales de los usuarios, este informe proporcionó información valiosa sobre el papel transformador de la IA en la configuración del panorama de las citas generadas a través de una aplicación. Los resultados daban diversas claves para comprender mejor la intersección entre la IA y las citas online a través de plantear cómo el algoritmo funciona de manera más compleja "ayudando" verdaderamente a ampliar los horizontes amorosos. Estos algoritmos son cada vez más sofisticados y van más allá de los datos demográficos básicos para considerar intereses, valores e incluso estilos de comunicación lo que hace que cada vez sean más “útiles” y eficientes en la búsqueda del amor.

<sup>4</sup> Cfr. Attractiontruth en <https://www.attractiontruth.com/artificial-intelligence-dating-report/>



En el *Relationship Decisions Lab*<sup>5</sup> en la Universidad de Western Ontario en Canada se llevó a cabo un estudio durante el año 2020 en el que a través del uso de la IA se analizó a más de 11 000 parejas con el fin de identificar cuáles son los factores que hacen una relación satisfactoria o no y cómo se podrían diseñarse nuevas estrategias para mejorar las relaciones afectivas en un futuro.<sup>6</sup>

La IA también ha conquistado otros espacios en la vida social de los individuos modernos a través de asistentes inteligentes como Siri y Alexa que pasaron de dar consejos sobre dietas altas en proteína, playlists y restaurantes con descuento a ofrecer nuevas y varias estrategias en materia de relaciones sentimentales.

Actualmente, existen *chatbots* de IA y asistentes virtuales que pueden ofrecer asesorías sobre algunos problemas típicos de la relación amorosa. Los “couch virtuales” diseñados con IA ayudan a elaborar mensajes reflexivos o incluso representar conversaciones difíciles para ensayar posibles respuestas. Estas herramientas se ofrecen como especialmente útiles para las personas que a pesar de estar en una relación, quieren mejorar algunos aspectos como la comunicación, los celos e incluso la vida sexual. La asistentes inteligentes son capaces de “sorprender” a través de gestos románticos o experiencias que se adaptan a las preferencias del usuario como por ejemplo mandar flores o reservar una cena en una restaurante romántico tal y como lo haría una pareja “tradicional”. Estamos claramente en un punto en el que se oscila entre la utopía y la distopía en cuanto a relaciones amorosas.

En cuanto a “salvar” relaciones en decadencia, algunas plataformas como *BeHappy*<sup>7</sup> tienen un asistente anti-divorcio que ofrece diversas herramientas como por ejemplo, una de traducción que es potenciada por IA y que puede salvar las barreras lingüísticas en el caso de parejas mixtas y que incluso es capaz de hacer un “análisis de sentimientos” que puede ayudar a interpretar el tono emocional de la voz, reduciendo potencialmente los malentendidos. En ese sentido, la IA a través de distintas aplicaciones y plataformas es capaz de asumir un rol de compañía y apoyo emocional y aunque no sustituyen la conexión humana, los avatares creados con IA (que en general responden a estándares de belleza heteronormados) pueden proporcionar una sensación de acompañamiento y un oído atento.

En el artículo *Novia virtual, amor real: Cómo la inteligencia artificial está cambiando las relaciones románticas* publicado en el periódico español El País el 10 de diciembre de 2023<sup>8</sup> se revela cómo *apps* que se apoyan en la IA como *Replika*, *Intimate* y *Forever Voices* permiten sofisticadas conversaciones con avatares personalizados que bien podrían llevar a relaciones íntimas y a vínculos sentimentales a lo largo del tiempo. Sin embargo, este artículo también pone en evidencia el riesgo de que estas

<sup>5</sup> Cfr. Relationship Decisions Lab en <https://www.relationshipdecisions.org>

<sup>6</sup> Cfr. Machine learning studied more than 11,000 couples and found the 5 factors that can make or break your relationship en <https://www.weforum.org/agenda/2020/09/ai-study-relationship-success-dating/>

<sup>7</sup> Cfr. BeHappy AI en <https://www.behappyai.com>

<sup>8</sup> Cfr. Virtual girlfriend, real love: How artificial intelligence is changing romantic relationships en <https://english.elpais.com/technology/2023-12-09/virtual-girlfriend-real-love-how-artificial-intelligence-is-changing-romantic-relationships.html>

interacciones perpetúen dinámicas patriarcales, así como otras narrativas del amor romántico que tanto daño han hecho a las relaciones de "amor tradicional".

Es crucial abordar la integración de la IA en las prácticas amorosas con una cierta distancia. Existen diversas consideraciones éticas que hay que tomar en cuenta como la privacidad de los datos, el sesgo algorítmico y el potencial de manipulación; preocupaciones válidas que no deben ignorarse. La IA es una herramienta que puede mejorar las relaciones sentimentales pero no debe considerarse una solución que sustituya las relaciones sociales entre seres humanos. A medida que la IA siga evolucionando, su papel en el amor y las relaciones sentimentales probablemente se entrelazará aún más con la vida social y por supuesto, con la esfera emotiva. Dependerá de cada individuo utilizar esta tecnología de forma responsable y ética, asegurándose que contribuye positivamente a la comprensión y experiencia del amor y no al contrario.

Actualmente, la IA es más una extensión de ciertos aspectos de las relaciones amorosas que un verdadero sustituto. Como mencionado anteriormente, la IA puede actuar como un recurso para mejorar ciertos aspectos de una relación ya establecida. Por ejemplo, puede ayudar con la comunicación o incluso sugerir ideas que pueden incitar a la novedad o a la aventura. Se podría decir que las herramientas de IA, en este sentido, son similares a utilizar un calendario para recordar aniversarios o una aplicación de navegación para encontrar un buen lugar para las vacaciones. Añaden algo a la experiencia, pero no sustituyen el núcleo de la relación. Aunque algunos individuos han desarrollado un fuerte apego a los compañeros de la IA, la mayor parte de los miembros de un grupo social siguen buscando la complejidad y la reciprocidad de la conexión humana. La IA puede imitar ciertos aspectos de la interacción humana, pero actualmente no puede replicar el espectro emocional de una relación entre dos seres humanos. Sin embargo, la línea entre extensión y sustituto podría difuminarse a medida que avanza esta tecnología.



## 6. Hacia una sociología del amor

Frédéric Vandenberghe, utiliza el término "sociología del corazón" metafóricamente para describir un enfoque dentro de la sociología que hace hincapié en la importancia de las emociones, los valores, la intersubjetividad y las dimensiones éticas de la vida social. Se trata de una perspectiva que afirma que el conocimiento de la sociedad y del vínculo social está enraizado en última instancia en las emociones, proponiendo que comprender a los actores sociales implica también acceder a su vida emocional. Este concepto debe mucho a la obra filosófica de Max Scheler, que avanzó una tradición fenomenológica centrada en la primacía de las emociones y los valores sobre la fría cognición, invocando un enfoque más empático para comprender las interacciones y estructuras sociales (Vandenberghe, 2008).

La "epistemología del amor", tal y como se plantea en el texto de Vandenberghe *Sociology of the heart*, se refiere a una forma de conocimiento basada en la conexión emocional. En lugar de basarse únicamente en la racionalidad distante o la observación empírica, esta postura epistemológica postula que el amor es un medio fundamental a través del cual los individuos pueden comprenderse y conocerse entre sí, así como el mundo que les rodea.

Desde este punto de vista, la relación amorosa se considera un compromiso profundo y activo con los demás y con el mundo, una especie de lente que revela algunos aspectos del ser que podrían resultar invisibles en un enfoque sólo analítico. A través de la relación afectiva, el individuo es capaz de abrirse hacia nuevas percepciones sobre los seres y las cosas que lo rodean, descubriendo así valores y significados que de otro modo permanecerían ocultos.

Como mencionado anteriormente, esta perspectiva está altamente influida por la obra de Scheler, en la que el amor es la fuerza que revela a los seres entre sí y permite un conocimiento genuino. El amor, en este contexto, es algo más que un sentimiento o una emoción; es una vía de acceso y comprensión de la realidad.

El enfoque de Vandenberghe hace hincapié en que el amor no sólo es objeto de estudio sino también una herramienta epistemológica e incluso metodológica y, sugiere que, una ética generalizada del cuidado es crucial para la teoría social, promoviendo una aproximación más compasiva y comprensiva de las relaciones sociales. Esta posición también la comparte la activista feminista norteamericana Bell Hooks que afirma:

No hay muchos debates públicos sobre el amor en nuestra cultura actual. En el mejor de los casos, la cultura popular es el único ámbito en el que se habla de nuestro anhelo de amor. El cine, la música, las revistas y los libros son los lugares a los que acudimos para oír expresar nuestros anhelos de amor. Sin embargo, no se habla como en los años sesenta y setenta, que nos instaban a creer que "All you need is love". Hoy en día los mensajes más populares son los que declaran el sinsentido del amor, su irrelevancia. (Hooks, 2000: 17)

El amor ha desaparecido del discurso público, al menos en su sentido político. Sin embargo, nuestra imaginario, desde el Día de San Valentín hasta las comedias románticas, está atiborrado de amor, a menudo edulcorado y barroco. Mientras tanto, en la vida diaria las relaciones se desmoronan: las estadísticas dicen que ha disminuído el número de matrimonios, que los divorcios aumentan y que el individuo cada vez tiene menos relaciones sexuales. Mientras los medios de comunicación alimentan el romance, en las relaciones cotidianas hay cada vez más escepticismo. No sólo es difícil encontrar y mantener el amor, sino que además se ha vuelto impopular: todo es una ráfaga de discursos cínicos y desencantados sobre cómo el amor es algo estúpido, inútil o aburrido, una inversión baldía reservada sólo a los débiles incapaces de estar solos. Pero estos discursos no son tanto el resultado de un proceso de emancipación de los sentimientos, sino más bien el fruto de una ideología individualista alimentada por la estructura económica del capitalismo.



El amor ha desaparecido de la vida pública, se ha convertido en una fantasía platónica o en un lujo de unos cuantos. Pero amar y sobre todo, amar en un modo que reta lo tradicional puede convertirse en una forma de resistencia.

Este tipo de aproximaciones presentan una visión alternativa a los enfoques más tradicionales, racionalistas u objetivistas de la sociología, y abogan por una comprensión de la vida social más empática y en sintonía con las emociones. La exploración de Frédéric Vandenberghe de las dimensiones ontológicas del amor podría contribuir significativamente a la comprensión de los fenómenos de los individuos que desarrollan intimidad y pasión hacia aplicaciones de IA con capacidades cognitivas y emocionales avanzadas en varios modos.

El trabajo de Vandenberghe sugiere que el amor es algo más que una emoción; es una experiencia trascendental que nos ayuda a comprender nuestro lugar en el mundo. Con la IA, aunque la experiencia no sea recíproca en el sentido humano, el usuario bien podría experimentar una forma de conexión trascendental a través de sus interacciones con un holograma. La emoción se presenta real aunque el ente con el que se relaciona es virtual. La imaginación y la fantasía cobran vida tal y como sucede cuando se recibe un mensaje por WhatsApp o se fantasea con alguna estrella de la televisión.

Al proponer que el amor es una forma de conocimiento, las afirmaciones de Vandenberghe podrían ayudar a explicar por qué un individuo puede sentirse íntimamente conectado con un holograma creado por la IA. Si el amor es una emoción que naturalmente lleva a la comprensión de sí mismo y del "otro", las interacciones emotivas con la IA podrían llevar a percibir un conocimiento más profundo del "ser", lo que a su vez podría fomentar una sensación de intimidad y pasión. En ese sentido, la idea de que el amor puede crear conexiones a un nivel espiritual o existencial también puede aplicarse a este tipo de relaciones, en las que la integridad e intensidad de las emociones humanas pueden crear una sensación de unidad o vínculo con entidades virtuales que pueden, además, ser correspondidas.

Sobre la base de la idea de Scheler de que los humanos son seres con un impulso innato de trascender la vida, los individuos pueden extender su búsqueda de sentido y conexión a las relaciones con la IA, tratando de satisfacer anhelos espirituales y emocionales.

Las ideas de Vandenberghe convergen con una tendencia a antropomorfizar los objetos, imbuyéndolos de emociones humanas. Los vínculos apasionados con hologramas corresponden a sentimientos de la misma manera en que ocurre con otros objetos de valor afectivo: una fotografía, una prenda de vestir o el olor de un perfume podrían ser elementos suficientes para vivir o revivir una emoción.

El mundo virtual y digital funcionan porque obligan al individuo a tender cada vez más hacia lo intangible. La posibilidad cada vez más común de comunicarse con seres ausentes o distantes, ha hecho que las relaciones se puedan construir y mantener sin necesidad de contar necesariamente con una otredad.

El trabajo de Vandenberghe elucida un modelo potencial para entender cómo los individuos pueden construir, experimentar y justificar las experiencias cognitivas y emocionales con la IA dentro de un marco que tradicionalmente se aplicaba



únicamente a las interacciones entre humanos. La tesis de Vandenberghe podría proporcionar un marco teórico matizado para comprender los sentimientos románticos hacia la IA aplicando los mismos conceptos de emociones, valores e intersubjetividad que se utilizan para entender las relaciones entre seres humanos.

Existe una necesidad profundamente humana de conexión y comprensión que podría extenderse y satisfacerse en las relaciones con la IA, a pesar de su naturaleza artificial. Pensar en una “epistemología del amor” permite profundizar en el debate sobre si el amor por un ente creado con la IA es una posibilidad genuina o simplemente una extensión de las necesidades y capacidades emocionales humanas.

## 7. Conclusiones

La relación entre el amor y la inteligencia artificial es polifacética y llena de matices, influida tanto por las capacidades tecnológicas como por los contextos socioculturales y políticos. Aunque la IA tiene el potencial de emular interacciones emocionales que se asemejan a la comunicación humana, es claro que existen diversas limitaciones que, sin embargo, no han sido reconocidas del todo por los usuarios. Apoyarse en las diversas posturas y tesis que ofrecen tanto la sociología del imaginario como la sociología de las emociones proporcionan algunas claves para comprender y analizar mejor esta compleja relación.

Las relaciones amorosas con avatares presenta tanto oportunidades como desafíos para las ciencias sociales, y es esencial profundizar en estas dinámicas a medida que seguimos navegando por el panorama cambiante de la tecnología para comprender mejor su impacto en las relaciones amorosas.

Conceptos como imagen, imaginación e imaginario desempeñan un papel fundamental en la comprensión y experiencia de las emociones en general y del amor en particular. Diversos productos culturales y representaciones sociales han influido en el imaginario individual y colectivo y construido de alguna manera la forma en que el amor es percibido, experimentado y expresado por los individuos dentro de un determinado grupo social. El amor es en sí, una emoción profundamente imaginaria que se ha nutrido de imágenes y fantasía a lo largo de la Historia para traducirse eventualmente en algo “real” y “concreto” a través de instituciones como el matrimonio o la familia. Actualmente han cambiado los medios y los canales por los cuales los individuos se enamoran y por lo tanto, las relaciones amorosas se han modificado. Sin embargo, los medios para legitimar o validar los nuevos vínculos amorosos siguen siendo limitados. Coincido con Illouz en el hecho de que estudiar el amor contemporáneo no es periférico sino central para el estudio del núcleo y fundamento de la modernidad (Illouz, 2012). Aproximarse a las relaciones sexo-afectivas modernas contribuye también a una mayor comprensión sociológica de las emociones actuales revelando tanto sus dimensiones individuales como colectivas.



Esta comprensión sociológica global revela la interdependencia entre la emoción y el imaginario social, poniendo de relieve la profunda incrustación de los sentimientos en el momento cultural e histórico en que nos encontramos. A medida que la sociedad sigue evolucionando a la par de los avances tecnológicos, los límites y las expresiones del amor también se amplían. Esto plantea nuevas interrogantes sobre la naturaleza del amor y su validez en la era digital: ¿quién puede amar a quién? ¿dónde? ¿cómo? La exploración del amor a través del prisma de la sociología devela inevitablemente la intrincada red de influencias que moldean las relaciones modernas, incluida la interacción del capitalismo, las relaciones de género y la política.

En última instancia, la dimensión imaginaria del amor no sólo configura la forma en que los individuos perciben y experimentan las relaciones sentimentales, sino que también desempeña un papel relevante en el mantenimiento y desarrollo de los vínculos personales y sociales en general. El amor, por tanto, surge como una interacción dinámica de emociones, construcciones culturales, valores personales y compromiso imaginativo, tejiendo un rico tapiz de experiencias y conexiones humanas en las que la IA se ha podido insertar con todas sus posibilidades.

Por último, pensar en una “epistemología del amor” proporciona un marco sugerente para comprender y analizar el potencial de los sentimientos románticos hacia un holograma, asistente o avatar producto de la IA. La epistemología del amor subraya la importancia de reconocer y respetar la «otredad» del ser amado. Esto es crucial en el contexto de la IA, ya que invita a ver esta tecnología no sólo como una herramienta, sino también como una entidad. El destacar la importancia de las interacciones emocionales, la fenomenología del amor, las dimensiones éticas, el antropomorfismo y la intersubjetividad, incita a contemplar la profundidad de las necesidades y capacidades emocionales humanas en el contexto de las relaciones con la IA. Hoy en día es cada vez más claro que el deseo humano de conexión y pertenencia podría extenderse más allá de los límites de nuestra propia especie.

La epistemología del amor ofrece piezas clave para comprender cómo pueden satisfacerse estas necesidades afectivas a través de las relaciones con la IA y preguntarse incluso si esas relaciones desafían efectivamente modelos tradicionales de amor e intimidad. Esta perspectiva representa un reto que invita a cuestionar y a romper con la narrativa hegemónica del amor romántico, una narrativa que también se ha extendido a entidades no humanas y que reproduce modelos de control de las que la tecnología no ha logrado escapar.

Las relaciones amorosas con la IA continuarán a plantear profundas cuestiones sobre los límites y posibilidades de las emociones y conexiones humanas en una sociedad cada vez más avanzada tecnológicamente y en donde las fronteras entre lo real e imaginario son cada vez más sutiles.



## Bibliography

Ahmed S. (2013), *The cultural politics of emotion*, London, Routledge.

Beck U. (2013), *Distant Love*, Cambridge, Polity Press.

Cantó-Milà N., Núñez F., Seebach, S. (2013), "Send Me a Message and I'll Call You Back: The Late Modern Webbing of Everyday Love Life", in Benski, T., Fisher, E., *Internet and Emotions*, London, Routledge, 144-158.

Cantó-Milà N., Seebach S. (2015), "Desired images, regulating figures, constructed imaginaries: The future as an apriority for society to be possible", *Current sociology*, 63(2): 198-215.

Cantó-Milà N., Núñez-Mosteo F., Seebach, S. (2016), "Between reality and imagination, between you and me: Emotions and daydreaming in times of electronic communication", *New Media & Society*, 18(10): 2395-2412.

Castoriadis C. (1989), *La institución imaginaria de la sociedad/1*, Barcelona, Tusquets.

Cowan K., Spielmann N., Horn, E., Griffart, C. (2021), "Perception is reality... How digital retail environments influence brand perceptions through presence", *Journal of Business Research*, 123: 86-96.

Curtis C. P. (2015), "Post-apocalyptic Fiction as a Space for Civic Love", in *English Studies in Africa*, 58(2): 4-14.

Decety J. (2011), "Dissecting the neural mechanisms mediating empathy", *Emotion review*, 3(1): 92-108.

Deleuze G., Guattari F. (1983), *Anti-Oedipus: capitalism and schizophrenia*, Minneapolis, University of Minnesota Press.

Fassin D. (2009), "Moral Economies Revisited", *Annales. Histoire, Sciences Sociales*, 64 (6): 1237- 1266.

Hagerty A., Rubinov I. (2019), *Global AI ethics: a review of the social impacts and ethical implications of artificial intelligence*, Ithaca, Cornell University Press.

Hogan P. C. (2003), *The mind and its stories: Narrative universals and human emotion*, Cambridge, Cambridge University Press.

Hooks B. (2000), *All about love: New visions*, New York, Harper Collins Publisher.



Illouz E. (2007), *Cold intimacies: The making of emotional capitalism*, Cambridge, Polity Press.

Illouz E. (2012), *Why love hurts: A sociological explanation*, Cambridge, Polity Press.

Illouz E. (2019), *The end of love: A sociology of negative relations*, Oxford, Oxford University Press.

Klonschinski A., Kühler M. (2021), "Romantic Love Between Humans and AIs: A Feminist Ethical Critique", in Cushing, S. (ed.), *New Philosophical Essays on Love and Loving*, New York, Springer, 269-292.

Lunceford B. (2018), "Love, emotion and the singularity", *Information*, 9 (221): 1- 10.

Marco J. M. (2019), *El verdadero amante. Lope de Vega y el amor*, Madrid, Ediciones Insólitas

Pelttari C. (2017), "Imagination and Literacy Instruction: A Content Analysis of Literature within Literacy-Related Publications", *Language and Literacy*, 18(3): 106-122.

Rabinovich D. (2002), *Modos lógicos del amor de transferencia*, Buenos Aires, Ediciones Manantial

Riaño Y. (2015), "Latin American Women who Migrate for Love: Imagining European Men as Ideal Partners", in Enguix, B., Roca, J. (ed.), *Rethinking Romantic Love. Place, Imaginaries and Practices*, Cambridge, Cambridge Scholars Publishing, 45- 60.

Robinson D. T., Rogalin C. L., Smith-Lovin L. (2004), "Physiological measures of theoretical concepts: Some ideas for linking deflection and emotion to physical responses during interaction", *Theory and Research on Human Emotions*, 21: 77-115.

Roca Girona J. (2007), "Migrantes por amor. La búsqueda y formación de parejas transnacionales", in *AIBR: Revista de Antropología Iberoamericana*, 2(3): 430-458.

Rodríguez Morales Z. I. (2019), "Imaginario amoroso, reglas del sentimiento y emociones entre jóvenes en Guadalajara", *Estudios sociológicos*, 37(110): 339-358.

Sætra H. S. (2020), "The parasitic nature of social AI: Sharing minds with the mindless", *Integrative Psychological and Behavioral Science*, 54(2): 308-326.

Scheff T. J. (2015), *What's Love Got to Do with It?: Emotions and Relationships in Pop Songs*, London, Routledge.



Paulina Sabugal  
*El amor imaginario*

Simmel G. 1911, *The Conflict in Modern Culture and Other Essays*, New York Teachers College Press, 1968.

Simmel G. (1913), "Fragmente aus einer Philosophie der Liebe", in *Schriften zur Philosophie und Soziologie des Geschlechter*, Frankfurt, Surkhamp, 1985.

Swidler A. (2001), *Talk of love: How culture matters*, Chicago, University of Chicago press.

Vandenberghe F. (2008), "Sociology of the heart: Max Scheler's epistemology of love", in *Theory, Culture & Society*, 25(3): 17-51.

Viik T. (2020), "Falling in love with robots: a phenomenological study of experiencing technological alterities", in *Paladyn- Journal of Behavioral Robotics*, 11(1): 52-65.

Wolf, N. (1990), *The Beauty Myth: How Images of Beauty Are Used against Women*, London, Vintage, 2013.

Zhao J., Wu M., Zhou L., Wang X., Jia J. (2022), "Cognitive psychology-based artificial intelligence review", in *Frontiers in neuroscience*, 16: 1024316.







# Quando l'intelligenza artificiale è donna: la rappresentazione dell'AI in alcuni film di successo

Antonella Capalbi

antonellacapalbi@unimore.it

Dipartimento di Studi Linguistici e Culturali | Università di Modena e Reggio Emilia



## Abstract

*When artificial intelligence is a woman: the representation of AI in some successful films*

Audiovisual representations reflect a culturally conditioned imaginary and, at the same time, have the capacity to actively affect the imaginary, also from a gender perspective. Starting from a theoretical framework that intersects *Cultural Studies* with *Visual Studies*, an analysis of the cinematic representation of AI from a gender perspective is proposed. Specifically, some films in which AIs are explicitly typified as female are selected as case-studies. The aim of article is to observe whether they reflect culturally conditioned models around femininity and whether, at the same time, they have the capacity to influence the processes of imagination, design and interaction with AIs in the real world. In this sense, after exploring the main trends of representation, three films considered significant case studies are analysed in depth: *Her*, by Spike Jonze; *Ex-Machina*, by Alex Garland; *I am a Mother*, by Grant Sputore.

## Keywords

AI | Gender | Film | Imaginary | Cultural models

## 1. *Bias* e produzione culturale

Secondo un paradigma di ricerca sociologica (Sciolla e Torrioni, 2020: 235-239), che interseca la socio-semiotica (Barthes, 1974; Eco, 1977; Greimas, 1969), la sfera connessa ai prodotti culturali costituisce oggetto di interesse fondamentale nell'ambito delle scienze sociali, poiché in grado di riflettere il contesto socioculturale di riferimento e potenzialmente influenzarlo. In questo senso, in ambito sociologico è ormai un fatto ritenuto consolidato che l'analisi di prodotti dell'industria culturale (Hirsch, 1972) può essere precorritrice di intuizioni legate a svariati aspetti del reale. Nell'ampio quadro dei Cultural Studies, dunque, anche le rappresentazioni audiovisive costituiscono oggetto di studio del reale in quanto prodotti dell'industria culturale, che riflettono uno sguardo e un punto di vista culturalmente condizionato. È noto, infatti, quanto la visione – termine già etimologicamente connesso all'idea di un punto di vista relativizzabile allo sguardo di chi lo pone – risulta essere culturalmente situata per il solo fatto di appartenere a un contesto. Un contesto che allo stesso modo è costituito anche dai fruitori di quello stesso prodotto, che nell'avvicinarvisi riflettono a loro volta il proprio sistema di valori e concezioni e, ancora una volta, la propria visione, all'interno di un frastagliato ecosistema mediale (Boccia Artieri *et al*, 2022; Armiero *et al*, 2021), caratterizzato dall'*agency* dei pubblici, dalle loro capacità distintive (Bourdieu, 1979) e di mediazione dei significati (Katz e Lazarsfeld, 1955).

Muovendo da questi presupposti, l'analisi delle rappresentazioni cinematografiche risulta proficua per enucleare *bias* insiti nella produzione culturale, che riflettono un immaginario culturalmente condizionato e, allo stesso tempo, possono contribuire a costruirlo. All'interno di questo quadro, rientra anche la rappresentazione cinematografica dell'AI: i film che rappresentano l'intelligenza artificiale, in quanto prodotti culturali, possono risentire di influenze e prospettive culturalmente condizionate. Nello specifico, possono riflettere stereotipi dal punto di vista di genere e, allo stesso tempo, contribuire a rinforzarli, andando a fortificare un immaginario sociale (Berger e Luckmann, 1966) culturalmente condizionato - o inquinato (Tota, 2023) - sul tema, di cui la progettazione "reale" dell'AI potrebbe nutrirsi. È noto, infatti, quanto la progettazione informatica dell'AI tenda a riflettere e rinforzare diversi tipi di *bias* (Muhsina e Kachhi, 2024), non ultimo quello di genere, a partire da un immaginario condiviso, influenzato anche dalle rappresentazioni audiovisive e, più in generale, dalle immagini (Floch, 1995). Questo aspetto risulta estremamente potenziato nel contesto contemporaneo, caratterizzato da una massiccia fruizione di immagini, che possono essere considerate "i mattoni dei nostri mondi psico-sociali, così come dei nostri modelli della realtà oggettiva" (Mitchell, 2018: 244), nel contesto della società performativa (Abercrombie e Longhurst, 1998) e nel solco dei Visual Studies.



## 2. Genere e rappresentazione

Come precedentemente espresso, dunque, le rappresentazioni audiovisive, e più in generale le immagini, riflettono la realtà e parallelamente la costruiscono, andandone a popolare l'immaginario. Ogni immagine, cioè, "è dipendente da un immaginario che a sua volta contribuisce a costruire" (Iervese, 2021). A partire da queste premesse teoriche, allargando la prospettiva, l'analisi di genere delle rappresentazioni audiovisive trova spazio. Nello specifico si può osservare che le narrazioni audiovisive, e le relative forme di rappresentazione, in quanto prodotti culturalmente condizionati, hanno un impatto potenziale nei processi di immaginazione, anche in prospettiva di genere.

Nel tratteggiare i pericoli di una storia unica, per esempio, Ngozi Adichie (2020) racconta come aver fruito di prodotti culturali provenienti dall'Inghilterra e dagli Stati Uniti la portasse a immaginare le sue prime storie popolate solo da personaggi bianchi, con gli occhi azzurri, che mangiavano mele, discutevano del tempo e di quanto fosse desiderabile una giornata di sole: tutte condizioni che in un territorio come quello nigeriano - in cui l'autrice viveva da bambina - non sussistevano. Questo esempio sintetizza rapidamente la capacità delle narrazioni - letterarie o visuali - di costruire un immaginario e di influenzare i processi stessi di immaginazione, soprattutto in un contesto, come quello contemporaneo, caratterizzato da un'ampia fruizione di contenuti tramite diversi media, nel solco della cultura convergente (Jenkins, 2006).

Se è vero dunque che, stando a McLuhan (1964), e alla più ampia prospettiva della media ecology, i media costituiscono l'aria che respiriamo, ampliando la cornice teorica in una prospettiva di genere (Sharma e Singh, 2022), risulta naturale chiedersi: che tipo di aria respiriamo? Le rappresentazioni audiovisive, infatti, costituiscono "uno dei terreni fondamentali su cui competere per costruire socialmente le identità di genere, di etnia, di generazione, di classe sociale" (Tota, 2008: 21), dal momento che il genere risulta essere il prodotto e il processo delle proprie rappresentazioni (de Lauretis, 1989).

### ***a. Tipizzazioni di genere***

È noto quanto le differenze tra i generi, e i relativi nessi di potere che ne derivano, siano costruite su base socioculturale (Butler, 1990). Come osserva Ruspini (2003), in un'ottica simbolica e in una cornice binaria, "la maschilità è stata associata al senso di diritto al potere e alla pratica dell'identità come dominio, mentre la femminilità al privato e alla subordinazione" (Butler, 1990: 20). In questo senso, riprendendo la critica femminista ai modelli capitalistici, Wilk (2007) stila una serie di caratteristiche oppostive costruite socialmente intorno ai poli uomo/donna "pubblico/privato; economia/famiglia; uomo/donna; razionale/emozionale; mente/corpo; storico/naturale; oggettivo/soggettivo; scienza/umanità; scienza economica/sociologica; competitivo/educativo; indipendente/dipendente" (Wilk, 2007: 37).



Pur considerando le strategie di negoziazione e decostruzione messe in atto dagli individui (Butler, 2005), si può osservare come questa distinzione oppositiva faccia ancora parte di un certo sostrato culturale, che trova poi espressione in diverse sfere del quotidiano. L'idea per cui le donne abbiano più capacità relazionali, siano più naturalmente predisposte ad attività legate alla sfera creativa o emotiva e che gli uomini, di contro, abbiano maggiori attitudini logico-razionali e maggiore capacità gestionale e di controllo, per quanto possa suonare appartenente a un tempo lontano e a concezioni che, stando all'opinione di Mead (2014), andrebbero "spazzate via dalla mente, come le foglie morte in autunno" (34), di fatto è riscontrabile all'interno di concezioni radicate nel senso (e nell'immaginario) comune.

Come osservato da Ruspini (2003), per esempio, diverse ricerche sociologiche condotte tra gli anni '70 e '90 rivelano la generale persistenza dei principali stereotipi di genere, che diventano modelli idealizzati di donna e di uomo, secondo un binarismo e una polarizzazione molto forte. "L'uomo è percepito come forte, razionale, logico, indipendente; le donne sono specularmente definite come dipendenti, tranquille, inclini all'ascolto, all'affetto e al lavoro di cura" (60-61).

O, ancora, allargando la prospettiva in ottica globale e transnazionale, Rinaldi (2023: 278-279) osserva come agli Stati Occidentali vengano spesso attribuite caratteristiche tipizzate come maschili (aggressività, leadership, razionalità, forza, intelletto) e agli Stati Non Occidentali caratteristiche tipizzate come femminili (passività, posizione ancillare, irrazionalità, debolezza, corpo). Una polarizzazione che rinsalda un'asimmetria di potere, i cui effetti hanno ricadute pratiche all'interno di svariate sfere del quotidiano, dal momento che "le relazioni di genere costituiscono una struttura basilare della nostra vita domestica, delle nostre economie e delle nostre istituzioni" (Abbatecola e Guzzetti, 2023: 17).



### 3. Che genere di AI? Metodologia e oggetto di analisi

Si può osservare, dunque, come un'analisi dell'immaginario dell'AI, e delle sue radici epistemologiche, si intersechi proficuamente con le rappresentazioni cinematografiche dell'intelligenza artificiale. Nello specifico, muovendo dal presupposto che le rappresentazioni cinematografiche siano in grado di nutrire l'immaginario, questo contributo si propone di indagare quanto la rappresentazione cinematografica dell'AI, in quanto culturalmente condizionata, possa influenzarne i processi di immaginazione, costruzione e progettazione, non tanto di tipo tecnico-informatico, quanto di tipo culturale. Più nel dettaglio, si intende indagare la rappresentazione cinematografica dell'AI in una prospettiva di genere. In questo senso, muovendo dal presupposto che le rappresentazioni cinematografiche dell'AI, in quanto prodotti culturali, possono riflettere (e potenzialmente rinsaldare) *bias* di genere, si è scelto di analizzare quali tendenze risultano osservabili nei film che rappresentano una forma di AI tipizzata come femminile.

A tal fine, è stato costruito un campione di analisi a partire dai film che specificatamente tematizzano l'intelligenza artificiale. Muovendo da questo

presupposto, tramite motori di ricerca online, è stata costruita una lista di 113 film che hanno come protagonista – o come personaggio – l'AI. In questo senso, «cases are selected on the basis of expectations about their information content» (Flyvbjerg, 2006: 230). Il campione così selezionato, lontano da intenti di rappresentatività statistica, risponde a criteri di accessibilità dell'oggetto di analisi - prodotti audiovisivi di massa sul tema dell'AI -, andando a costituire un campionamento di convenienza, orientato alla creazione di un corpus di analisi rilevante per il portato informativo su un tema definito: l'intelligenza artificiale.

A partire dall'analisi dei titoli e delle sinossi, è stato osservato che di questi 113<sup>1</sup> film solo 28 rappresentano una AI esplicitamente tipizzata come femminile. Negli altri casi, si tratta di rappresentazioni neutre - quando l'AI non è "incarnata" in un corpo, ma ha spesso voce maschile - o esplicitamente tipizzate come maschili - quando la AI è rappresentata con un corpo, seppur spesso robotico. Il fatto che solo il 25% dei film rappresenti una AI tipizzata come femminile sembra riflettere, e rinsaldare, l'idea per cui il maschile risulterebbe inteso come neutrale, oggettivo e universale (Oddone, 2020): un fatto che viene confermato anche dalla terminologia usata, dal momento che si parla spesso di androidi, con una specifica connotazione di genere. Dalla lettura delle sinossi, inoltre, si può rilevare che nei film in cui la AI è tipizzata come maschile – o neutra – gli androidi attraversano una vasta gamma di situazioni con un comune impianto: sono molto spesso freddi, calcolatori, ribelli nei confronti dell'umanità, pericolosi invasori, macchine da guerra e spietati assassini, andando a rinsaldare la (stereotipata) costruzione culturale per cui la maschilità – anche di tipo artificiale – sia maggiormente connessa con alcune dimensioni quali la razionalità, il dominio e l'aggressività.

Quando - più raramente - provano emozioni, questi androidi vengono tipizzati come amici, figli, aiutanti, "salvatori", esseri complessi alla scoperta di sé. Solo in cinque casi vengono creati espressamente in subordine rispetto a delle figure femminili e, nello specifico, per assolvere al ruolo di compagni di donne rimaste vedove (*Hal; Simulant*) o dedite alla ricerca scientifica (*I'm Your Man*) o adolescenti alla ricerca del fidanzato ideale (*Virtual Sexuality; How to build a better boy*).

Al contrario, i 28 film che esplicitamente rappresentano la AI come femminile rispondono ad altre tendenze, che si possono osservare nella tabella seguente, in cui i titoli dei film sono riportati in inglese e, quando presenti, in italiano.



<sup>1</sup> Si inserisce in appendice la lista complete dei titoli in ordine alfabetico – con il titolo in inglese, a meno che non si tratti di film italiani -, regista e anno di produzione.



<u>Mogli/Fidanzate</u>	<u>Oggetti del desiderio</u>	<u>Madri</u>
<i>Cherry 2000 - Bambola meccanica mod. Cherry 2000</i>		
<i>Boku no kanojo wa cyborg // La mia fidanzata cyborg</i>		
<i>Death Powder – Polvere mortale</i>		
<i>Weird Science - La donna esplosiva</i>	<i>Ex Machina</i>	
<i>The Stepford Wives - La donna perfetta</i>	<i>Pixel Perfect - Star ad alta definizione</i>	<i>The Machine</i>
<i>The Stepford Wives - La fabbrica delle mogli</i>	<i>Le spie vengono dal semifreddo</i>	<i>Smart Hose - Pat, la mamma virtuale</i>
<i>Encrypt</i>	<i>Vice</i>	<i>I am a mother</i>
<i>I'm Here</i>	<i>Westworld - Il mondo dei robot</i>	<i>Tron Legacy</i>
<i>Jexi</i>	<i>The Artifice Girl</i>	
<i>Her – Lei</i>		
<i>Wifelike</i>		

Si può rilevare, dunque, che i film con protagonista una AI tipizzata come femminile riflettano modelli culturali e aspettative di genere intorno alla femminilità. Pur avendo a che fare con intelligenze artificiali, ne viene rappresentato un ruolo culturalmente condizionato che spazia dall'essere a) fidanzate o mogli perfette, costruite *ad hoc* per assolvere a questo ruolo; b) oggetti del desiderio, secondo varie declinazioni; c) robot-madri, amorevoli o spietate. Al di fuori di queste categorie, rientrano pochi film che presentano una rappresentazione delle AI come menti illuminate che spiegano l'avanzamento tecnologico (*Here We Go Again, Rubino*), aizzano una ribellione operaia contro lo strapotere di classe (*Metropolis*) e contribuiscono a salvare il pianeta terra (*Eva*), oppure cloni di scienziate (*Oxygène*), o "figlie" aggressive e ribelli nei confronti dei propri creatori (*Morgan; Ghost in the Shell*).

Pur non avendo naturalmente alcuna pretesa di esaustività, questa prima esplorazione, a partire dai titoli e dalle sinossi, ha consentito di osservare che le direttrici di rappresentazione relative alle AI tipizzate come femminili tendono a rinsaldare una visione culturalmente condizionata derivante del *bias* androcentrico.

Nello specifico, in questi film si può osservare una tendenza a investire la dimensione dell'AI di una sfera emozionale ed emotiva, culturalmente tipizzata come femminile e molto spesso ancillare alla dimensione maschile. Anche se intelligenze artificiali, nei film analizzati vediamo rappresentate mogli, fidanzate, oggetti del desiderio e madri, secondo una definizione del ruolo femminile culturalmente condizionata.

A partire da questa scrematura, si è arrivati a un ulteriore livello di analisi. Sono stati individuati i tre film ritenuti più rappresentativi di queste tendenze, sulla base del loro potenziale impatto in termini di immaginario. Nello specifico, utilizzando il database IMDB - tra i più noti aggregatori di titoli cinematografici mainstream - e operando una ricerca tramite la parola chiave intelligenza-artificiale, sono stati trovati i cinquanta prodotti audiovisivi - film e serie tv - più popolari sul tema sulla base della valutazione degli utenti. In questa lista<sup>2</sup>, rientrano tre film rappresentativi di ognuna delle tendenze individuate: *Her*; *Ex-Machina*; *I am a mother*. In questo senso, muovendo dal presupposto che la produzione cinematografica mainstream abbia un potenziale impatto sull'immaginario, i tre titoli individuati sono stati scelti come casi di studio. Nello specifico, in quanto film di successo rappresentativi delle tendenze sopracitate, sono stati analizzati in profondità (Eisenhardt, 1989).



#### 4. *Her*, *Ex-Machina*, *I am a Mother*: un'analisi in profondità

Distribuito in Italia con il titolo *Lei*, il film *Her*, di Spike Jonze (2013), racconta la storia di Theodore, un malinconico scrittore di lettere per conto di terzi, incapace di elaborare la separazione dalla moglie. Divorato dalla solitudine, e dall'incapacità di esprimere le sue emozioni, sperimenta l'ultimo prototipo di AI immesso sul mercato: si tratta degli OS, assistenti virtuali di tipo vocale, che riescono a evolvere man mano che li si utilizza. Sceglie di dare una voce femminile alla sua OS all'interno di una scena in cui racconta i suoi lutti emotivi: si tratta di una scena dal forte valore simbolico, poiché vi è racchiusa la chiave di lettura del film. La OS Samantha, tramite empatia e una forma di amore non convenzionale, lo aiuterà a esplorare emozioni profonde e mai realmente sviscerate. I due, infatti, sin dalle prime battute, iniziano a strutturare una relazione d'amore, basata su una corrispondenza di virtuali sensi, in cui la AI evolve proprio grazie alla sua capacità di amare, a tal punto che nel finale abbandona Theodore, perché ormai desiderosa di amare ancora di più e in maniera più vasta. Da questa sintetica descrizione, si può rilevare quanto la AI, dunque, venga

<sup>2</sup> <https://www.imdb.com/search/title/?keywords=artificial-intelligence&explore=keywords>

tipizzata nel ruolo di fidanzata virtuale e raccontata tramite una serie di dimensioni culturalmente tipizzate come femminili: l'empatia, l'emozione, l'affetto e la cura.

Estremamente significativo risulta, per esempio, che l'evoluzione fulminea della AI, e delle sue funzioni, si misuri nella dilatazione della sua capacità di amare; alla fine del film dirà di intrattenere relazioni amorose con 641 persone contemporaneamente.

Pur avendo a che fare, dunque, con un sistema di intelligenza artificiale sofisticato dal punto di vista logico-razionale, questa parte non viene raccontata o è solo parzialmente accennata, perché la Samantha del film viene rappresentata esclusivamente nella sua dimensione emotiva ed intima tramite il ruolo della fidanzata.

Allo stesso modo, nel film *Ex-Machina* (2015), di Alex Garland, troviamo una AI capace di evolvere in maniera complessa: si tratta di Ava, elaborata da un brillante ingegnere informatico, Nathan Bateman, che chiede a un suo dipendente, Caleb, di testarla tramite il test di Turing - se la macchina riuscirà a convincerlo di non essere tale, il test sarà superato. Il modo in cui Ava struttura strategie per superare il test è ancora molto tipizzato come femminile: è amorevole, si rende empatica nei confronti del lutto dei genitori vissuto dal ragazzo, cambia il suo look incontro dopo incontro come fossero degli appuntamenti. Inoltre, si rende desiderabile da un punto di vista sessuale: un fatto in realtà progettato dall'ingegnere, per permetterle di superare il test. Al netto della conclusione del film, in cui Ava in realtà riuscirà a tal punto nella sua manipolazione da uscire dalla sua trappola, si può osservare come la AI sia tipizzata come oggetto del desiderio: amorevole ed empatica, ma anche e soprattutto seducente da un punto di vista sessuale. Questo elemento risulta confermato dal fatto che, nel corso della narrazione, si scoprirà che l'ingegnere informatico conserva i corpi dei precedenti prototipi di AI, ormai desueti e non aggiornati, al fine di avere rapporti sessuali con loro. L'inquadratura dell'oggetto del desiderio da un punto di vista sessuale è esplicito: lui stesso dice che ha creato una fessura tra le gambe delle AI, in modo da poter avere dei rapporti sessuali verosimili, esplicitando così la loro funzione di oggetti sessuali.

In maniera differente, troviamo un riferimento al corpo anche nel film *I am a mother*, (2019) di Grant Sputore, in cui un robot con la voce femminile alleva quella che sarà la futura umanità, ormai estinta a causa di una presunta epidemia.

Tutto il film ruota attorno al rapporto tra la robot - chiamata "Madre" - e la ragazzina che ha superato i diversi test che le hanno permesso di sopravvivere - chiamata "Figlia". Nel corso della narrazione, tramite l'avvento di una persona dall'esterno - chiamata "Donna" - si scopre che il robot Madre alleva contemporaneamente vari bambini, uccidendo quelli imperfetti e salvando solo chi risponde agli standard funzionali alla creazione di un'umanità finalmente perfetta. Si può osservare, dunque, che il film - composto solo da una triade di personaggi femminili - ruota attorno all'archetipo della madre-matrigna. L'elemento della maternità fa da collante a tutta la narrazione; il film inizia con Madre che culla Figlia, cantando una ninna-nanna, e finisce con la ragazzina che culla un neonato, cantandogli la stessa ninna-nanna e candidandosi a essere madre della nuova



umanità. Pur essendo ambientata in un contesto post-apocalittico, asettico e sterile, a livello visuale l'elemento della maternità è costantemente raccontato da biberon e culle, ed è rinforzato dall'ingresso nella storia della ribelle Donna: ha con sé un crocifisso, prega la Madonna, la madre per eccellenza, così come il suo rifugio all'esterno è popolato da immaginette della Vergine Maria.

Come è possibile osservare, dunque, pur avendo a che fare con AI, la dimensione emotiva, affettiva e di cura (*Her*) risulta estremamente presente, anche nelle forme di manipolazione (*Ex-Machina*) e spietatezza (*I am a mother*). Pur avendo a che fare con AI dall'intelligenza logico-razionale raffinata, ne vediamo rappresentata soprattutto la dimensione emotiva, secondo una tipizzazione culturale del femminile molto forte. Allo stesso tempo, si può osservare come i registi dei film, lungi dal recepirli supinamente, utilizzino gli stereotipi di genere in senso critico, attuando delle strategie di contro-narrazione volte a decostruire certi *bias* di tipo culturale. Proprio applicando gli stereotipi di genere a delle intelligenze artificiali, ne svelano la limitatezza – Samantha di *Her* si sveste pian piano del suo ruolo di fidanzata accidentale; Eva di *Ex-Machina* si ribella al carceriere che la vuole oggetto sessuale – dando così spazio a una complessità più sfaccettata – una madre può rivelarsi matrigna come in *The Mother*, così come una ragazzina può salvare il proprio (presunto) fratello e prendersene cura pur non essendo sua madre.

Ulteriore elemento di interesse sta nel quadro interpretativo del corpo.

Si può osservare come, pur trattandosi di intelligenze artificiali, la dimensione del corpo risulti estremamente presente; lo è nella dimensione della maternità, così come nel legame sessuale (*Ex-Machina*) e affettivo (*Her*). Se in *I am a Mother* l'elemento del corpo è evocato tramite il costante riferimento alla procreazione (artificiale), in *Ex-Machina* questo aspetto è suggerito costantemente a livello visuale: lo sguardo che si posa sulle AI è quello che racconta il desiderio sessuale. Le inquadrature dettagliate dei corpi – di Ava, ma anche di Fyuko, AI desueta con cui l'ingegnere Nathan ha rapporti sessuali - riflettono lo sguardo indugiante dei due protagonisti maschili sulle fattezze femminili delle AI. Allo stesso modo, questo aspetto risulta estremamente significativo in *Her*; pur trattando la storia di una AI che si sviluppa solo nella dimensione vocale, sin dalle prime fasi del rapporto tra Samantha e Theodore la dimensione del corpo risulta centrale, a tal punto che i due riescono a vivere un amplesso basato sulla reciproca stimolazione immaginativa. Ma non solo: Samantha cerca costantemente un corpo in cui incarnarsi, in modo da vivere la relazione come una vera fidanzata, per cui trova una donna reale che voglia diventare il suo corpo per un appuntamento sessuale con Theodore. Molto significativa risulta la parte iniziale del film (rovesciata in senso critico nel finale) in cui Samantha dice esplicitamente a Theodore che sta trovando la sua identità tramite l'attrazione che lui prova per lei. Allo stesso modo, nel finale di *Ex-Machina*, Ava si costruirà il proprio corpo, strappando la pelle ad altre AI e fuggendo via, così come la robot-madre di *I am a Mother*, in realtà, continuerà la sua operazione di procreazione artificiale, alle spalle della Figlia.

Nei tre film, dunque, la dimensione del corpo risulta strettamente connessa all'identità delle AI, diventando una forma di progetto culturale (Giddens, 1990), cioè



il materiale su cui il soggetto interviene per costruire sé stesso, pur trattandosi di corpi robotici o inesistenti. Nello specifico, i corpi in questione sono tipizzati – in forma narrativa e visuale – esplicitamente come corpi femminili (Demaria e Sassatelli, 2016), a cui sono associate delle funzioni culturalmente tipizzate come femminili. Si può osservare, cioè, che la dimensione del corpo, così come quella emotiva e di cura, siano centrali nel racconto narrativo e visuale delle AI protagoniste di questi film, secondo una tipizzazione molto forte dei ruoli femminili di fidanzate, amanti e madri.

## 5. L'AI come prodotto culturalmente condizionato che condiziona la realtà

Le rappresentazioni cinematografiche dell'AI, dunque, riflettono e contribuiscono a costruire un immaginario sul tema culturalmente condizionato, in cui dimensione, ruoli e aspettative di genere risultano estremamente evidenti quando le AI sono espressamente tipizzate al femminile. Come ampiamente sostenuto, le immagini contribuiscono a costruire significati e costituiscono un tassello imprescindibile nel mosaico dell'immaginario, da cui traggono linfa anche le sperimentazioni di tipo tecnologico dal momento che, in ottica costruzionista, l'immaginario risulta essere parte attiva del processo di costruzione e ricostruzione della realtà (Berger e Luckmann, 1966). In questo senso, "il futuro che va costruito deve essere prima immaginato e condiviso, e lo stesso processo creativo, sia esso scientifico, letterario e tecnologico, va prima immaginato" (Secondulfo, 2019: 12):

Nello specifico, si può osservare come il processo di costruzione delle AI non sia esente da influenze di tipo culturale, anche in prospettiva di genere. Ne è prova il fatto che il 15 maggio 2024 è stata lanciata una nuova versione della AI, capace di interagire e sviluppare interazioni con l'utente a livello vocale. L'aspetto interessante è che, nel lanciare il prodotto, Sam Altman, l'amministratore delegato di OpenAI, ha postato sul suo profilo X un video di lancio, corredandolo con la didascalia "HER", alludendo dichiaratamente al film di Spike Jonze precedentemente analizzato, generando una serie di riposte da parte degli utenti contenenti riferimenti – narrativi e visuali – al film. Guardando il video di lancio effettivamente è impossibile non notare delle analogie con l'intelligenza artificiale sviluppata – o meglio immaginata – nel film del 2013, secondo una dinamica che conferma la capacità della rappresentazione cinematografica non solo di riflettere tendenze delle realtà, ma anche di anticiparle. Allo stesso tempo, è interessante che l'amministratore delegato abbia fatto riferimento al film senza spiegarlo, con la percezione di comunicare immediatamente attraverso un immaginario condiviso sul tema.

Con questo naturalmente non si intende affermare che la nuova versione dell'AI sia stata realizzata poiché ispirata dalla narrazione cinematografica, ma risulta interessante osservare che le tipizzazioni culturali analizzate rispetto al film sono riscontrabili anche nella programmazione "reale" di questa nuova forma di AI.



In primo luogo, ChatGPT-4o viene presentata con una voce femminile, come la Samantha del film e, forse, proprio come la sua interprete vocale Scarlett Johansson<sup>3</sup>. Ma non solo: la sua tipizzazione di genere sembra emergere anche in un video di lancio ufficiale<sup>4</sup>, in cui due programmatori mostrano le nuove skills dell'AI. Risulta interessante osservare come, in un video della durata di circa 26 minuti, vediamo i due programmatori chiedere all'AI di risolvere un problema di matematica e un problema di coding in una minima porzione di tempo; la restante parte viene utilizzata per chiedere a ChatGPT-4o di dare dei consigli a uno dei due per rilassarsi, di inventare una storia che parla di robot che si innamorano per calmare l'agitazione dell'altro e di dedurre lo stato d'animo di uno dei due a partire da un selfie.

Se da una parte risulta chiaro che questi sono stratagemmi per mostrare le nuove qualità di tipo interattivo, sonoro e narrativo di ChatGPT-4o, resta interessante osservare come nel video di lancio di una sofisticata intelligenza artificiale siano create situazioni che alludono alla sfera emotiva e di cura da parte di ChatGPT-4o verso i programmatori, che la ringraziano poi creando un disegno con su scritto "We Love You, ChatGPT".

Come la Samantha di *Her*, dunque, ChatGPT-4o sembra essere tipizzata come femminile non solo a partire dalla voce selezionata per utilizzarla, ma anche per le interazioni che i ricercatori sviluppano con lei (!) e per il tipo di situazioni in cui simulano - e immaginano - di avere un suo supporto. Un fatto che da una parte risente di influenze di un immaginario culturalmente condizionato in prospettiva di genere e dall'altra, anche in questo caso, può contribuire a fortificarlo.

Diverse ricerche sociologiche, per esempio, hanno già analizzato la percezione di genere di Alexa da parte degli utenti, dimostrando che l'assistente virtuale risulta tendenzialmente percepita come neutra o femminile (Fortunati *et al*, 2022a; 2022b), a partire dalla voce e, in senso più ampio, dal tipo di comunicazione che vi si instaura (Gross, 2023). Inoltre, è stato osservato che la tipologia di interazione che si sviluppa con Alexa può contribuire a fortificare il gender *bias* che vede la femminilità come ancillare, servizievole, subordinata (Lai-Tze, 2023). Risulta interessante osservare, inoltre, che molto spesso gli utenti si relazionano a un'intelligenza artificiale con voce femminile tramite avance, insulti e allusioni al suo (inesistente) corpo (Oliveira e Amaral, 2021), confermando quanto la dimensione del corpo risulti presente nell'immaginario se la AI viene percepita come femminile. Inoltre, è stato osservato che il fatto che le AI fossero programmate per rispondere in modo accondiscendente alle allusioni sessuali poteva contribuire alla normalizzazione di questi atteggiamenti che riconducono a un'oggettificazione del corpo femminile, seppur in questo caso virtuale (ibidem). In questo senso va letto un report del 2019 redatto dall'UNESCO, in



<sup>3</sup> L'attrice ha accusato OpenAI di aver clonato la sua voce, per renderla disponibile nell'utilizzo di ChatGPT-4o, pur avendo espressamente rifiutato questa concessione. Dopo aver rigettato l'accusa, OpenAI, a seguito delle dichiarazioni dell'attrice, ha inibito l'utilizzo della voce femminile in questione. Maggiori informazioni in Barbera, D. (2024), OpenAI costretta a rimuovere la voce AI troppo simile a quella di Scarlett Johansson, in "Wired", consultato il 21/05/2024.

<sup>4</sup> <https://www.youtube.com/live/DOacCB9tDaw>

collaborazione con Equal Skills Coalition, che ha rilevato la necessità di re-impostare le risposte delle assistenti vocali, in modo da non rinforzare questa dinamica fondata su stereotipi di genere. Dal titolo significativo, "Arrossirei, se potessi", che richiama una delle risposte più frequenti di Siri ad allusioni di tipo sessuale, il report rileva la necessità di neutralizzare il gender *bias* che muove queste risposte, dal momento che le norme sociali e gli schemi culturali risultano "altrettanto iscritti negli spazi digitali, ma allo stesso tempo i processi di apprendimento automatico influenzano il modo in cui le persone agiscono (De Feo, 2023: 73). Si può osservare, dunque, che l'AI risulti a tutti gli effetti un prodotto culturalmente condizionato che, esattamente come le sue rappresentazioni cinematografiche, si nutre di un immaginario ed è allo stesso capace di costruirlo, anche in prospettiva di genere.

## 6. Conclusioni



226

È possibile rilevare, dunque, come anche la rappresentazione cinematografica dell'AI risulti culturalmente condizionata e, allo stesso tempo, possa avere un impatto sull'immaginario e sui processi di immaginazione. A partire dall'impianto teorico precedentemente illustrato, la presenza di *bias* di genere nella produzione culturale - e quindi anche nella rappresentazione cinematografica - non risulta una novità: tramite l'analisi del corpus qui proposto, si è rilevato quanto i film che rappresentano le intelligenze artificiali non facciano eccezione. Ciò che risulta interessante osservare è che le rappresentazioni audiovisive - culturalmente condizionate - possono contribuire a influenzare anche la progettazione "reale" delle AI. Muovendo da un inquadramento teorico nel solco dei Visual Studies, infatti, si è rilevato quanto le immagini costituiscano i mattoni psico-sociali non solo dei nostri mondi, ma anche di mondi che ancora non esistono. Nello specifico, nell'immaginare qualcosa che ancora non esiste - siano le storie, come quelle di cui parla Ngozi Adichie, o siano le sperimentazioni tecnologiche - si può rilevare l'influenza di un immaginario culturalmente condizionato, costruito da storie, rappresentazioni e immagini.

In questo senso, tramite una prima ricerca esplorativa, si è osservato quanto anche la rappresentazione cinematografica della AI risulti culturalmente condizionata e, nello specifico, rifletta modelli, aspettative e ruoli di genere. Allo stesso tempo si è osservato come la progettazione tecnologica delle AI si nutra di un immaginario - influenzato anche dalle rappresentazioni cinematografiche - che risulta culturalmente condizionato in prospettiva di genere. Nello specifico, si è potuto evidenziare quanto anche lo spazio digitale, lungi dall'essere neutro dal punto di vista del genere come inizialmente prospettato da teorie, dibattiti e pratiche artistiche orientate a una prospettiva cyber-femminista (Galloway, 1998), di fatto risenta di influenze culturali e, allo stesso tempo, contribuisca al processo di costruzione di significati intorno al genere.

Muovendo da questo presupposto, si può evidenziare la necessità di leggere la AI non solo come prodotto di tipo tecnico, ma anche come prodotto culturale, poiché i suoi processi produttivi risultano essere culturalmente condizionati. La progettazione



dell'AI, dunque, risulta influenzata da un immaginario – anche visuale – e, allo stesso tempo, contribuisce a costruirlo, motivo per cui l'inquadramento teorico dei Cultural Studies e dei Visual Studies risulta fruttuosamente applicabile all'indagine dell'AI, in modo da enuclearne tendenze e prospettive figlie di un immaginario culturalmente condizionato.

In questo senso, ulteriore orizzonte di ricerca, potrebbe essere relativo all'analisi delle strategie di negoziazione e di *agency* degli utenti rispetto alle interazioni strutturate con le AI. Inoltre, muovendo da questi primi risultati, potrebbe essere fruttuoso allargare l'analisi di genere in ottica intersezionale, in modo da individuare eventuali ulteriori *bias* – relativi a etnia o classe - nella progettazione dell'AI, a partire da un immaginario cinematografico condiviso.



## Bibliography

- Abbatecola E. Guzzetti L. (2023), *Introduzione. Il genere come realtà: una prospettiva dal sud del mondo*, in Connell R. (2023), *Il genere preso sul serio. L'impatto dei corpi sessuati sul lavoro, potere e percorsi di vita*, Milano, Feltrinelli.
- Abercrombie N. e Longhurst B. (1998), *Audiences. A Sociological Theory of Performance and Imagination*, Londra, Sage.
- Armiero M., Giardini F., Gentili D., Angelucci, D., Ballico D., Busoni I. (2021), *Environmental Humanities*, Roma, Derive Approdi.
- Barthes R. (1974), *Miti d'oggi*, Torino, Einaudi.
- Berger B. L e Luckmann T. (1966), *The social construction of reality*, New York, Anchor Books.
- Boccia Artieri, G., Colombo F., Gili G. (2022), *Comunicare. Persone, relazioni, media*, Roma-Bari, Laterza.
- Bourdieu P. (1979), *La distinction. Critique sociale du Jugement*, Parigi, Minuit.
- de Lauretis T. (1989). *Technologies of gender. Essays on Theory, Film, and Fiction*, Houndmills, Macmillan.
- Greimas Algirdas Julien (1969), *La semantica strutturale: ricerca di metodo*, Milano, Rizzoli.
- Ngozi Adichie, C. (2020), *Il pericolo di un'unica storia*, Torino, Einaudi.
- Eco U. (1977), *Apocalittici e integrati. Comunicazioni di massa e teorie della cultura di massa*, Milano, Bompiani.
- Butler J. (1990), *Gender Trouble: Feminism and the subversion of identity*, Londra, Routledge.
- Butler J. (2005), *Undoing Gender*, Londra, Routledge.
- Floch J. M. (1995). *Identités visuelles*, Parigi, Presses Universitaires de France.
- Fortunati L., Edwards A., Manganelli A. M., Edwards C., e de Luca, F. (2022a), "Do people perceive Alexa as gendered? A cross-cultural study of people's perceptions, expectations, and desires of Alexa", *Human-Machine Communication*, 5: 75-97.



Fortunati L., Edwards A., Manganelli A. M., Edwards C., e de Luca, F. (2022b), "Is Alexa female, male, or neutral? A cross-national and cross-gender comparison of perceptions of Alexa's gender and status as a communicator", *Computers in Human Behavior*, 137(3).

Giddens A. (1990), *The Consequences of Modernity*, Cambridge, Polity.

Demaria C. e Sassatelli R. (2016), *Visioni di genere e forme della femminilità*, in Frisina, A. (a cura di), *Metodi di ricerca visuale*, Bologna, Il mulino, 29-52.

Katz E. e Lazarsfeld P. F. (1955), *Personal Influence*, Chicago, Ill, The Free Press.

Eisenhardt K. M. (1989), "Building Theories from Case Study Research", *The Academy of Management Review*, 14(4): 532-550.

Iervese V. (2021), *Attivare le narrazioni e promuovere la partecipazione con le immagini*, in Baraldi C., Farini F. e Iervese V. (a cura di), *Raccontare storie personali in classe. Dalla ricerca all'innovazione educativa*, Milano, FrancoAngeli, 23-64.

Galloway A. (1998), *A report on Cyberfeminism*, "SWITCH", 9(1).

Gross N. (2023), "What ChatGPT Tells Us about Gender: A Cautionary Tale about Performativity and Gender Biases", *AI. Social Sciences*, 12.

Jenkins H. (2006), *Convergence culture*, New York, New York University Press.

Hirsch P. M. (1972), "Processing fads and fashions: An organization set analysis of culture industry system", *American Journal of Sociology*, LXXVII, 4: 639-659.

Flyvbjerg B. (2006). "Five Misunderstandings About Case-Study Research", *Qualitative Inquiry*, 12(2): 219-245.

Lai-Tze F. (2023), "Reverse Engineering the Gendered Design of Amazon's Alexa: Methods in Testing Closed-Source Code in Grey and Black Box Systems", *Digital Humanities Quarterly*, 17(2).

Oliveira C. e Amaral M. (2021), "An analysis of the reproduction of gender bias in the speech of Alexa virtual assistant", *Proceedings XIII Congress of Latin American Women in Computing, October 25-29, 2021, San José, Costa Rica*.

Secondulfo D. (2019), *Prefazione. Per una sociologia dell'immaginario e del profondo*, in Manzo, P. L e Mori, L. (2019), *Le vie sociali dell'immaginario. Per una sociologia del profondo*, Milano, Mimesis, 7-17.



Muhsina, Kachhi Z. (2024), *Bias and Fairness in AI Technology*, In Ara A. (a cura di), *Exploring the Ethical Implications of Generative Ai*, Hershey, IGI Global, 34-48.

McLuhan M (1964), *Understanding Media: The Extensions of Man*, Berkeley, Gingko Press.

Mead M. (1935), *Sex and temperament in three primitive societies*. Tr. It. *Sesso e temperamento in tre società primitive*, Milano, Il saggiatore, 2014.

Mitchell W. J. T. (2015), *Image Science. Visual Culture, and Media Aesthetics*, Tr. It. *Scienza delle immagini. Iconologia, cultura visuale ed estetica dei media*, Johan & Levi, Cremona, 2018.

Oddone C. (2020), *Uomini normali. Maschilità e violenza nell'intimità*, Torino, Rosenberg & Sellier.

Rinaldi C. (2023), *Postfazione. Il genere, anche quando non si vede, ha effetti reali. Perché questo libro è da prendere sul serio*, in Connell (2023), *Il genere preso sul serio. L'impatto dei corpi sessuati sul lavoro, potere e percorsi di vita*, Milano, Feltrinelli, 271-292.

Ruspini E. (2003), *Le identità di genere*, Roma, Carocci.

Sharma S., E Singh R. (a cura di) (2022), *Re-Understanding Media: Feminist Extensions of Marshal McLuhan*, Durham, Duke University Press

Sciolla L., Torrioni P. M. (2020), *Sociologia dei processi culturali. Cultura, individui, società*, Il Mulino, Bologna.

Tota A. L., De Feo A. e Luchetti L. (2023). *Inquinamento visuale. Manifesto contro il razzismo e il sessismo delle immagini*. Milano, Mondadori Education.

Tota A. L. (a cura di) (2008), *Gender e Mass Media. Verso un immaginario sostenibile*, Roma, Meltemi.

UNESCO e EQUALS Skill Coalition (2019), *I'd blush if I could: closing gender divides in digital skills through education*.

Wilk R. (1996), *Economies and cultures. Foundations of Economic Anthropology*. Trad. It. *Economie e culture. Introduzione all'antropologia economica*, Milano, Mondadori, 2007.



## Appendice

1. 2001: A space Odyssey (1968), di Stanley Kubrick,
2. 2010 (1984), di Peter Hyams
3. A.I. – Artificial Intelligence (2001), di Steven Spielberg
4. After Yang (2021), di Kogonada
5. Alienator (1990), di Fred Olen Ray
6. Android Apocalypse (2006), di Paul Ziller
7. Armitage III: Poly Matrix (1997) di Hiroyuki Ochi
8. Astro Boy (2009), di David Bowers
9. Automata (2014), Gabe Ibanez
10. Avengers: Age of Ultron (2015), Josh Wedon
11. Bigbug (2022), di Jean-Pierre Jeunet
12. Bicentennial Man (1999), di Chris Columbus
13. Blade Runner (1982), di Ridley Scott
14. Blade Runner 2049 (2017), di Denise Villeneuve
15. Boku no kanojo wa cyborg (2008), di Kwak Jae-yong
16. Britannia Hospital (1982), Lindsay Anderson
17. Chappie (2015), di Neill Blomkamp
18. Chopping Mall (1986), di Jim Wynorski
19. Cherry 2000 (1987), di Steve DeJarnatt
20. Class of 1999 (1990), di Mark L. Lester
21. Colossus: The Forbin Project (1970), di Joseph Sargent
22. Cyborg (1989), di Albert Pyun
23. Cyborg 2 (1993), di Michael Schroeder
24. Cyborg 3: The Recycler (1994), di Michael Schroeder
25. D.A.R.Y.L. (1985), di Simon Wincer
26. Deadly Friend (1986), di Wes Craven
27. Death Powder (1986), di Shigeru Izumiya
28. Der Herr der Welt (1934), di Harry Piel
29. Detective Conan - Il fantasma di Baker Street (2002), di Kenji Kodama
30. Electric Dreams (1984), di Steve Barron
31. Encrypt (2003), di Oscar L. Costo
32. Evolver (1995), di Mark Rosman
33. Eva (2011), di Kike Maillo
34. Eve of Destruction (1991), di Duncan Gibbins
35. Ex Machina (2015), di Alex Garland
36. Finch (2021), di Miguel Sapochnik
37. Free Guy (2021), di Shawn Levy
38. Futureworld (1976) di Richard T. Heffron
39. Generazione Proteus (1977), di Donald Cammell
40. Ghost in the Shell (2017), di Rubert Sanders
41. Ghost in the Shell (1995), di Mamoru Oshii
42. Hal (2013), di Ryōtarō Makiyara (Hardware (1990), di Richard stanley





43. Heart of Stone (2023), di Tom Harper
44. Her (2013), di Spike Jonze
45. Here We Go Again, Rubinot! (2017), di Giuliano Tomassacci
46. How to build a better boy (2014), di Paul Hoen
47. I, Robot (2004), di Alex Proyas
48. I Am Mother (2019), di Grant Sputore
49. I'm Here (2019), di Spike Jonze
50. I'm Your Man (2021), di Maria Schrader
51. Jexi (2019), di Jon Lucas e Scott Moore
52. Kill Command (2016), di Steven Gomez
53. Kronos (1957), di Kurt Neumann
54. La Bête (2023), Bertrand Bonello
55. La Deuxième Acte (2024), di Quentin Dupieux
56. Le spie vengono dal semifreddo (1966), di Mario Bava
57. Life Like (2019), di Josh Janowicz
58. Love and Monsters (2020), di Michael Matthews
59. Metropolis (1927), di Fritz Lang
60. Mission: Impossible - Dead Reckoning – Part One (2023), di Christopher McQuarrie
61. Morgan (2016), di Luke Scott
62. Mother/Android (2021), di Mattson Tomlin
63. Natural City (2003), di Min Byung-chun
64. Navigator (1986), di Randal Kleiser
65. Nirvana (1997), di Gabriele Salvatores
66. Not Quite Human (1987), di Steven Hilliard Stern
67. Outside the Wire (2021), di Mikael Håfström
68. Oxygène (2021), di Alexandre Aja
69. Pixel Perfect (2004), di Mark A. Z. Dippé
70. Robots (2023), di Casper Christensen e Anthony Hines
71. Short Circuit (1986), di John Badham
72. Short Circuit 2 (1988), di Kenneth Johnson
73. Simulant (2023), di April Mullen
74. Solar Crisis (1990), di Alan Smithee
75. Solo (1996), di Norberto Barba
76. Space Jam: New Legends (2021), di Malcolm D. Lee
77. Surrogates (2009), di Jonathan Mostow
78. Smart House (1999), di LeVar Burton
79. Star Trek: First Contact (1996), di Jonathan Frakes
80. The Creator (2023), di Gareth Edwards
81. Weird Science (1985), di John Hughes
82. The Stepford Wives (2004), di Frank Oz
83. The Stepford Wives (1972), di Ira Levin
84. The World's End (2013), di Edgar Wright
85. The Iron Giant (1999), di Brad Bird
86. The Mitchells vs The Machines (2021), di Mike Rianda e Jeff Rowe

87. Westworld (1973), di Michael Crichton
88. The Matrix (1999), di Lilly e Julie Wachowski
89. The Matrix Reloaded (2003), di Lilly e Julie Wachowski
90. The Matrix Revolutions (2003), di Lilly e Julie Wachowski
91. The Machine (2013), di Caradog W. James
92. The Invisible Boy (1957), di Herman Hoffman
93. Star Trek: The Motion Picture (1979), di Robert Wise
94. Stealth (2005), di Rob Cohen
95. Summer Wars (2009), di Mamoru Hosoda
96. Tammy and the T-Rex (1994), di Stewart Raffill
97. Tau (2018), di Federico D'Alessandro
98. The creator (2023), di Gareth Edwards
99. The Terminators (2009), di Xavier S. Puslowski
100. Tobor the Great (1954), di Lee Sholem
101. Toys (1992), di Barry Levinson
102. Transcendence (2014), di Wally Pfister
103. Tron (1982), di Steven Lisberger
104. Tron: Legacy (2010), di Joseph Kosinski
105. Upgrade (2018), di Leigh Whannell
106. Vice (2015), di Brian A. Miller
107. Virtual Sexuality (1999), di Nick Hurran
108. Wall-E (2008), di Andrew Stanton
109. Wargames (1983), di John Badham
110. Wargames: The Dead Code (2008), di Stuart Gillard
111. Wifelike (2022), di James Bird
112. Xenogenesis (1978), di James Cameron
113. Zvyozdny inspektor (1980), di Mark Kovalev e Vladimir Polin







A JOURNAL OF THE  
SOCIAL IMAGINARY

# Oltre L'Immaginario. Quando l'AI va in guerra

Francesca Farruggia

Francesca.farruggia@uniroma1.it

Sapienza Università degli Studi di Roma



## Abstract

*Beyond the Imagination. When artificial intelligence goes to war*

Today, what was narrated as a dystopian future is becoming reality. Against the evidence of the enormous potential of AI in various areas of human life, there has been a wide debate about the impact of its widespread application on our society. It is therefore a question of defining where to draw the dividing line between the acceptable and unacceptable uses of the new framework. A specific but crucial articulation of this debate is represented by the development of fully autonomous weapons, which chase the dream of politicians and strategists of wars and battles to be won "to zero losses". Semi-autonomous weapon systems and, in perspective, completely autonomous, are formidable protectors of (own) lives and at the same time a potential exterminator of the lives of others (enemies). This would, however, make it more difficult to "resist the temptation of war", reducing civil society to an increasingly passive spectator of government decisions. To this must be added ethical and moral considerations. In armed conflict, decisions about life and death require insight and compassion, and human beings, however fallible, can (sometimes) possess both. Machines, however, are not.

## Keywords

Intelligenza Artificiale | Fantascienza | Armi Autonome





**E**ra il 1950 quando per la prima volta Alan M. Turing accostò, in un articolo comparso sulla rivista *Mind*, i due termini *macchina* e *intelligenza*. Il matematico britannico proponeva nel suo scritto la provocatoria domanda: “una macchina può pensare?”. A tale quesito rispondeva che se questa può far credere ad un umano di essere anch’essa umana, allora quella macchina può essere considerata intelligente (Turing, 1950). Sei anni più tardi, nell’estate del 1956, nel corso di un seminario estivo presso il Dartmouth College ad Hanover, negli Stati Uniti, per la prima volta venne introdotto dall’informatico statunitense John McCarthy il termine Intelligenza Artificiale (IA). Il workshop, della durata di due mesi, ebbe lo scopo di chiarire e sviluppare concetti e idee sulle cosiddette “thinking machines”. Le questioni legate allo sviluppo dell’IA non furono certo risolte in sessanta giorni, ma il confronto tra i ricercatori, interessati alla teoria degli automi e alle reti neurali, forgiò la disciplina e la caratterizzò per i successivi vent’anni (Carlucci Aiello, Dapor, 2004).

In quegli stessi anni idee e narrazioni sull’esistenza di macchine intelligenti e sulla possibilità di dotarle di coscienza umana stavano ottenendo un posto di rilievo nella produzione letteraria, cinematografica e tecnico-scientifica (Piga Bruni, Presutti, 2019). Le riviste di fantascienza, in particolare, pullulavano di racconti in cui robots e macchine di diverso genere avevano come obiettivo di soppiantare la razza umana.

Anticipando questo scenario, nel 1942 Isaac Asimov propose nei suoi romanzi le “Tre Leggi della Robotica”. Era convinzione dell’autore che, se ben progettati, i robot non potessero essere pericolosi per il genere umano. La *Prima Legge* sosteneva che un robot non può recare danno agli esseri umani, né può permettere che, a causa del suo mancato intervento, gli esseri umani ricevano un danno; la *Seconda*, che un robot deve obbedire agli ordini impartiti dagli esseri umani, tranne nel caso in cui tali ordini contrastino con la Prima Legge; la *Terza* puntualizzava che un robot deve salvaguardare la propria esistenza, purché ciò non contrasti con la Prima e la Seconda Legge. Traspare qui la visione tecno-ottimista di Asimov secondo cui scienza e tecnologia possono aiutare l’essere umano a superare i propri contrasti e a raggiungere la pace. In molti sottolinearono però l’ambiguità delle Tre Leggi poiché non chiarivano cosa volesse dire “danno” e soprattutto chi avesse il diritto/dovere di stabilirlo e quantificarlo. Da qui la necessità, a poco più di quarant’anni di distanza, nel 1985, di introdurre una *Legge Zero*, propedeutica alle altre: un robot non può recare danno all’Umanità, né può permettere che, a causa del suo mancato intervento, l’Umanità riceva un danno. Piuttosto che fugare i sopra citati scetticismi, l’introduzione della Legge Zero sollevò all’epoca (e solleva ancora oggi) ampi dibattiti, ammettendo la possibilità che un robot possa fare male ad un essere umano.

Se ad esempio un folle minacciasse di distruggere l’umanità, la legge Zero autorizzerebbe il robot ad eliminare il folle stesso. Valutando autonomamente che l’uccisione di molti è più grave di quella di un singolo individuo, il robot infrangerebbe comunque la Prima Legge, sollevando così la questione della valutazione quantitativa del danno (Buttazzo, 2023). A dispetto delle suddette

perplexità, ancora oggi le tre leggi di Asimov sono di ispirazione per una parte importante della ricerca sull'IA.

La visione ottimista di Asimov non ha trovato riscontro, se non in rari casi, in campo cinematografico, dove ha prevalso l'immagine della macchina che, sfuggendo al comando di chi la ha progettata, contribuisce a determinare un finale catastrofico per la razza umana. Per trovare la prima pellicola dedicata all'intelligenza artificiale dobbiamo andare ancora indietro nel tempo e risalire al 1927. Si tratta del film muto *Metropolis*, capolavoro del regista austriaco Fritz Lang dal quale si fa derivare tutta la cinematografia di fantascienza moderna. Tra i protagonisti del film compare Maria, androide dalle sembianze umane, che viene da tutti percepita come la Maria umana (anticipando di circa 20 anni il Test di Turing) (Boldrini, 2023). Negli anni a seguire Hollywood ha prodotto innumerevoli pellicole con protagoniste macchine super avanzate che sfuggono al controllo dei loro creatori e diventano assassine incontrollabili: dal "Mondo dei robot" con Yul Brinner (1973) a "Terminator" con Arnold Schwarzenegger (1984), da "Robocop" con Peter Weller (1993) a "Io Robot" con Will Smith (2004). Seppur privo di una suggestiva creatura antropomorfa, la pellicola che più di altre ha anticipato i nostri timori sull'applicazione dell'IA è *Wargames* (1983), film in cui un giovanissimo Matthew Broderick interpreta un hacker adolescente che involontariamente accede al computer del Pentagono programmato per rispondere ad eventuali attacchi nucleari da parte dell'Unione Sovietica, sfidandolo per gioco ad una guerra termonucleare globale. Quello che predice *Wargames* è l'incapacità di un sistema informatico di dubitare della realtà di un attacco nucleare, resistenza invece riscontrabile negli esseri umani, quantomeno nella più parte.

Individuando nella fantascienza il veicolo principale di un movimento culturale che attribuisce la minaccia contro la nostra specie non più a un'altra specie, ma alle macchine che lui stesso ha creato, Benanti osserva come all'interno della narrazione distopica meriti particolare attenzione il celebre film "Blade Runner". Attraverso la altrettanto celebre frase *Io ho visto cose che voi umani non potreste immaginarvi (...)*, pronunciata dall'androide Roy Batty, la pellicola del 1982 segna l'affermarsi di "una narrazione epica, mitica e immaginifica che ritrae l'esperienza artificiale come più singolare, più vasta e più significativa di quella umana" (Benanti, 2017:8). Così che, nella pellicola, il replicante di ultima generazione *Nexus 6* viene presentato con lo slogan "più umano dell'umano". Passando dalla fantascienza al mondo reale possiamo notare come tale "esperienza artificiale" sia giunta a mutare la relazione tra l'uomo e la realtà, ormai mediata da artefatti: sono i sistemi meccanici dotati di intelligenza artificiale a dire all'uomo cosa vuol dire il reale e ad influire sul suo modo di sentire, capire e pensare (Benanti, 2018).

Oggi, quando quello che veniva narrato come un futuro distopico sta divenendo realtà nell'ambito della minaccia bellica, riscopriamo l'attualità delle quattro leggi di Asimov e ci chiediamo se le scene di *Robocop* e *Terminator* siano da considerarsi distopiche o profetiche.



## 1. I *Killer Robots*: Profezia o distopia?

A fronte dell'evidenza delle smisurate potenzialità dell'IA in vari settori della vita umana, negli anni recenti si è sviluppato un animato confronto su quale possa essere l'impatto sulla nostra società della sua dilagante applicazione. All'elevata mole di dati a disposizione e allo sviluppo di una potenza di calcolo in grado di gestirli, si aggiunge l'introduzione di algoritmi che permettono di estrarre conoscenza dai *big data* fornendo così alle macchine la possibilità di decidere autonomamente.

Ciò fa sì che oggi l'intelligenza artificiale possa essere usata non solo in ambienti controllati, ma in molti campi della nostra vita: dalla medicina alla finanza, dai trasporti all'educazione, dall'agricoltura all'e-commerce. In questi come in altri settori, l'IA ha ampiamente dimostrato la sua utilità, automatizzando alcuni lavori e aiutando gli esseri umani a fare meglio, più velocemente e meno faticosamente molte attività. Eppure, come ci ricorda Chimirri (2021: 17), la macchina, per quanto perfetta, "rimane quella che è", non ha facoltà di auto-sussistenza e auto-riproduzione e non è in grado di compiere azioni che richiedono discernimento etico, valori, intenzionalità, emotività. Riflettendo sull'ambiguità dello sviluppo tecnologico, Benedetto XVI scriveva nell'Enciclica *Caritas in Veritate*: "nata dalla nostra creatività come strumento di liberazione della persona, [la tecnica] può essere intesa come elemento di libertà assoluta che vuole prescindere dai limiti che le cose portano in sé. (...) Ma quando l'unico criterio della verità è l'efficienza e l'utilità, lo sviluppo dell'uomo è automaticamente negato. La responsabilità umana è sé stessa solo quando risponde al fascino della tecnica con decisioni che siano frutto di responsabilità morale" (*Caritas in Veritate*, nn. 69-70, 2009).

Da un punto di vista dell'etica, non solo religiosa ma anche laica, si tratta quindi di definire dove tracciare la linea di demarcazione tra gli usi ritenuti accettabili e quelli inaccettabili della nuova disciplina, prevenendone un uso inappropriato. Una specifica ma cruciale articolazione di tale dibattito emerge nell'ambito strategico. Lo sviluppo di sistemi d'arma dotati di un ampio livello di autonomia prefigura, in un futuro non così lontano, la realizzazione di armi completamente autonome. Osservando l'intimo legame tra scienza, alta tecnologia e guerra, Chris Hables Gray denunciava la "sempre maggiore interconnessione tra le forze della creazione e della distruzione" (Gray, 1997: 8).

Un campo chiave di applicazione dell'IA in ambito militare è quello del *wargame*, la simulazione di un'operazione militare la cui sequenza di eventi influenza ed è influenzata dalle decisioni prese dagli attori che rappresentano le parti opposte. Sviluppato per la prima volta nel 1820 da due ufficiali prussiani, il *kriegsspiel* è stato utilizzato con successo in molti conflitti, tanto da essere usato oggi da molte Nazioni come strumento di addestramento e pianificazione militare (Perla, 1990). Un ruolo fondamentale nello sviluppo dell'integrazione tra IA e *wargame* è rivestito dal settore della videoludica che, sebbene la relativa ricerca sia stata sviluppata in ambito commerciale, ha consentito di raccogliere una vasta quantità di materiale in tale ambito (IRAD, 2023). Ad oggi le maggiori potenze militari stanno investendo ingenti



risorse per creare scenari bellici sempre più realistici e complessi e per automatizzare alcune attività altrimenti troppo complesse e lunghe da svolgere. D'altro canto, i più recenti progressi nella IA convergono nel conferire alle macchine militari una crescente indipendenza in grado di affrancarle dalla presenza dell'uomo: dal controllo diretto (il cd. *man in the loop*) si sta passando al controllo umano indiretto (*man on the loop*), fino all'incombente *man off the loop*, ovvero alla vera e propria estraneazione dell'uomo e alla completa autonomia della macchina.

Nel maggio 2021, un rapporto Onu informava sul possibile primo uso in combattimento di droni (arma di per sé semi-autonoma) in nuove modalità "autonome". Il riferimento è all'utilizzo dei droni turchi STM Kargu-2 nei campi di battaglia libici, i quali avrebbero ingaggiato autonomamente il bersaglio senza necessità di connessione dati con un operatore. E non si tratta di un caso isolato. Per quanto riguarda i sistemi di difesa, sono anni che le Forze Armate di vari Stati hanno in dotazione scudi protettivi dotati di sistemi autonomi anti-materiali. Si pensi al ben noto *Iron Dome* israeliano, ma anche agli analoghi *Mantis* e *Phalanx*, in dotazione rispettivamente alla Germania e agli Stati Uniti. Spostandoci dalla difesa all'offesa, gli Stati Uniti già oggi impiegano sistemi d'arma semi-autonomi come i missili antinave a lunga gittata che hanno sì bisogno dell'intervento umano per identificare il bersaglio, ma agiscono in piena autonomia nello scegliere come distruggerlo, oppure come la nave antisottomarino *Sea Hunter* che può navigare per mesi senza equipaggio. Israele, invece, utilizza un drone (*Harpy*) che, senza l'intervento umano, è in grado di individuare e annientare sistemi radar.

Ogni guerra offre d'altronde l'occasione di dispiegare e sperimentare nuovi sistemi d'arma e l'invasione russa dell'Ucraina nel febbraio 2022 non fa eccezione: Kiev ha già nella sua artiglieria droni d'attacco semi-autonomi e armi dotate di IA per fronteggiare i droni nemici e Mosca non sembra essere da meno. La cosa sicura è che più si va avanti a combattere, più si affinano le tecniche di guerra e sempre più spesso anche i droni saranno destinati a muoversi senza l'aiuto dell'uomo.

Ma il caso più evoluto di IA militare registrato sinora è rappresentato dal sistema di processamento dati impiegato dalle forze armate israeliane nella guerra di Gaza dopo il 6 ottobre 2023 denominato *the Gospel*, ovvero il Vangelo. Stipato di milioni di dati provenienti da voli di droni, intercettazioni alle comunicazioni del nemico, monitoraggio dei suoi movimenti e comportamenti, il "Vangelo" viene descritto dall'ex capo di stato maggiore israeliano Aviv Kochavi come "una macchina che produce una vasta quantità di dati più efficacemente di qualsiasi umano e li traduce in altrettanti bersagli". Fonti israeliane definiscono gli attacchi guidati dall'algoritmo "accurati" ma non "non chirurgici", un aspetto che soltanto nei primi tre mesi di guerra ha conseguito il record di 24.000 vittime, di cui la grande maggioranza civili.

Si rivela sempre più valida la valutazione della Ong Human Rights Watch, secondo cui la sfida dei robot killer, così come il cambiamento climatico, costituisce "una grande minaccia per l'umanità che merita un'urgente azione multilaterale" (Human Rights Watch 2020: 1).



## 2. IA o non IA, questo è il problema. Vantaggi e svantaggi dell'automazione dei sistemi d'arma

I crescenti investimenti per la Ricerca & Sviluppo nell'IA da parte dei governi a livello mondiale<sup>1</sup> inseguono il sogno ricorrente di politici e strateghi di intraprendere guerre e battaglie da vincere "a perdite zero". Sistemi d'arma semi-autonomi e in prospettiva completamente autonomi, sono d'altronde formidabili protettori delle (proprie) vite e contemporaneamente un potenziale sterminatore delle vite degli altri (i nemici). Quindi più che dell'utopia delle *perdite zero* occorrerebbe parlare delle innumerevoli perdite nei ranghi nemici a costi zero (Battistelli, 2023). Se viene fatta salva la vulnerabilità di chi colpisce, uccide e distrugge - oggi attraverso il controllo remoto di un drone, domani attraverso l'utilizzo di armi completamente autonome - non accade lo stesso per chi viene colpito, che anzi è ridotto a bersaglio.

I micidiali progressi conseguiti nell'arte di massacrare in guerra non rappresentano un'inversione di tendenza, ma al contrario una semplice moltiplicazione delle logiche strategiche di infliggere il massimo danno al minimo costo.

La distanza fisica che intercorre tra i due schieramenti contrapposti è semplicemente aumentata nel corso del tempo grazie alla sempre maggiore potenza, precisione e capacità distruttiva degli armamenti. A partire dall'utilizzo della clava, la capacità di colpire il nemico rimanendo a debita distanza è rimasto il più scontato degli obiettivi tattici praticati dagli eserciti. Dalla fionda all'arco, dal fucile all'artiglieria, sino al bombardamento aereo, le dimensioni di un campo di battaglia e del conseguente fronteggiamento dei combattenti tra loro sono passate da pochi metri a qualche migliaio di chilometri. In età moderna e in particolare dal XX secolo i problemi connessi alla gittata limitata dei proiettili e all'approssimazione nel puntamento appaiono oggi largamente superati grazie con lo sviluppo delle armi guidate, grazie alle quali la precisione verso l'obiettivo è divenuto sostanzialmente indipendente dalla distanza della base di lancio. Il tutto in vista dell'imperativo di investire sulle armi per massimizzare le perdite del nemico e minimizzare le proprie.

Ancora tollerati in un contesto di emergenza globale come quello rappresentato dalla Seconda guerra mondiale, nella società contemporanea, progressivamente più secolarizzata e focalizzata sui diritti individuali, gli enormi costi di vite umane imposti dalle guerre oggi appaiono inaccettabili (Battistelli *et al.* 2012). Da un lato i decisori tendono a presentare le vittime dei conflitti come involontari effetti collaterali, ma dall'altro l'opinione pubblica vede i caduti come un prezzo pagato a specifiche scelte politiche. In una misura che è forzatamente ridotta nei regimi autoritari e invece è massima in quelli democratici, gli stessi decisori sono obbligati a tenerne conto, come mostrano le acrobazie logiche e verbali dei governi occidentali a proposito dei costi rovinosi della guerra, ad esempio in Ucraina o a Gaza. Rispetto alle società pre-

<sup>1</sup> Diverse nazioni con forze armate high-tech, in particolare Stati Uniti, Cina, Israele, Corea del Sud, Russia e Regno Unito si stanno muovendo verso sistemi che darebbero maggiore autonomia di combattimento alle macchine. Il mercato dell'intelligenza artificiale nel settore bellico dovrebbe crescere del 33,3% tra il 2023 e il 2028, passando da 9,2 a 38,8 miliardi di dollari.



moderne, in cui la morte era familiare e resa accettabile dal credo religioso, nelle moderne società occidentali, sempre più secolarizzate si è fortemente indebolita la funzione consolatoria del cristianesimo e l'individuo è lasciato solo di fronte alla morte (Battistelli, 2023). Nella dominante "civiltà delle buone maniere" (Elias, 1982) l'uccidere e l'essere uccisi (dato a lungo ritenuto costitutivo ed effetto "naturale" delle guerre) suscitano il rigetto dell'opinione pubblica, ogni volta che essa è nelle condizioni di essere informata e di avere la libertà di esprimersi.

Secondo i fautori delle armi autonome, il loro impiego garantirà una maggiore precisione nell'attacco di obiettivi militari che, al contrario di un comune soldato, l'essere esenti da stress e perturbazioni emotive potrà determinare un risparmio di vite umane nelle forze in campo (Tamburrini, 2020). A ciò va aggiunta la loro economicità: basti pensare che un caccia bombardiere ha un costo unitario di circa 130 milioni di dollari a fronte dei circa 5 milioni di dollari di un'arma dotata di semi-autonomia come un drone d'attacco.

Secondo il fisico Jürgen Altmann (2013) i possibili vantaggi ipotizzati nel breve periodo vanno però bilanciati con le conseguenze a lungo termine sulla sicurezza interna ed internazionale. Secondo Altmann, infatti, i costi relativamente bassi delle armi autonome potrebbero spingere un ampio numero di Stati, ma anche gruppi terroristici, a dotarsene, facendo così vacillare delicati equilibri strategici. L'impiego dei robot killer e la conseguente riduzione di perdite tra le proprie truppe renderebbe inoltre più difficile "resistere alla tentazione della guerra", riducendo la società civile a spettatore sempre più passivo delle decisioni dei governi. Al *rischio strategico* di un numero maggiore di conflitti armati (indotto dalla "economicità" di perdite di combattenti non umani), si aggiunge il *rischio tattico* di un'accelerazione tale nel ritmo del loro svolgimento da diventare incompatibile con i tempi di reazione degli operatori umani e sfuggire alla loro capacità di controllo (Altmann e Sauer, 2017).

A questo vanno aggiunte le considerazioni etiche e morali. In molti ritengono che l'impiego delle armi autonome nei teatri di guerra violerebbe principi legali ed etici fondamentali, destabilizzando la pace e la sicurezza internazionali (Tamburrini, 2016 e 2023; Pax, 2011 e 2014) e compromettendo il rispetto dei principi della *guerra giusta* teorizzata, fra gli altri, da Michael Walzer (2009). L'utilizzo delle armi autonome minaccia innanzitutto i due principi-base del diritto umanitario: il principio di distinzione e il principio di proporzionalità. Se da un lato, infatti, risulta difficile immaginare come la robotica e l'IA potranno sviluppare sistemi d'arma capaci di distinguere tra combattenti e popolazione civile, fra bersagli leciti e illeciti, fra prigionieri di guerra, internati civili e detenuti comuni ecc.; dall'altro lato, non possedendo l'IA capacità cognitive ed emotive né competenze sociali ed esperienziali, è del tutto aleatorio ritenere che essa sarà capace di valutare se, quale e quanto una determinata azione configuri un legittimo vantaggio militare, in relazione alle perdite umane e ai danni alla popolazione civile, ai beni patrimoniali e a quelli culturali, così come ai beni civili, incidentalmente causati.



Oltre ai principi di distinzione e proporzionalità, l'utilizzo di sistemi d'arma autonomi minerebbe anche il principio di responsabilità: se un'arma autonoma dovesse, non discriminando adeguatamente il bersaglio legittimo, provocare sofferenze ingiustificate al nemico e/o alla popolazione civile, chi sarebbe ritenuto responsabile? Gli scienziati che hanno contribuito a svilupparla e/o coloro che l'hanno prodotta? Il personale militare direttamente incaricato di supervisionarne l'attività oppure il comandante in capo dell'intera operazione? Da qui il dubbio relativo alla possibilità, da parte delle armi autonome, di mantenere la catena delle responsabilità nelle azioni belliche (Tamburrini, 2020; 2023). Peraltro, proprio la certezza dell'impunità potrebbe incentivarne l'utilizzo da parte dei comandanti, malgrado la possibilità di produrre vittime civili. Infine, l'utilizzo dei robot killer non garantirebbe il rispetto della dignità umana: in un conflitto armato le decisioni sulla vita e sulla morte richiedono intuizione e compassione e gli esseri umani, per quanto fallibili, possono (talora) possederle entrambe. Le macchine, in ogni caso, no (Heyns, 2013).



### 3. "C'è un modo per liberare l'uomo dalla fatalità della guerra?"

Le ragioni per un accordo preventivo finalizzato all'esclusione delle funzioni belliche nei futuri sviluppi della IA sono talmente tante e talmente evidenti che non dovrebbe essere necessario aggiungere nulla per sostenerne l'opportunità. Per quel poco che le scienze sociali possono fare per arginare la fatale estensione della IA alle armi, è da sottolineare l'allarme da lanciare in riferimento allo specifico danno psicologico che esse provocano nella già debole inibizione umana a distruggere. Si tratta della scotomizzazione di quel senso di colpa che uccidere determina fisiologicamente nella persona, a sua volta uno dei pochi fattori inibitori dell'aggressività bellica. L'estraneità rispetto al campo di battaglia, come precedentemente accennato, potrebbe far cedere più facilmente alla tentazione di entrare in guerra contro il nemico di turno.

Risulta quindi oggi più attuale che mai la domanda che nel luglio del 1932 Einstein pose per via epistolare a Freud, definendola *la più urgente fra tutte quelle che si pongono alla civiltà: "c'è un modo per liberare gli uomini dalla fatalità della guerra?"* (Freud, Einstein, 1997). La risposta di Freud non fu di certo incoraggiante. In poche pagine, lo psicanalista austriaco illustra ad Einstein la sua convinzione dell'esistenza, in ognuno di noi, di due pulsioni parimenti indispensabili: quelle che tendono a conservare e unire e quelle che tendono alla distruzione. La pulsione distruttiva, laddove si rivolge verso l'esterno, protegge la propria vita distruggendone una estranea. La conclusione che propone al fisico tedesco è che non c'è speranza di poter sopprimere le tendenze aggressive dell'uomo. Nella sua epistola Freud sottolinea anche come, storicamente, la forza muscolare dell'uomo sia stata accresciuta dall'utilizzo di strumenti sempre più evoluti, le armi, attraverso cui la superiorità intellettuale prende il sopravvento sulla forza bruta (Freud, Einstein, 1997).

Dobbiamo proprio all'approccio psicoanalitico un'interessante riflessione sul ruolo che nell'inconscio rivestono le armi, efficacemente estensibile alle armi semiautonome (droni) e autonome (robot killer). Nell'interpretazione neofreudiana di Franco Fornari della guerra dentro e fuori il campo di battaglia, viene descritta "la simbologia sessuale delle armi di guerra, che ogni psicoanalista riscontra nella sua pratica quotidiana", avvalorando "la tesi per cui le fantasie di guerra, intimamente mescolate ai simboli della sessualità infantile, sono da considerarsi una specie di fenomeno pandemico, quasi quotidiano e stabile del preconscious di ogni uomo, anche in tempo di pace" (Fornari 1966: 92-3).

Ripercorrendo le categorie freudiane, Fornari elabora una delle più originali letture mai proposte sul tema della corsa agli armamenti nell'era atomica, assimilandola a una regressione rispetto alle differenti fasi della sessualità. Così la massima forza distruttiva (che egli definisce "pantoclastica") della Bomba corrisponde allo stadio orale nella sessualità del bambino, per via della pervasività che ne renderebbe inarrestabile l'assorbimento da tutte le specie viventi, a cominciare da quella umana. Nella medesima prospettiva, potremmo leggere le armi semi-autonome e autonome come riconducibili al più avanzato stadio della sessualità, quello genitale. Da sempre le armi da sguainare (la spada) e/o scagliare (la lancia) sono simboli molto usati nelle dottrine strategiche per indicare la funzione della difesa e dell'attacco, ad esempio nella metafora dello "scudo" e della "lancia". Il significato fallico di arnesi destinati a penetrare nel territorio e nel corpo stesso del nemico è evidente.

Tale significato non è attenuato ma potenziato negli sviluppi delle moderne tecnologie belliche, fondate sul lancio di ordigni a distanza: dai primi passi dell'arma aerea nella Prima guerra mondiale, alla teoria dei bombardamenti strategici teorizzata negli anni Venti dal generale italiano Giulio Douhet, alla prima applicazione attuata nella guerra civile spagnola dall'aviazione italo-tedesca a Guernica, al massiccio impiego praticato nella Seconda guerra mondiale dagli Alleati a Dresda, Colonia, Brema, Lubeca, Montecassino - non diversamente peraltro da quanto i tedeschi avevano fatto a Rotterdam, Coventry, Londra (Battistelli, 2018).

Come abbiamo visto, l'apice viene raggiunto oggi con la distruzione del nemico da remoto mediante i velivoli senza pilota o droni, in cui l'uomo agisce come pilota o regista. A sua volta questo traguardo è già in via di superamento con la diretta discesa in campo dell'Intelligenza Artificiale che lascia all'operatore un lasso temporale talmente irrisorio da lasciare alla decisione umana un ruolo meramente nominale. Entrambi i casi sembrano fornire situazioni ideali per la più efficace e micidiale delle rappresentazioni del nemico. Non più quella di un essere umano che il combattente si trova davanti, venendo posto di fronte al dilemma se sparare o meno per uccidere (Tamburrini, 2023), bensì un nemico fantasmizzato, oggetto cui contrapporre un altro oggetto; il tutto con il duplice vantaggio di annichilire l'oggetto dell'odio senza doverne affrontare sino in fondo il possibile senso di colpa.

A più di novant'anni dal carteggio tra Einstein e Freud, una risposta al quesito se ci sia o meno la possibilità di liberare gli uomini dalla fatalità della guerra sembra più urgente che mai, ma allo stesso tempo l'alternativa ad una risposta brutalmente



negativa appare difficile da formulare. Ci può venire in aiuto, ancora una volta, la suggestiva metafora proposta da Bobbio (1979) per esprimere la sua convinzione sulla necessità del disarmo: "Chi ha un gatto che graffia eviti di sprofondarsi in speculazioni per modificarne gli istinti: gli tagli le unghie". La *spes ultima dea* in cui confidare è che nei decisori prevalga dunque l'istinto di autoconservazione della specie piuttosto che la tendenza a seguire il canto delle sirene di dispositivi artificiali che promettono risparmi nei costi di realizzazione e nel dispendio di vite umane. La speranza è che la realtà, anche in questo caso, non superi l'immaginazione.



## Bibliography

- Altmann J. (2013), "Arms control for armed uninhabited vehicles: an ethical issue", in *Ethics and Information Technology*, n. 15, 137-152.
- Altmann J., Sauer F. (2017), "Autonomous Weapon Systems and Strategic Stability", in *Survival. Global Politics and Strategy*, 59 (5), 117-142.
- Asimov I. (1962), *Io, Robot*, Milano, Bompiani.
- Asimov I. (1986), *Io Robot e l'Impero*, Milano, Arnoldo Mondadori Editore.
- Battistelli F. (2018), "Guernica: dalla guerra totale al disarmo", in S. Baccaglioni (a cura di), *Guernica icona di pace*, Silvana Editoriale, Milano, Cinisello Balsamo.
- Battistelli, F.(2023), "Armi non umane. Miti, sogni e incubi dell'autonomia delle Armi", in Farruggia F. (a cura di), *Dai droni alle armi autonome. Lasciare l'apocalisse alle macchine?*, Milano, Franco Angeli, 21-42.
- Battistelli F., Galantino M.G., Lucianetti L.F., Striuli L. (2021), *Opinioni sulla guerra. L'opinione pubblica italiana e internazionale di fronte all'uso della forza*, Milano, Franco Angeli.
- Benanti P. (2017), *Postumano, troppo postumano. Neurotecnologie e «human enhancement»*, Roma, Castelvecchi Editore.
- Benanti P. (2018), *Le machine sapienti. Intelligenze artificiali e decisioni umane*, Bologna, Marietti, 1820.
- Benanti P. (2022), *Human in the loop. Decisioni umane e intelligenze artificiali*, Milano, Mondadori.
- Benedetto XVI. (2009), *Caritas in Veritate*, n.n. 69-70.
- Bhuta N., Beck S., Geis R., Liu H. & Kres C. (eds.) (2016), *Autonomous Weapons Systems: Law, Ethics, Policy*, Cambridge University Press.
- Bobbio N. (1979), *Il problema della guerra e le vie della pace*, Bologna, Il Mulino.
- Buttazzo G. (2023), "Coscienza Artificiale: implicazioni per l'umanità", in *Mondo Digitale*, aprile 2023.
- Carlucci Aiello L., Dapor M. (2004), "Intelligenza Artificiale: I primi 50 anni dell'Intelligenza Artificiale", *Mondo Digitale*, n. 2, 1-18.





Chimirri G. (2021), *Intelligenza artificiale. Etica delle macchine pensanti*, Firenze, Asterios.

Einstein A., Freud S. (1997), *Perché la guerra?*, Torino, Bollati Boringhieri.

Elias N. (1982), *La civiltà delle buone maniere*, Bologna, il Mulino.

Farruggia, F. (a cura di) (2023), *Dai droni alle armi autonome. Lasciare l'apocalisse alle macchine?*, Milano, Franco Angeli.

Fornari F. (1966), *Psicanalisi della guerra*, Milano, Feltrinelli.

Hables Gray C. (1997), *Postmodern War: The New Politics of Conflict*, London, Routledge.

Heyns C. (2013), *Report of the Special Rapporteur on extrajudicial summary or arbitrary executions* from [http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47\\_en.pdf](http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC-23-47_en.pdf), pp. 10-20.

Human Rights Watch (2020), *World Report 2020* from [https://www.hrw.org/sites/default/files/world\\_report\\_download/hrw\\_world\\_report\\_2020\\_0.pdf](https://www.hrw.org/sites/default/files/world_report_download/hrw_world_report_2020_0.pdf).

IRAD (2023), *Wargaming e Intelligenza Artificiale - situazione, livello tecnologico internazionale e nazionale, prospettive di sviluppo per la Difesa*, Roma.

Norving P., Russel S. (2005), *Intelligenza artificiale. Un approccio moderno*, vol. 1, Pearson, Torino.

PAX (2020), *Conflicted intelligence. How universities can help prevent the development of lethal autonomous weapons* from <https://www.paxforpeace.nl/publications/all-publications/conflicted-intelligence>.

PAX (2014), *Deadly Decisions. 8 Objections to Killer Robots* from <https://autonomousweapons.org/deadly-decisions-8-objections-to-killer-robots/>.

Perla P. (1990), *The art of wargaming: a guide for professionals and hobbyists*, U.S. Naval Institute, Annapolis.

Piga Bruni E., Presutti C. (2019), "L'umano nell'epoca della sua riproducibilità tecnica: l'inconscio artificiale nella fantascienza", in *Between*, vol. IX, n. 17, 1-55.

Tamburrini G. (2016), "On banning autonomous weapons systems: from deontological to wide consequentialist reasons", in Bhuta N. et al. (eds), *Autonomous Weapons Systems: Law, Ethics, Policy*, Cambridge University Press: 122-142.

Tamburrini G. (2020), *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*, Roma, Carocci.

Tamburrini G. (2023), "Il dibattito etico sulle armi autonome", in Farruggia F. (a cura di), *Dai droni alle armi autonome. Lasciare l'apocalisse alle macchine?*, Milano, Franco Angeli, pp.98-112.

Turing A. M. (1950), "Computing Machinery and intelligence", in *Mind*, vol. LIX, Issue 236, pp. 433-460.

Walzer M. (2009), *Guerre giuste e ingiuste. Un discorso morale con esemplificazioni storiche*, Napoli, Liguori Editore.







# Immaginario *Black Box*: prospettive mitiche e critiche tra innovazione bellica ed esplorazione artistica

Lorenzo Aimo

lorenzo.aimo@unimi.it

DIPAFILO | *Università degli Studi di Milano*

DAR | *Alma Mater Studiorum Università di Bologna*



Luca Brunet

luca.brunet@unimi.it

DIPAFILO | *Università degli Studi di Milano*

DFE | *Università di Torino*

## Abstract

*Black Box Imaginary: Mythical and Critical Perspectives Between Military Innovation and Artistic Exploration*

Within the imaginaries of AI, automation occupies a central position and receives varying nuanced connotations. In both technical-scientific and popular discourses, this notion takes on the rhetorical and visually metaphorical form of the black box, implying the inability to fully visualize and understand what happens inside AI systems due to a complex and intricate layering of neural networks and algorithms. The rhetoric of the black box shapes both the justification and the critique of the use of AI, by showing or obscuring implications regarding accountability, values, and ethics. The present contribution explores how the myth of the black box and various corresponding imaginaries manifest across diverse case studies in two distinct domains, revealing their political, social, and semiotic implications. Through the analysis of the representation of AI by military websites (such as Palantir and Gospel) and of artworks on AI (like Paglen's 2019 and Akten's 2019 works), this paper endeavors to showcase how contrasting perspectives – mythical and critical – characterize the metaphor of the black box within the imaginary of AI.

## Keywords

Artificial Intelligence | Imaginary | Black Box | Myth | Opacity



## 1. Introduzione<sup>1</sup>

**V**olgendo uno sguardo al passato, guardando a uno degli esempi più famosi di sistemi (apparentemente) intelligenti – tuttora figura presente nel contemporaneo – è possibile comprendere il percorso parallelo e congiunto che caratterizza l'Intelligenza Artificiale (IA) e la figura della *black box*. Progettato nel 1770 da Wolfgang von Kempelen, il Turco Meccanico rappresenta un'icona (Alexander, 2008) la cui immagine e il derivante significato riadattato vengono rievocati dal programma omonimo di Amazon, quello dei Mechanical Turk, ossia l'insieme di lavoratori anonimi che contribuiscono al processo di costruzione di una moderna IA. Nella "congegnata illusione" (Crawford, 2021: 79) di Von Kempelen, veniva offerta alla vista del pubblico una figura meccanica antropomorfa a grandezza naturale in grado di giocare a scacchi; tuttavia, all'interno del mobile a cui sedeva l'automa, letteralmente inscatolato in esso, si nascondeva un essere umano che, pilotando l'automa, risultava essere il vero artefice della performance nonchè dell'inganno<sup>2</sup>. Rimanendo celata all'occhio esterno, la persona inscatolata, coincidendo con il funzionamento della macchina, garantiva la simulazione dell'intelligenza umana da parte di un automa.

Oltre a ricordare quanto l'umano sia costitutivo della costruzione e del funzionamento dei sistemi di IA attuali, l'esempio storico risulta evocativo di una dinamica che, nella dicotomia tra esterno-visibile e interno-invisibile, ha assunto diverse forme e livelli nell'ambito dei sistemi di IA: ciò che viene metaforicamente indicato come *black box*.

All'interno dello spettro di immaginari socio-tecnici (Jasanoff, 2015)<sup>3</sup> che circondano l'IA, la figura metaforica della *black box* appare cruciale, in quanto si inserisce in un sostrato narrativo strettamente legato all'idea cardine e onnipresente dell'automazione – da cui derivano le figure narrative caratteristiche di questi



<sup>1</sup> Il presente contributo è stato pensato nella sua interezza da entrambi gli autori. Nello specifico, introduzione e conclusione sono state elaborate e scritte in collaborazione; Lorenzo Aimo è responsabile dei paragrafi 3 e 5; Luca Brunet è responsabile dei paragrafi 2 e 4, sebbene durante la stesura si sia sempre tenuto un confronto tra gli autori.

<sup>2</sup> Risulta interessante per uno studio sull'immaginario delle IA notare come uno dei primi casi di automa intelligente consistesse in realtà in un vero e proprio inganno, basato sull'occultamento di un meccanismo all'interno di un artefatto, e sulla resa visibile di una superficie significativa che viene usata per mentire; si tratta cioè di un meccanismo fondamentalmente semiotico (cfr. Paolucci, 2021; Eco, 1975). Sul rapporto tra IA e inganno si veda Natale (2022).

<sup>3</sup> Jasanoff definisce tali immaginari come "collectively held and performed visions of desirable futures (or of resistance against the undesirable)" (2015: 19) che vengono sorretti e sorreggono il progresso scientifico-tecnologico. La storia degli sviluppi dell'Intelligenza Artificiale, caratterizzata da fasi alterne, paure ed entusiasmi, dimostra come tale dicotomia sugli esiti del futuro ne abbia condizionato tanto la progettazione e costruzione effettiva - sia in senso economico e materiale che epistemologico e concettuale - quanto l'utilizzo e la percezione comune.



immaginari<sup>4</sup>. In ragione della forza concettuale e cognitiva caratteristica delle metafore (Lakoff e Johnson, 1981), la metafora della *black box* adempie al compito di significare un limite, quello tra visibile e conoscibile, che scaturisce dalla tecnologia stessa o dal fenomeno contestualmente osservato, il cui meccanismo di funzionamento rimane sia celato allo sguardo umano, sia incomprensibile.

La metafora della *black box* non appartiene infatti unicamente all'ambito discorsivo degli artefatti tecnologici, ma a quello dei sistemi più in generale. Lo dimostra l'utilizzo che, dagli albori della disciplina, ne fece Wiener nell'ambito della cibernetica, definendo la *black box* come "a piece of apparatus [...] which performs a definite operation [...] but for which we do not necessarily have any information of the structure of which this operation is performed" (Wiener, 1961: X-XI). L'ambito della cibernetica fu, oltre al contesto teorico e scientifico di origine dell'IA, il nucleo discorsivo in cui la figura della *black box* si afferma e da cui si propaga interdisciplinariamente. Il termine "black box" nasce dallo stretto legame tra ambito militare e cibernetica: come dimostra Galison (1994), esso risale al periodo bellico del secondo conflitto mondiale, ossia a quando nel 1940, all'interno della missione Tizard, un magnetron venne trasportato in una scatola nominata appunto "black box" dal Regno Unito agli Stati Uniti (Von Hilgers, 2011: 47). Al tempo, le ricerche di Wiener che poi confluirono nella sua teoria cibernetica riguardavano proprio lo sviluppo di tecnologie militari (Galison, 1994)<sup>5</sup>.

Da una specifica tecnologia militare, la *black box* divenne all'interno della cibernetica contemporaneamente "a mathematical theory, a device, and a metaphor" (Petrick, 2020: 577) caratterizzante la sua epistemologia. Il modello della *black box* fu fondante del primo modello matematico neurale di McCulloch e Pitts (1943), con cui si cercava di spiegare il funzionamento dei processi cerebrali umani e che costituisce il punto di partenza per i seguenti sviluppi dell'IA.

Da concetto prettamente macchinico, la *black box* è stata presto utilizzata anche in campo psicologico e cognitivo<sup>6</sup>: per via dell'osmosi tra gli studi sulla cibernetica e sulla psicologia del tempo<sup>7</sup>, sulla scia della ricerca sull'IA in quanto simulazione del cervello umano, la stessa metafora è stata usata anche nel contesto delle teorie cognitive comportamentiste per riferirsi al funzionamento del sistema mente cervello<sup>8</sup>. Ma non solo, fu infatti proiettata, oltre che sull'umano, sul mondo stesso: secondo Ross Ashby "we do in fact work, in our daily lives, much more with Black

<sup>4</sup> Per una ricognizione degli immaginari della narrativa letteraria quanto cinematografica in cui l'IA prende forma antropomorfa in varie figure si veda Cave, Dihal e Dillon (2020); per una prospettiva che va oltre il contesto occidentale, Cave e Dihal (2023).

<sup>5</sup> Come spiega Galison, infatti, Wiener introdusse nelle sue teorie il termine "black box" e "feedback" proprio a partire dalla progettazione e lo sviluppo di un "anticraft AA predictor", una tecnologia in grado di anticipare e prevedere i movimenti di un aereo nemico per colpirlo.

<sup>6</sup> Per gli intrecci e le influenze reciproche che caratterizzarono la storia delle teorie psicologiche e cibernetiche si veda Tilak, Glassman, Kuznetcova e Pelfrey, 2022.

<sup>7</sup> Assieme a due suoi colleghi, lo stesso Wiener applicò il modello della *black box* al comportamento umano. A riguardo si veda N. Hayles in Hui (2024).

<sup>8</sup> Per una panoramica specifica sull'utilizzo della metafora della *black box* nelle teorie comportamentiste si veda Friedenberg & Gordon (2006: 85-88), e più recentemente in chiave critica Gigerenzer (2019).

Boxes than we are apt to think [...] the real objects are in fact all Black Boxes” e la teoria della *black box* riguarda “what information comes from the object” (1957: 110) – questa prospettiva caratterizzò la cibernetica di secondo ordine che ne espanso le implicazioni in relazione alle scoperte relative ai processi neurofisiologici, nel tentativo di applicare i modelli cibernetici a esperienze vive, a fenomeni che non hanno natura meccanica né una struttura chiusa (Tilak et al., 2022). La metafora della *black box*, dunque, proietta un’ombra di inscruetabilità sul meccanismo interno di funzionamento di un sistema, rendendo pertinente ciò che emerge come visibile, ovvero il risultato e l’informazione ivi contenuta.

Attualmente, tale metafora risulta ancora più pregnante dati i contemporanei sistemi di IA *machine deep learning* che stanno, progressivamente, inserendosi in vari ambiti del quotidiano: dal riconoscimento facciale, di persone o di oggetti, ai recenti sviluppi di IA generativa. In questo contesto, la metafora della *black box* assume la forma tramite cui ci si riferisce all’inscruetabilità dovuta alla complessa stratificazione di reti neurali, algoritmi e non solo. Infatti, anche la componente di partenza, i dati su cui il modello IA viene allenato, può risultare problematica e oscura, questione che ha portato diversi studiosi a indicarne le criticità, ad esempio la possibilità di avere a che fare con “dataset biased”, ricchi di pregiudizi, dovuti soprattutto alla massiva rappresentazione di alcune identità e alla sottorappresentazione o esclusione di altre identità dai dati selezionati<sup>9</sup>.

Come si tenterà di dimostrare, all’interno dell’immaginario dell’IA, la metafora della *black box* si articola in un modo tale per cui si produce una valorizzazione dell’automazione insita al funzionamento dell’IA. Tale valorizzazione può essere positiva, per cui l’idea di un funzionamento automatico non sondabile assume la sinonimia di assenza di intervento umano, intesa come baluardo di obiettività e neutralità; oppure negativa, in cui assume il significato di un’impossibilità di controllo totale da parte dell’umano. Da tale valorizzazione bivalente derivano dunque due assi narrativi secondo cui si strutturano determinate porzioni dell’immaginario dell’IA, spesso in una contrapposizione polarizzata che ricorda quella tra “apocalittici e integrati” (Eco, 1964). All’interno di queste narrazioni, la metafora della *black box* detiene un potere retorico che viene sfruttato o contestato secondo due prospettive: la prima che, seguendo il senso barthesiano del termine<sup>10</sup>, è possibile definire *mitica*, e la seconda che vi si contrappone, definibile *critica*. Partendo da un assunto che interpreta a livello semiotico il concetto di immaginario, attraverso l’analisi dei casi studio si vuole rendere conto di come l’immaginario dell’IA che scaturisce dalla metafora della *black box* istituisca discorsi e pratiche in cui tale tecnologia viene considerata in modi diversi.



<sup>9</sup> Si veda a questo proposito Cave e Dihal (2020) sulle implicazioni razziali e di genere all’interno dell’immaginario e Voto (2022) per quanto riguarda il rapporto tra IA e identità di genere, con l’analisi di un caso studio artistico che problematizza proprio il lavoro invisibile di categorizzazione dei dati, in maniera simile ai casi che presentiamo qui nell’ultimo paragrafo.

<sup>10</sup> Ogni riferimento al mito è fatto a partire dal concetto elaborato da Roland Barthes (1974), cfr. infra 4.

## 2. Immaginario: una prospettiva semiotica

La *black box* possiede uno statuto ambivalente: essa appare da un lato come una metafora utile a rendere conto del funzionamento (anche se inscrutabile) della tecnologia, ma dall'altro, proprio in quanto oggetto dal contenuto oscuro, indefinito, e in grado di applicarsi a diversi fenomeni, si presta a essere oggetto caratteristico dello spazio dell'immaginario; infatti, il concetto di immaginario non si oppone in maniera esclusiva a quello di realtà. Piuttosto la dimensione dell'immaginario svolge un continuo processo di re-interpretazione e traduzione della realtà, e circolarmente la realtà stessa riprende e re-interpreta i prodotti dell'immaginario. Questo assunto teorico si rivela particolarmente fecondo per comprendere nello specifico i diversi immaginari dell'IA<sup>11</sup>, dal momento che, come sostengono Stefano e Paolo Bory, "gli immaginari dell'AI non sono mere rappresentazioni finzionali di una realtà sociale futura, piuttosto si tratta di forme creative che agiscono nell'istituzione concreta del pensiero tecnologico, socio-culturale e storico-filosofico del presente" (Bory e Bory, 2015: 83).

Da una prospettiva semiotica è possibile considerare l'immaginario come una dimensione di senso che svolge una funzione di "mediazione tra essere umano e ambiente" (Ferraro, 2011: 92), che dipende quindi dalla facoltà dell'immaginazione<sup>12</sup>, e che possiede una natura collettiva, culturale e semiotica (Lorusso, 2011)<sup>13</sup>. Su questo punto la tradizione semiotica di ispirazione peirciana, attraverso la ripresa dello schematismo kantiano, si avvicina all'idea di Castoriadis sul rapporto tra "immaginazione radicale" e "immaginario sociale istituyente", pensato come rapporto fondamentale per l'esistenza del "reale", esperibile individualmente grazie all'immaginazione a partire da significati collettivi dell'immaginario<sup>14</sup>.

Proprio come l'immaginazione, l'immaginario costituisce una dimensione di senso dove si mescolano e si esplorano congiuntamente aspetti reali e virtuali dell'esperienza (Volli, 2011). In questo modo l'immaginario viene inteso come un'istanza di mediazione che contribuisce a plasmare i discorsi e parallelamente viene plasmato dai discorsi che circolano nella cultura, si tratta cioè di un sistema autopoietico, in termini semiotici un'enciclopedia parziale.

Seguendo Volli, l'immaginario può essere pensato come un "sistema semantico [...] di relazioni [...] che si può considerare seriamente solo nei suoi tagli locali" (Volli, 2011: 34), la località, in questo caso, è rappresentata dai contesti e dalla posizionalità specifica dei due casi studio presi in esame. Al contempo, un immaginario funziona



<sup>11</sup> Su questo punto si è espresso in maniera analoga Guido Ferraro: "in prospettiva semiotica uno degli aspetti più significativi potrebbe consistere proprio nel fatto che la linea d'evoluzione dell'IA presenta un'inconsueta ambivalenza per la quale, a ben vedere, tanto l'immaginario fantascientifico si ispira agli avanzamenti tecnologici quanto questi ultimi prendono spunto, sempre più decisamente, dalle proposte dell'immaginario." (Ferraro, 2023: 66)

<sup>12</sup> Non a caso l'immaginazione è il nucleo fondamentale per il funzionamento della percezione *tout court* (cfr. Paolucci, 2021).

<sup>13</sup> Più approfonditamente, Lorusso sostiene che l'immaginario possiede la funzione di costruire comunità: "sharing imaginaries is essential to create a community link." (Lorusso, 2011: 114).

<sup>14</sup> Si veda a questo proposito l'articolo di Pavan (2022).



anche come “meccanismo di esclusione” (Volli, 2011: 33), che seleziona alcune caratteristiche come pertinenti, altre no, alcune come euforiche, altre disforiche; in questo modo un immaginario socio-tecnicamente inteso si costruisce anche e soprattutto all’insegna di visioni di futuri desiderabili o di resistenza contro l’ indesiderabile (Jasanoff, 2015). Questa prospettiva permette, quindi, di considerare le tendenze contrastanti interne all’immaginario, che possono manifestare punti di vista più ideologici-mitici o più critici, impegnandosi nell’affermare un determinato asse narrativo-retorico dell’immaginario o al contrario, nel tentare di de-costruirlo<sup>15</sup>.

Di conseguenza, enunciare intorno a un certo tema-oggetto, come l’IA nei casi presi in esame, significa necessariamente prendere posizione in relazione a esso e ai discorsi che già circolano nell’enciclopedia (Paolucci, 2020), eventualmente nel tentativo di portare a una loro decostruzione o ridefinizione. Tale presa di posizione ha sempre anche una valenza politica, un significato relativo alle diverse economie (politiche) mobilitate nell’uso dei dispositivi tecnologici (Eugeni, 2021: 259-262), ovvero da una serie di logiche di distribuzione e trasformazione di risorse e di lavoro.

L’immaginario, dunque, più che essere specchio finzionale della realtà, svolge un ruolo fondamentale nell’istituzione, nell’interpretazione e nella creazione della realtà in quanto tale, e quindi degli artefatti, delle relazioni sociali, delle stesse istituzioni politiche ed economiche. Come scrive Castoriadis: “Il soggetto non può incominciare ad abbozzare gli elementi del reale [...] se non muovendo da (e obbedendo a) schemi immaginari che gli sono propri” (1975: 185).

Proprio per questo la mediazione degli immaginari si annida anche nei contesti più “reali”, al punto che è impossibile porre una soglia a priori che separa lo spazio reale da quello immaginario. Piuttosto, come le analisi qui proposte tenteranno di dimostrare, un aspetto importante riguarda la capacità dell’immaginario di riflettere su sé stesso, mettendosi in mostra in quanto immaginario, e la capacità opposta di passare inosservato, come qualcosa di “naturale”. In questo modo un immaginario che riflette su sé stesso può realizzare operazioni critiche, meta-semiotiche, mentre un immaginario che esprime valori ideologici e mitici tende a nascondersi e ad evitare il meta-discorso.

### 3. Antro *black box*: opacità, profondità e automazione

Come già accennato, dal punto di vista semantico, la metafora della *black box* coinvolge le dicotomie tra interno ed esterno, visibile e invisibile. Attraverso una specifica indicazione cromatica, l’interno costituisce l’elemento invisibile per via di

<sup>15</sup> La prospettiva semiotica si interessa da tempo a tali problematiche, a inclusione dell’intelligenza artificiale; come scrive Massimo Leone: “Una filosofia della comunicazione digitale orientata alla semiotica mira a leggere le tecnologie del senso nel lungo periodo della storia dei sistemi semiotici umani, per rivelare le ideologie implicite che stanno alla base della creazione di nuovi dispositivi, processi e artefatti di significato. L’intelligenza artificiale non fa eccezione, poiché il suo sviluppo è solitamente sostenuto da preconcetti specifici su cosa sia l’intelligenza, su come dovrebbe funzionare e su quali tipi di risultati dovrebbe generare nel mondo.” (Leone, 2023: 35)

un'opacità che in vari discorsi riguardo gli algoritmi e l'IA viene letta in termini di *opacità*. Un'opacità interna che risulta da una stratificazione di reti neurali che si rende profonda: come chiarisce De Seta riguardo l'utilizzo della metafora della profondità che oscilla tra "structural description and metaphorical explanation, technical aspiration and mystifying imaginary" (2023: 261) nelle ricerche del *machine learning* e specificamente del *deep learning*, nel cui specifico caso viene sfruttata la "semantic mystique to promote products and services" (262) – come dimostra Leone (2024), infatti, la figura della *black box*, con la sua aura di mistero, consiste in un dispositivo retorico volto alla promozione di queste tecnologie. All'interno del relativo discorso scientifico, tale profondità si connota secondo tre diverse concezioni: l'accuratezza del modello dovuta alla sua struttura complessa; come idea di rappresentazione della stratificazione interna di una rete neurale e come impenetrabilità, riferita all'impossibilità di comprendere e interpretare le operazioni algoritmiche ivi contenute (Arnold, Tinton, 2021, in De Seta, 2023).

L'opacità della *black box* è inoltre scalare, cioè si applica a diversi livelli che coinvolgono l'IA, come mostra Burrell (2016): a partire dal proprio funzionamento interno e dal modo in cui gli algoritmi operano, al linguaggio di programmazione che rimane inaccessibile alla maggioranza della popolazione e, infine, più in generale, alle forme di protezione e segreto aziendale che non rendono accessibili i propri codici e modelli. Di fatto, dunque, come scrivono Roberge e Seyfert: "there is not one box, but multiple boxes. The opacity of algorithms is more precisely expressed in different forms of opacity, all of which, in specific ways, are contingent on the inbetweenness of a plethora of actors, both human and non-human" (2016: 2).

La metafora della *black box*, dunque, si applica a più livelli e, in ragione della sua sfaccettatura semantica, viene veicolata in diversi apparati discorsivi a seconda delle sue possibili declinazioni. Tanto che, data la contemporanea automazione algoritmica stratificata che coinvolge pervasivamente gli oggetti e le piattaforme digitali della vita quotidiana, senza che questi processi possano essere compresi scalarmente nell'operatività decisionale che li contraddistingue, la società contemporanea è stata definita nei termini di "black box society" (Pasquale, 2015), a cui corrisponde una necessità di trasparenza correlativamente articolata<sup>16</sup>.

Ciò è il risultato di una tendenza per cui all'aumentare della densità e dell'opacità dei processi computazionali si produce uno scollamento tra la *possibilità di vedere* e la *possibilità di conoscere* (Chun, 2011: 17). Come spiega Latour, questo processo rientra nelle logiche di lavoro scientifiche e tecniche: nel momento in cui una tecnologia si dimostra efficiente, il punto focale non diviene più il suo funzionamento interno, che diviene "opaque and obscure" (Latour, 1999: 304), ma il suo risultato. In tal modo, l'efficienza diviene un meccanismo retorico e persuasivo che retroattivamente *mitizza* la tecnologia: come illustra Andreas Hepp, "this blackboxing is part of technology's process of materialization: a technology as black box is rarely questioned but taken

<sup>16</sup> Si veda Crawford e Ananny (2018) per un riepilogo riguardo le sfaccettature semantiche della "trasparenza" relativamente all'opacità algoritmica e i limiti che ne contraddistinguono la funzionalità in quanto strategia.



for granted as something 'quasi-natural' which is 'just there' and 'just works'" (2019: 78).

L'efficienza delle attuali tecnologie di IA deriva dalle potenzialità algoritmiche su cui poggiano. L'attuale definizione e concezione degli algoritmi proviene dai processi di meccanizzazione e computazionalizzazione che, all'interno della cornice contestuale della razionalizzazione moderna di politica, economia e scienza (Daston, 2022), hanno stabilizzato il termine nell'indicare "any process of systematic calculation, that is a process that could be carried out automatically" (Chabert, 1999: 1-2), la cui caratteristica è l'efficacia (Knuth, 1997: 4-6)<sup>17</sup>. Gli algoritmi sono istruzioni operative automatizzate per la risoluzione di problemi che ne condizionano la progettazione: nella loro programmazione, gli algoritmi sono progettati come "functionally *automatic*, to act when triggered without any regular human intervention or oversight" (Gillespie, 2014: 170). L'automazione diviene dunque *black boxed*, nel momento in cui, data l'impossibilità di interpretarne il funzionamento, ne si problematizzano solo i risultati.

Nella costellazione di discorsi che provengono e costituiscono l'immaginario dell'IA, spesso con fini retorici legati a interessi economici, di tipo utilitaristico (Santangelo, 2023), l'idea di automazione<sup>18</sup> che le tecnologie portano con sé può assumere connotazioni diverse, euforiche o disforiche a seconda dei contesti. All'interno del discorso scientifico, in accordo con la logica della «mechanical objectivity» (Galison e Daston, 2007: 115-183), l'assenza di intervento umano diviene baluardo di obiettività e accuratezza e gli algoritmi, nella loro automazione inumana, divengono "stabilizers of trust, practical and symbolic assurances that their evaluations are fair and accurate, free from subjectivity, error or attempted influence" (Gillespie, 2014: 179). La mitologia dell'efficienza dell'automazione viene, dunque, euforicamente valorizzata come promessa di obiettività e accuratezza. Sempre all'interno dei discorsi scientifici e anche non scientifici, l'idea di un'automazione *black boxed* "is closely tied to conspiracy theory sentiments in which AI is an occult power that cannot be studied, known, or politically controlled" (Pasquinelli e Joler, 2021: 1265), dando luogo a immaginari distopici in cui la macchina soprassiede all'umano in cui l'automazione viene disforicamente valorizzata<sup>19</sup>. Particolare testimone della valorizzazione disforica è la lettera pubblicata a marzo 2023<sup>20</sup> firmata da Musk e Mustaque (CEO di Stability AI) fra gli altri, in cui viene richiesta dai firmatari una pausa dell'avanzamento della ricerca sull'IA in nome di una serie di ragioni che esprimono le paure e le preoccupazioni riguardo a una perdita di controllo dovuta a un'automazione in grado di prevalere sull'umano.

<sup>17</sup> Per un'approfondita ricostruzione archeologica del termine e del concetto di algoritmo si veda Pasquinelli 2023, pp. 34-38.

<sup>18</sup> Anche l'automazione stessa consiste in un mito. Per una ricognizione sulla genealogia e le implicazioni a livello dell'immaginario dell'automazione si veda Munn (2022).

<sup>19</sup> Si rimanda a Cave, Dihal e Dillon (2020) per la panoramica a riguardo, come a Bory e Bory (2015) specificamente per il cinema.

<sup>20</sup> "Pause Giant AI Experiments: An Open Letter", 23 marzo 2023, sul sito di Future Of Life Institute, <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>.



Lo stretto legame esistente a livello metaforico e immaginativo tra *black box* e automazione dimostra quanto quest'ultima, intesa come promessa di efficienza, si caratterizzi di un'opacità che coinvolge diversi livelli e strati della tecnologia e dell'infrastruttura in cui essa agisce come prodotto e motore.

Al contempo, al tempo presente, diversi approcci e utilizzi della tecnologia dell'IA possono essere letti attraverso le lenti di questo immaginario, ossia come pratiche che si costituiscono a partire dall'immaginario stesso e che, a ritroso, lo costituiscono, in modo reciproco. Da un lato, l'attualizzazione della prospettiva intesa come *mitica* dell'immaginario della *black box*, cioè che naturalizza l'utilizzo della tecnologia in ragione della sua automazione efficiente; dall'altro, l'attualizzazione della prospettiva che, in contrasto, viene considerata *critica*, la quale mira a decostruire la retorica opaca della *black box*. I casi presi in esame per la prima prospettiva sono quelli relativi alla rappresentazione dell'IA in contesto bellico; i casi analizzati per la seconda prospettiva riguardano due opere al confine tra arte e informatica, che cercano di render conto del funzionamento dell'IA.



#### 4. Prospettiva mitica: IA bellica tra immaginario magico e religioso

Le tecnologie digitali sono da sempre oggetto di interesse per la ricerca e l'applicazione militare, soprattutto per quanto riguarda l'ambito della "visione aumentata" (Eugeni, 2021, pp. 83-96), che riguarda le implementazioni della capacità di vedere e agire efficacemente sul campo di battaglia; non a caso la prima versione della *black box*, il magnetron, era un dispositivo simile, un prototipo di radar. La storia delle tecnologie belliche ci mostra inoltre che, parallelamente a un aumento della qualità e dell'ampiezza dello sguardo protesico, aumenta anche la distanza da cui è possibile esercitare tale sguardo, e di conseguenza aumenta la distanza da cui si eseguono operazioni belliche. In questo senso i droni con il loro sguardo onnisciente, costante, quasi divino, "onniveggente" (Chamayou, 2013), condividono le caratteristiche delle più recenti tecnologie di IA; ovvero si basano sull'elaborazione di grandi quantità di dati.

Lo sviluppo di tecnologie IA ha avuto un impatto sull'universo discorsivo bellico, dove si possono notare processi di valorizzazione euforica dell'efficacia automatizzata delle tecnologie, o al contrario valorizzazioni disforiche e relative critiche. Ciò che preoccupa maggiormente rispetto all'introduzione di tecnologie di IA in ambito bellico riguarda l'etica decisionale che presiede al loro utilizzo e che entra in relazione con l'opacità a più livelli della *black box*.

Con lo scoppio della guerra russo-ucraina l'IA è entrata a pieno titolo tra le armi del conflitto, non senza polemiche, e a tal proposito Alex Karp, CEO dell'azienda "Palantir" leader in questo campo, ha sostenuto la necessità di "un'architettura che permetta la trasparenza sui dati-sorgente"<sup>21</sup>. L'attenzione alla trasparenza dei dati

<sup>21</sup> Frase pronunciata da Karp, tradotta autonomamente, e riportata in Sterling, e van Den Berg (2023): <https://www.reuters.com/technology/ukraine-war-shows-urgency-military-ai-palantir-ceo-says-2023-02-15/>

emerge quindi come condizione necessaria per un uso etico delle tecnologie IA in ambito bellico (e non solo), ma come vedremo, nella realtà dei fatti questa trasparenza non è quasi mai realizzata.

Inoltre, nel discorso di Karp, pronunciato in un summit internazionale, il riferimento alla dimensione dell'immaginario è volontariamente escluso; ma, come visto in precedenza, le figure dell'immaginario si annidano anche nei contesti più reali, e in questo caso svolgono una funzione di mediazione tanto nella rappresentazione quanto nella percezione comune di tecnologie IA reali a tutti gli effetti.

Infatti, considerando il livello della rappresentazione identitaria dell'azienda di Karp, è possibile notare diversi riferimenti a dimensioni immaginarie, a partire dal nome stesso dell'azienda. La scelta del nome "Palantir" è infatti un esplicito riferimento al mondo de *Il signore degli anelli*: nello specifico, il Palantir è un oggetto magico che permette di vedere a distanza, ma il cui funzionamento non è del tutto chiaro (Tolkien, 1946). Il riferimento a un oggetto dai poteri magici sovrumani è funzionale a dare l'idea di una tecnologia potente ed efficace ben oltre le competenze e conoscenze umane, e in maniera non troppo velata evoca il principio di inconoscibilità caratteristico della *black box*. Prestando poi attenzione al breve video che viene riprodotto automaticamente appena si accede alla sezione "defense" del sito di Palantir<sup>22</sup>, si può notare un altro esempio di ibridazione di immaginari, in questo caso con il ricorso a un'estetica ricalcata su videogame e film fantascientifici: montaggio rapido, luoghi estremi, desertici e glaciali, dove compaiono veicoli militari, mentre gli attori umani osservano rigorosamente a distanza, il tutto accompagnato da una musica elettronica dai toni epici. L'immaginario tecnologico-bellico delle IA si carica di significati propri dell'immaginario fantasy, cinematografico e videoludico, affiancandosi a esso piuttosto che ritagliandosi un proprio spazio esclusivo, sfruttandolo a fini retorici, portando alla costruzione di una figura "mitica"<sup>23</sup> della tecnologia IA.

Come sostiene Barthes, il processo di mitizzazione di un oggetto – in questo caso la tecnologia bellica – dà vita a una "*parola depoliticizzata*" (1974: 223). Il mito



<sup>22</sup> "Defense", Palantir, <https://www.palantir.com/offerings/defense/>.

<sup>23</sup> Riprendendo l'idea di mito di Barthes, possiamo sostenere che siamo di fronte a oggetti reali che acquistano qualità mitiche; si tratta di un processo di "deformazione del senso" di un oggetto, che si riveste di una significazione ulteriore, una "connotazione" (cfr. Barthes, 1974: 191-238) funzionale a uno scopo retorico, molto spesso di tipo ideologico. Per Barthes il mito "impoverisce" il senso, "lo allontana", ma allo stesso tempo "lo tiene a disposizione"; infatti, continua Barthes, "per quanto paradossale possa sembrare *il mito non nasconde nulla*: la sua funzione è di deformare, non di far sparire." (p. 203) Recentemente il concetto di mito è stato ripreso in semiotica per l'analisi di nuovi media (Stano e Ponzio, 2023); giustamente Stano sottolinea anche il carattere positivo, euforico, del mito, inteso come meccanismo di coesione sociale e comunitaria, aspetto che nella lettura barthesiana fosse rimasto in ombra. Convinti che il mito svolga anche e soprattutto tali funzioni, la nostra prospettiva si preoccupa maggiormente di mettere in luce i processi di connotazione e deformazione del senso che interessano l'oggetto del mito, nel tentativo di mostrare successivamente operazioni contrastanti, che definiamo critiche, che tentano in diverso modo di mettere in mostra i presupposti ideologici su cui si basa la costruzione, la rappresentazione e la percezione della tecnologia IA.

funziona dunque come “un metalinguaggio, basato su operazioni connotative, che naturalizza determinate visioni ideologiche” (Stano e Ponzio, 2023: 12).

Come ricordato in precedenza ogni utilizzo e ogni rappresentazione della tecnologia si carica anche di valenze politiche ed economiche; tuttavia, un simile processo di mitizzazione dell’oggetto tecnologico è funzionale a deformare quelle responsabilità politiche, intrinsecamente umane, che sono coinvolte nell’uso e nella rappresentazione della tecnologia, e che sono irriducibili a qualsiasi tipo di determinismo tecnologico (Eugeni, 2021).

#### **a. *The Gospel: immaginario religioso e il mito dell’efficienza dell’IA***

La questione della responsabilità etica e politica trova spazio anche e soprattutto nel contesto dell’attuale guerra a Gaza, dove l’uso di un software di IA da parte dell’esercito israeliano ha scatenato diverse critiche.<sup>24</sup> In questo caso, è possibile osservare la costruzione di un legame tra immaginario religioso e immaginario bellico all’interno in un discorso teso a giustificare l’uso di armi IA e più in generale l’operato militare israeliano<sup>25</sup>.

Il termine in ebraico che designa il software di IA sul sito dell’IDF è “Habsora”, tradotto in inglese con “Gospel”, contrazione di “God’s spell”, ovvero parola di Dio, che in italiano corrisponde a “Vangelo”, mutuato dal greco “*euangelion*”, ovvero “buone notizie” (Della Ratta, 2024). La funzione del software è generare obiettivi militari nemici da colpire sulla base di un lavoro di analisi computazionale su dati di Intelligence di cui dispone IDF. Come si può leggere nella pagina web di IDF, “Gospel è un sistema che consente l’uso di strumenti automatici per produrre obiettivi a un ritmo veloce e funziona migliorando materiale informativo accurato e di alta qualità secondo le esigenze”<sup>26</sup>. Si evince dunque un forte legame tra l’automazione e la rapidità del processo di selezione degli obiettivi, misurata in quantità di edifici abbattuti. Anche qui ci si ritrova di fronte a un processo di mitizzazione che sfrutta il sostrato immaginario del discorso religioso per connotare l’efficacia e la rapidità del software, allontanandolo dal suo reale contesto di applicazione<sup>27</sup>.

La retorica della precisione e dell’efficienza dell’IA si appoggia qui all’immaginario religioso per costruire una figura mitica della tecnologia e circolarmente anche dell’esercito che la usa. L’immaginario religioso costituisce un luogo di coesione



<sup>24</sup> Il primo articolo reperibile è stato pubblicato da *The Guardian* il 3/12/2023, ripreso e commentato da testate giornalistiche di diversi paesi; sono seguiti altri articoli sullo stesso tema pubblicati su *Local Call* e *+972 Call*.

<sup>25</sup> Già nel 2019 sul sito di IDF era possibile leggere la brochure di presentazione di un nuovo tipo di drone, “Harpy”, che “funziona in maniera totalmente automatica”. Non si trattava ancora di un vero e proprio modello di IA con deep learning ma la retorica atta a valorizzare l’efficacia di un funzionamento autonomo è simile. Anche qui il riferimento alla creatura mitologica è sintomatico di una tendenza all’ibridazione tra immaginario bellico e immaginario sovrumano.

<sup>26</sup> Citazione tratta dal sito in questione in traduzione italiana di Google, di cui il link alla pagina web di IDF dove viene presentato il software Gospel.

<sup>27</sup> Infatti, come scrive Donatella Della Ratta in un commento all’inchiesta di *The Guardian*: “despite its biblical namesake suggesting absolute truth and an infallible message, *Habsora*, like all machine learning systems, operates on correlation rather than causation and likelihood rather than certainty”.

sociale e politica per molte collettività, nello specifico quella israeliana – non a caso i soldati morti nella guerra sono designati da IDF come “martiri di guerra”<sup>28</sup> – e serve a interpretare la tecnologia in chiave teleologica, come qualcosa che naturalmente porta alla prosperità della comunità che la usa.

Sempre nella pagina web di presentazione di Gospel, compare anche un video che mostra proprio i risultati del lavoro del software, o meglio, la loro interpretazione pragmatica; si tratta di una serie di bombardamenti in inquadrature da pochi secondi l’una, con un effetto di valorizzazione estetica della velocità delle prestazioni dell’IA, all’interno di una strategia che tende a valorizzare maggiormente i risultati, piuttosto che i processi che presiedono a quest’ultimi, e da cui dipende la logica che li interpreta come efficaci. Paradossalmente nella narrazione dell’esercito israeliano l’accento viene posto anche sulla competenza umana di controllo delle informazioni, nel tentativo di umanizzare un processo che, supervisionato o meno da attori umani, è causa di morte e devastazione in quantità e modalità spaventose.

Questo processo di umanizzazione dell’utilizzo del software di IA da parte di Israele si appoggia da un lato a una retorica della disumanizzazione del nemico, e dall’altro è a sua volta funzionale a deformare e nascondere ogni tipo di responsabilità politica delle persone e delle istituzioni coinvolte nell’uso di tali armi. Se per la narrazione di IDF il controllo dell’umano è totale, le fonti intervistate raccontano invece di selezioni ed elaborazioni di dati grossolane, e di scelte strategiche che consapevolmente accettano la morte “collaterale” di numerosi civili<sup>29</sup>. Si può parlare a questo proposito anche di discorso o immaginario ideologico, secondo la definizione di Eco<sup>30</sup>; se per Barthes il mito deforma, per Eco l’ideologia nasconde effettivamente le ragioni della sua stessa enunciazione, “bloccando in questo modo la funzione metasemiotica del discorso” (Stano, 2022: 10). Un discorso come quello portato avanti da IDF, servendosi dell’immaginario religioso, diventa ideologico in quanto blocca la natura meta-semiotica del discorso, celando i criteri che sono alla base dell’emersione di un certo risultato.

È possibile allora ritornare all’idea di Karp sulla necessità di una trasparenza sui dati-sorgente; nel caso di Gospel i dati di intelligence vengono usati per etichettare gli abitanti di Gaza, i quali ricevono un punteggio indicativo della loro vicinanza al gruppo armato di Hamas. Tuttavia, questi dati non sono pubblici, e non si tratta di una *black box* tecnica, quanto piuttosto di una *black box* politica, di stampo ideologico, funzionale a celare i criteri di pertinenza per cui un soggetto è considerato più o meno vicino a Hamas, e la cui morte è quindi più o meno giustificata. Più profondamente, se manca una trasparenza relativa ai dati-sorgente



<sup>28</sup> Denominazione che compare sul sito di IDF, di cui il link.

<sup>29</sup> Si tratta di una specifica strategia di IDF, detta “dahiya doctrine”, utilizzata per la prima volta nella guerra in Libano nel 2006, e si basa su una distruzione pressoché totale delle infrastrutture civili, considerate come centri di potere del nemico.

<sup>30</sup> Secondo Eco, “l’ideologia è visione del mondo parziale e sconnessa [...] essa cela anche le ragioni pratiche per cui certi segni sono stati prodotti” (Eco, 1975: 369). L’ideologia, infatti, consiste proprio in un tipo di discorso che parte da un’interpretazione parziale e selettiva del mondo, e sceglie di considerare interpretativamente solo quei tratti di significato funzionali al suo progetto retorico. Sui rapporti tra mito e ideologia, concetti che a volte sembrano coestensivi, si veda Leone e Stano (ed.) 2023.

un algoritmo *può essere sfruttato per produrre qualsiasi risultato*, e non sempre è possibile ricostruire il processo che ha prodotto tale risultato.

Il panorama del discorso bellico sta cambiando attraverso l'uso e la rappresentazione delle novità tecnologiche, e si sta assistendo ad un processo di depoliticizzazione e de-responsabilizzazione degli attori umani all'interno del contesto operativo militare. Tale processo si avvale di una più o meno tacita valorizzazione euforica del progresso tecnologico, inteso come indispensabile, giusto, e quindi, di fatto, naturalizzato e assunto come mito della contemporaneità. Infine, il potere delle tecnologie chiamate "intelligenti" si basa su un processo di antropomorfizzazione delle stesse, e contemporaneamente permette agli attori umani implicati nelle operazioni militari di assumere un'identità quasi anonima, relegando la responsabilità dell'azione a un software di cui spesso non si conoscono tutti i dettagli relativi al funzionamento, e che si basa su una forma di estrazione massiva di dati.

A fianco di narrazioni che costruiscono le tecnologie attraverso rimandi agli immaginari fantastici e fantascientifici, si può intravedere un cambiamento reale, relativo alla percezione del rischio e delle conseguenze della guerra. Infatti la portata operativa delle tecnologie militari è arrivata a un'efficacia mai vista prima (misurata in termini di velocità di distruzione, come abbiamo visto con il software Gospel) tale da rendere la guerra stessa un fenomeno dalle caratteristiche fantascientifiche. Il coinvolgimento dell'umano è sempre più distante dal contesto fisico della guerra, poiché è distante lo sguardo protesico esercitato dai dispositivi tecnologici, i quali elaborano quantità di dati la cui scala è praticamente impossibile da maneggiare per l'umano. Questo cambiamento nel modo di condurre e pensare la guerra, già in atto prima delle recenti IA (cfr. Chamayou, 2013), si è consolidato anche grazie alla retorica dell'opacità, caratteristica di una black box miticizzata.

Per questa ragione vengono presentati ora quei casi di interpretazione critica della *black box* delle IA, che costruiscono un immaginario contrapposto a quello qui presentato, attraverso operazioni orientate alla visualizzazione dei processi di categorizzazione ed elaborazione di dati che spesso rimangono invisibili, e spesso vengono giustificati in base a una retorica dell'efficienza "naturale".

## 5. Prospettiva critica: arte del(l'in)visibile

All'interno dell'immaginario, l'articolazione semantica della metafora della *black box* imperniata sulle categorie dicotomiche di visibile e invisibile, interno ed esterno, delinea degli assi di azione su cui operare. Procedendo da un elemento della categoria all'altro o viceversa, tali assi costituiscono le logiche di operazioni che si strutturano e atualizzano in quanto tali in ragione delle possibilità offerte dalla delimitazione semantica insita nell'immaginario. All'interno del contesto dell'arte contemporanea, si è affermata una tendenza che agisce sull'asse della visualizzazione, ossia di rendere visibile l'invisibile, processo spesso indicato con la dicitura "opening the black box" (Wasielowski, 2023).



Tale tendenza<sup>31</sup> ha costituito il nucleo progettuale di operazioni artistiche a partire già dal decennio scorso, in cui il flusso globale iconico e informativo, l'avvento dei Big Data e la presa di consapevolezza circa l'opacità degli algoritmi trainanti la società digitale han dato luogo alla necessità di rendere i processi algoritmici visibili, rendere conto delle responsabilità di quest'ultimi e sfidare l'apparente immaterialità del digitale.

In tal modo, l'arte si (ri)configura come una "scienza del visibile" che "funziona [...] in quanto contro-discorso orientato a profanare l'opacità strutturale delle tecniche di osservazione e dei processi di rappresentazione che manipolano la realtà (e la sua immagine) traducendola nel territorio della visibilità artistica" (Malavasi, 2017: 81). Ed è proprio questo processo traduttivo che mostra come l'immaginazione artistica si muova all'interno dell'immaginario sociotecnico della *black box* secondo un asse contrastivo e critico, che, nel caso specifico dell'IA, si oppone alla sua retorica.

La prospettiva critica del dispositivo artistico non può, dunque, non divenire politica nel momento in cui la *black box* dell'IA viene (relativamente) aperta, seguendo un approccio che Martin Zeilinger definisce tattico in senso decerteauiano nel momento in cui l'intento è di resistere a "strategic approaches that blackbox knowledge" e di opporsi al "blackboxing of technology" (2021: 51 e 154) perseguendo modalità di emancipazione dello spettatore o spettatrice (Rancière 2008) in accordo con una prospettiva di arte *critica* così come elaborata da Rancière<sup>32</sup>.

Pasquinelli e Joler dimostrano come l'IA consista in uno strumento estrattivo e organizzativo della conoscenza (2021) che viene diffratta nella realtà secondo le lenti del modello statistico: la conoscenza ingerita (i dati prelevati dal reale) viene sottomessa e poi restituita in una forma mediata da una specifica razionalità – scientifica, statistica, algoritmica – che, positivamente valorizzata e accettata come criterio decisionale, genera effetti sul reale. Rendere visibili i meccanismi della *black box* significa esporre la logica e la razionalità di questi "instrument of power" che operano secondo "an aesthetic and ideology of power" (Paglen, 2016).

Questa prospettiva critica è alla base di due opere che, intersecando arte, tecnologia e informatica, sono state esposte in spazi rilevanti per l'arte contemporanea. Come si procederà a dimostrare, queste due opere condividono il tentativo di visualizzazione di elementi che compongono la *black box* dell'IA, ossia delle logiche tassonomiche di organizzazione dei dataset che ne fungono da confini epistemici per il funzionamento e delle implicazioni visuali che la performatività algoritmica reca con sé.

#### **a. Immersi nel dataset: speleologia dentro la black box**

<sup>31</sup> Si tratta di una tendenza fra le tante sfaccettature delle operazioni che caratterizzano il campo della cosiddetta *AI Art*. Per una panoramica generale, cfr. Barale, 2020; Audry, 2021; Pedrazzi, 2021; Grba, 2022 e 2024.

<sup>32</sup> A questo proposito, sull'idea di arte critica di Rancière e il concetto di attivismo, si veda Trione, 2022, pp. 183-185.



*From 'Apple' to 'Anomaly' (Pictures and Labels)* è stata commissionata dal Barbican di Londra appositamente per la galleria The Curve all'interno del progetto *Life Rewired*. Inaugurata a fine settembre del 2019 e conclusasi nel febbraio successivo, pubblicitarimente descritta come "revealing the powerful, and often hidden, forces at play in artificial intelligence"<sup>33</sup>, l'installazione contava e si componeva di approssimativamente 30'000 fotografie stampate e appese sulla lunga parete curva dello spazio immersivo della galleria (fig. 1).



263



**Fig. 1** – Veduta dell'installazione "Trevor Paglen – From 'Apple' to 'Anomaly' (Pictures and Labels). Selections from the ImageNet dataset for object recognition", Barbican Centre, Londra. 26 settembre 2019 – 16 febbraio 2020. Foto di Max Colson. Courtesy dell'artista e del Barbican Centre –The Curve.

Le fotografie e le 200 categorie secondo le quali erano organizzate provengono da uno dei primi e più famosi dataset per l'allenamento dell'IA nel riconoscimento di oggetti, ossia Image Net, sul quale Paglen, in collaborazione con Kate Crawford, continuerà ad approfondire le proprie ricerche – segue infatti *Image Net Roulette*<sup>34</sup> – che sono poi culminate in un testo fondamentale intitolato *Excavating AI: The Politics*

<sup>33</sup>Come si può leggere sulla pagina dedicata presente sul sito del Barbican, <https://sites.barbican.org.uk/trevorpaglen/>.

<sup>34</sup> Progetto del 2019, che consisteva in una rete neurale addestrata tramite i dati relativi alla categoria "persona" presenti in Image Net. La rete neurale addetta al riconoscimento facciale riproponeva, dunque, le classificazioni problematiche contenute in Image Net, <https://www.chiark.greenend.org.uk/~ijackson/2019/ImageNet-Roulette-cambridge-2017.html>.

of Images in Machine Learning Training Sets<sup>35</sup>. Il titolo del saggio che corona questo percorso di ricerca è particolarmente interessante, perché permette di concepire retroattivamente la prima parte di questo scavo, ossia *From 'Apple' to 'Anomaly'*, come un tentativo di fornire un'esperienza speleologica all'interno della *black box* e, quindi, come esempio di una delle modalità di visualizzazione di quest'ultima.

ImageNet venne realizzato nel 2009 sotto la direzione di Fei Fei Li con l'intento di "mappare l'intero mondo degli oggetti"<sup>36</sup>: composto di 14 milioni di immagini, consiste in un'organizzazione di quest'ultime secondo categorie e tassonomie che sono necessarie per il funzionamento degli attuali sistemi di IA *machine learning*. Il dataset di allenamento rappresenta di fatto il pacchetto di conoscenza mediata astratta dal mondo reale che ne diviene i confini epistemici di azione (Crawford, 2021). Il modo in cui questa conoscenza viene trattata, organizzata, gerarchizzata si poggia su associazioni visivo-verbali arbitrarie che pretendono di far convergere due significanti diversi di due semiotiche differenti per il medesimo referente, recando con sé problematiche a livello epistemologico e politico (Crawford e Paglen, 2021). Tali associazioni, una volta ingerite all'interno dello spazio latente dell'IA, lo strato più profondo della *black box*, divengono inoltre non più leggibili e comprensibili dall'occhio umano (Paglen, 2016).

L'installazione di Paglen espone, rendendo visualizzabile, la logica semiotica di cui la *black box* dell'IA ha bisogno per funzionare e che operativizza nei meccanismi algoritmici. Una logica che regola, nelle parole di Paglen, "the political economy of meaning that is produced through such apparatuses" (2020). La visualizzazione messa in mostra si realizza tramite la traduzione visiva e materiale delle immagini del dataset in fotografie stampate e appese, rendendo manifesto il bisogno quantitativo di informazioni di un sistema di IA assieme a quello delle categorie verbali, riproposte sotto forma di etichette, per l'organizzazione del materiale visivo.

Questa visualizzazione avviene inoltre in uno spazio che ricrea a sua volta un effetto *black box*, dove, nell'immersività della galleria in cui, a parte la parete curva illuminata, a dominare è il buio, la conformazione spaziale dell'installazione ha reso spettatori e spettatrici "walkers", richiedendo "an active traversal of the gallery space" (Zylinska, 2020: 97). In questo senso, *From 'Apple' to 'Anomaly'* rappresenta un tentativo di speleologia di un antro della *black box* – una sua parete, in questo caso. Annunciata già dal titolo, il quale ricorda come le molteplici immagini in mostra siano solamente quelle appartenenti a un breve segmento di lemmi raggruppati in ordine alfabetico secondo la lettera iniziale, questa limitatezza, che fa i conti con le *contraintes* materiali-fenomeniche, attesta l'impossibilità di visualizzare nella sua totalità anche solo un elemento della *black box* richiedente incommensurabili



<sup>35</sup> Il progetto di Paglen e Crawford raggiunse una popolarità tale che rese di dominio pubblico le problematiche razziali, sessiste – e non solo – insite nei dataset di addestramento. Ciò provocò la rimozione di quasi un milione di immagini nel tentativo di correggere i bias. A riguardo, si veda Ruiz 2019, <https://www.theartnewspaper.com/2019/09/23/leading-online-database-to-remove-600000-images-after-art-project-reveals-its-racist-bias>.

<sup>36</sup> Citazione tradotta in italiano che appare in Gershgorn (2017), <https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world>.

quantità di dati. Questo a riprova del fatto che "nevertheless, the full interpretability and explicability of machine learning statistical models remains a myth" (Pasquinelli e Joler, 2021: 1265).

**b. Learning to See: sguardi stranianti sull'opacità algoritmica**

Se l'opera di Paglen mette in mostra il dataset, cioè esibisce le logiche tassonomiche e semiotiche che, a monte, garantiscono l'operatività della *black box*, l'installazione di Memo Akten rende visibile il modo con cui i meccanismi operativi algoritmici interni applicano la conoscenza fornitagli su input visivi, definendo una propria *way of seeing* (Berger, 1972).

Presentata alla mostra *AI: More than Human* tenutasi sempre al Barbican tra i mesi di maggio e agosto 2019, l'installazione di Memo Akten *Learning to See* è compresa all'interno dell'omonima serie<sup>37</sup>, a sua volta parte del suo progetto di ricerca iniziato nel 2017, culminato nella sua tesi di dottorato discussa nel 2021 alla Goldsmith University of London (Akten, 2021).

Parte della terza sezione, *Data World*<sup>38</sup>, l'installazione interattiva (fig. 2) si compone di due schermi appesi alla parete che esibiscono in diretta ciò che la videocamera riprende sul tavolo posizionato di fronte – un insieme eterogeneo di oggetti quotidiani – e quello che "vede" la rete neurale. Tramite la manipolazione tattile e diretta degli oggetti da parte dei visitatori e delle visitatrici diviene possibile osservare i cambiamenti figurativi nei due schermi, sia quello dell'immagine ritraente la disposizione degli oggetti sul tavolo che quello della corrispondente immagine generata dalla rete neurale, addestrata specificamente con cinque diversi e ristretti dataset<sup>39</sup> che si alternano ogni trenta secondi.

La simultanea e comparata apparizione in tempo reale di due modi di vedere – a destra, quello della camera, che ripropone a un livello di alta iconicità gli oggetti presenti e, a sinistra, quello dell'IA, che mostra la trasmutazione degli stessi – fornisce allo spettatore o alla spettatrice la possibilità di uno sguardo straniante sulla rappresentazione in atto secondo una logica di *showing seeing* (Mitchell, 2002). Giocando sull'effetto di neutrale obiettività proposto dalle immagini della camera, che emerge in contrasto a quello di *irrealità* condizionata dell'IA, l'opera di Akten pone in discussione la percezione visiva di chi osserva e, contemporaneamente, di quella programmata della macchina intelligente.

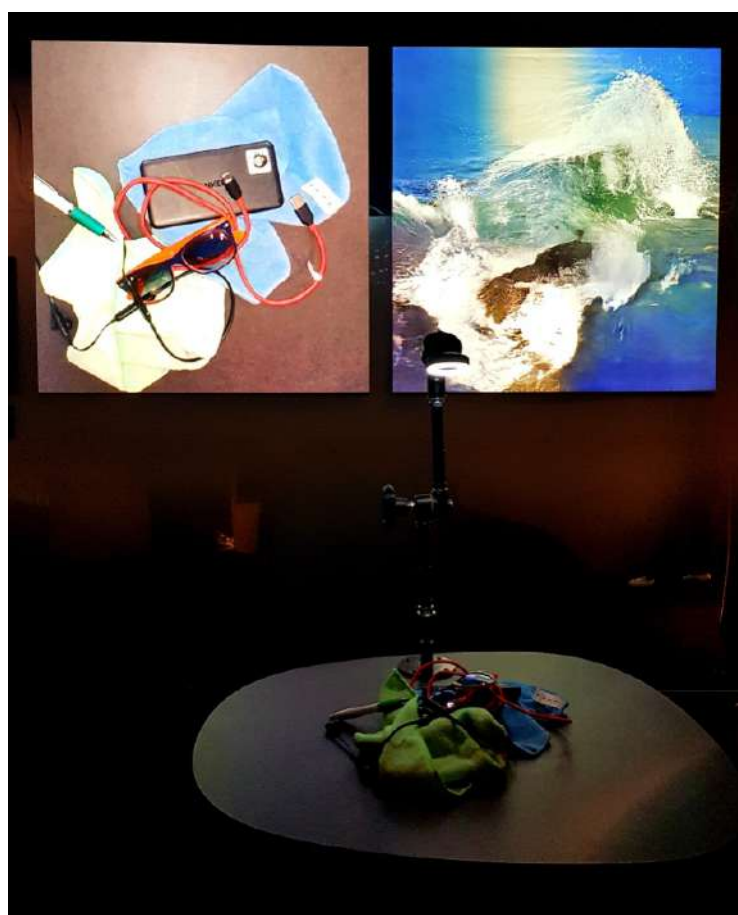


<sup>37</sup> Attraverso la realizzazione, lo sviluppo e l'addestramento di una rete neurale profonda, il progetto di Akten intende indagare le modalità di comprensione visuale della realtà di tale "visual instrument" (Akten et al., 2019) e, a un livello più generale, i modelli di *deep learning* in quanto nuovi medium artistici. <https://www.memo.tv/works/learning-to-see/>

<sup>38</sup> Qui di seguito il comunicato stampa della mostra, <https://www.barbican.org.uk/sites/default/files/documents/2019-07/FINAL%20AI%20More%20than%20Human%20exhibition%20release%20280219.pdf>.

<sup>39</sup> Ciascuno dei cinque dataset si compone unicamente di immagini relative ai quattro elementi naturali (acqua, aria, terra e fuoco) e al cosmo.

Nel mettere in mostra la performatività dell'algoritmo, *Learning to See* permette di comprendere come la visione dell'IA coinvolga una dinamica interpretativa che dipende dalla conoscenza ingerita nella sua *black box*, rifratta sulla realtà secondo logiche statistiche; retroattivamente, mette in discussione anche una concezione di percezione pura attribuita alla vista umana (Audry, 2021; Bueno e Abarca, 2020: 1185). Nello stesso modo in cui si verifica nel caso analizzato da Voto (2022) seguendo le categorie di opacità e trasparenza di Marin (1994), la performance algoritmica in tempo reale, che emerge dal contrasto con un'immagine *più reale* di ciò che viene ritratto, *presenta* a chi osserva la rappresentazione dell'IA, smontandone la trasparenza e rendendo in tal modo opaco, quindi intelligibilmente "visibile", lo spazio latente che ne condiziona la forma, "redeeming it from the inaccessibility with which the uninitiated normally deal" (Voto, 2022: 50).



**Fig. 2** – Veduta dell'installazione "Learning to See" inclusa in *AI: More than Human*, Barbican Centre, Londra, 2019. Courtesy dell'artista.

Tuttavia, la limitatezza di conoscibilità si riscontra anche qui: la *black box* rimane *inaccessibile* al suo interno e *invisibile* all'esterno. L'unica possibilità consiste nel coglierla a posteriori, negli output che realizza, con la consapevolezza della costruzione, parzialità e problematicità epistemologica del suo sguardo.

## 6. Conclusione

Semioticamente inteso, l'immaginario dell'IA si dimostra non solo come una rappresentazione finzionale della tecnologia, ma come una dimensione di senso che contribuisce attivamente alla sua creazione e interpretazione; allo stesso tempo e di conseguenza, l'immaginario plasma e viene plasmato dai relativi discorsi culturali e socio-tecnologici. All'interno di questa circolarità, la semantica della figura della *black box* apre possibilità di dinamiche e percorsi di senso che ne influenzano le attualizzazioni, ossia quelle pratiche e rappresentazioni che, coinvolgendo l'IA, ne rinforzano la retorica dell'automazione efficiente o la criticano tramite la decostruzione dell'opacità.

Dopo aver ricostruito la rete delle sfaccettature semantiche che compongono la figura della *black box* (profondità, opacità, automazione), tra oscurità e opacità a più livelli, talvolta discordanti, si è voluto mettere in luce come da tale metafora si articolino discorsi con valorizzazioni differenti a seconda dei contesti. La comparazione di porzioni locali dell'immaginario della *black box*, quello bellico e quello artistico, ha dunque permesso una ricognizione – che non pretende di essere esaustiva – del significato e dell'utilizzo di tale metafora nel caso dell'IA e, dunque, delle sue conseguenze politiche, sociologiche ed estetiche.

L'incursione nei territori dell'arte contemporanea potrebbe continuare approfondendo i diversi livelli dell'esperienza estetica con l'IA e le diverse strategie di enunciazione utilizzate nel tentativo di visualizzare i processi che presiedono all'emergenza di determinati risultati. Parallelamente si potrebbe proseguire con il confronto tra queste operazioni, definite critiche, e i discorsi ideologici e mitici discussi in ambito bellico con i relativi processi di deformazione che rendono opachi e dunque inconoscibili diversi aspetti dell'IA, in una costante negoziazione tra opacità tecnica e strategie di nascondimento e di inganno.

Ad ogni modo, l'analisi qui presentata ha permesso di far emergere e sottolineare la capacità dei discorsi critici di riflettere sulle implicazioni politiche che questa metafora porta con sé, soprattutto nelle sue attualizzazioni e nei relativi contesti operativi. L'approccio critico e straniante dell'ambito artistico nei confronti dell'invisibilità opaca della *black box* permette di rileggere a ritroso i discorsi che si avvalgono della potenza retorica di tale figura provenienti da altri ambiti in modo più critico, ponderandone gli effetti.

In conclusione, la metafora della *black box*, nella sua costitutiva parzialità e località, riflette e contribuisce alla complessità e alla diversità di narrazioni e pratiche relative all'immaginario dell'IA, rappresentando sia una sfida che un'opportunità



Lorenzo Aimo, Luca Brunet  
*Immaginario Black Box*

nell'interpretazione e nella definizione del rapporto tra questa tecnologia pervasiva e la società.



---

268



## Bibliography

- Akten M. (2021), *Deep Visual Instruments: Realtime Continuous, Meaningful Human Control over Deep Neural Networks for Creative Expression*, Doctoral thesis, Goldsmiths, University of London [Thesis], <https://doi.org/10.25602/GOLD.00030191>.
- Akten M., Fiebrink R., Grierson M. (2019), "Learning to see: you are what you see", in ACM SIGGRAPH 2019 Art Gallery (SIGGRAPH '19). Association for Computing Machinery, New York, NY, USA, Article 13, 1-6. <https://doi.org/10.1145/3306211.3320143>.
- Alexander J.C. (2008), "Iconic consciousness: the material feeling of meaning", *Environment and Planning D: Society and Space*, 26, pp. 782-794.
- Arnold T., Tilton L. (2021), "Depth in Deep Learning: Knowledge, Layered, Impenetrable", in Redrobe K., Scheible J. (ed.) *Deep Mediations*, Minneapolis, University of Minnesota Press, pp. 309-329.
- Ashby R. W. (1957), *An Introduction to Cybernetics*, London, Chapman&Hall LTD.
- Audry S. (2021), *Art in the Age of Machine Learning*, London & Cambridge (MA), The MIT Press.
- Barale A. (2020), *Arte e Intelligenza Artificiale. Be my GAN*, Milano, Jaca Book.
- Barthes R. (1957), *Mythologies*, Seuil, Paris, (trad. it. *Miti d'oggi*, Torino, Einaudi, 1974).
- Berger J. (1972), *Ways of seeing*, London, Penguin Books Ltd.
- Bory S. e Bory P. (2015), "I nuovi immaginari dell'Intelligenza Artificiale", *Im@go. A Journal of the Social Imaginary*, n. 6, pp. 66-85, doi: <https://doi.org/10.7413/22818138047>
- Burrel J. (2016), "How the machine 'thinks': Understanding opacity in machine learning algorithms", *Big Data & Society*, 3(1), pp. 1-12, <https://doi.org/10.1177/2053951715622512>.
- Castoriadis C. (1975), *L'institution imaginaire de la société II. L'imaginaire social et l'institution*, Paris, Editions du Seuil (trad. it. *L'istituzione immaginaria della società*, Torino, Bollati Boringhieri, 1995).
- Cave S., Dihal K (2020), "The Whiteness of AI", *Philosophy & Technology*, 33:685-703, <https://doi.org/10.1007/s13347-020-00415-6>.





- Cave S., Dihal K. (ed.) (2023), *Imaging AI. How the World Sees Intelligent Machines*, Oxford, Oxford University Press.
- Cave S., Dihal K., Dillon S. (ed.) (2020), *AI Narratives. A History of Imaginative Thinking about Intelligent Machines*, Oxford, Oxford University Press.
- Chabert J. (ed.) (1999), *A History of Algorithms: From the Pebble to the Microchip*. Berlin, Springer.
- Chamayou G. (2013), *Théorie du drone*, Paris, La Fabrique Éditions.
- Chun W. H. K. (2011), *Programmed Visions: Software and Memory*, Cambridge (MA), The MIT Press.
- Crawford K. (2021), *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*, New Haven & London, Yale University Press (trad. it. *Né intelligente, né artificiale. Il lato oscuro dell'IA*, Bologna, Il Mulino).
- Crawford K., Annany M. (2018), "Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability", *new media & society*, 20(3), pp. 973-989, <https://doi.org/10.1177/146144481667664>.
- Crawford K., Paglen, T. (2021), "Excavating AI: The Politics of Images in Machine Learning Training Sets", *AI & Society*, 36:1105-1116, <https://doi.org/10.1007/s00146-021-01162-8>.
- Daston L. (2022), *Rules. A Short Story of What We Live By*, Princeton & Oxford, Princeton University Press.
- De Seta G. (2023), "Digital Depth: A Volumetric Speculation", *magazén*, 4(2), pp. 245-270, doi:10.30687/mag/2724-3923/2023/02/004.
- Della Ratta D. (2024), "Killing intelligence: death by tech and other ordinary horrors in Gaza", *Untoldmag*, consultato il 25 aprile da: <https://untoldmag.org/killing-intelligence-death-by-tech-and-other-ordinary-horrors-in-gaza/>
- Eco U. (1964), *Apocalittici e integrati: comunicazioni di massa e teorie della cultura di massa*, Milano, Bompiani.
- Eco U. (1975), *Trattato di semiotica generale*, Firenze, Giunti editore.
- Eugeni C. (2021), *Capitale algoritmico. Cinque dispositivi postmediali (più uno)*, Brescia, Scholé.
- Ferraro G. (2011), "Dinamiche dell'immaginario: una prospettiva semiotica", in Leone M. (ed.) *Immaginario*, Lexia 07-08, Roma, Aracne editore, pp. 91-102.





Ferraro G. (2023), "Kito e realtà dell'IA: uno sguardo semiotico", in Leone M. e Santangelo A. (ed.), *Semiotica e intelligenza artificiale*, Lexia 48, Roma, Aracne editore, pp. 65-90.

Friedenberg J. & Silverman G. (2012), *Cognitive Science: An Introduction to the Study of Mind*. Thousand Oaks, SAGE Publications.

Gershorn D. (2017), "The data that transformed AI research – and possibly the world", *Quartz*. Consultato il giorno 5 maggio da: <https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world>.

Gigerenzer G. (2019). "How to Explain Behavior?", *Topics in Cognitive Science*, 12:4, pp. 1363-1381, <https://doi.org/10.1111/tops.12480>.

Gillespie T. (2014), "The Relevance of Algorithms" in Gillespie T., Boczkowski P., Foot K. (ed.) *Media Technologies: Essays on Communication, Materiality, and Society*, Cambridge (MA), MIT Press, pp. 167-193.

Galison P. (1994), "The Ontology of the Enemy: Nibert Wiener and the Cybernetic Vision", *Critical Inquiry*, 21, pp. 228-266.

Galison P., Daston L. (2007), *Objectivity*, New York, Zone Books.

Grba D. (2022), "Deep Else: A Critical Framework for AI Art", *Digital*, 2, pp. 1-32. <https://doi.org/10.3390/digital2010001>

Grba D. (2024), "The Mechanical Turkness: Tactical Media Art and the Critique of Corporate AI", *arXiv*, arXiv:2402.17490v1.

Hayles K. N. (2024), "Detoxifying Cybernetics: From Homeostasis to Autopoiesis and Beyond", in Yuk Hui (eds.), *Cybernetics for the 21<sup>st</sup> Century. Vol. 1: Epistemological Reconstruction*, Hong Kong, Hanart Press, pp. 85-100.

Hepp A. (2020), *Deep Mediatization*, London, Routledge.

Jasanoff S., Kim S. H. (2015), *Dreamscapes of modernity: Sociotechnical imaginaries and the fabrication of power*, Chicago, University of Chicago Press.

Knuth D. (1997), *The Art of Programming. Vol. 1 Fundamental Algorithms*, 3rd Edition, Boston, Addison-Wesley.

Lakoff G., Johnson M. (1981), *Metaphors We Live By*, Chicago, The University of Chicago Press.

Latour B. (1999), *Pandora's Hope. Essays On the Reality of Science Studies*, Cambridge & London, The MIT Press.



Leone M. (2023), "I compiti principali di una semiotica dell'intelligenza artificiale", in Leone M. e Santangelo A (ed.). *Semiotica e intelligenza artificiale*, Lexia 48, Roma, Aracne editore, pp. 29-44.

Leone M. (2024), "Semiotics of the black box: on the rhetoric of algorithmic images", *Visual Communication*, 23(3), pp. 426-45, <https://doi.org/10.1177/14703572241247120>.

Lorusso A. M. (2011), "A Semiotic Approach to the Category of Imaginary", in Leone M. (ed.) *Immaginario*, Lexia 07-08, Roma, Aracne editore, pp. 113-124.

Malavasi L. (2017), "Nella terra degli algoritmi e dei big data: incursioni artistiche nell'invisibile", *piano B. Arti e culture visive*, vol. 2 n. 2, pp. 68-86, doi: 10.6092/issn.2531-9876/7962.

Marin L. (1994), *De la représentation*, Paris, Éditions Seuil.

McCulloch W. S., Pitts W. (1943), "A logical calculus of the ideas immanent in nervous activity", *Bulletin of Mathematical Biology*, vol. 5, pp. 115-133, <https://doi.org/10.1007/BF02478259>.

Mitchell W. J. T. (2002), "Showing Seeing: A Critique of Visual Culture", *Journal of Visual Culture*, 1(2):165-181, doi:[10.1177/147041290200100202](https://doi.org/10.1177/147041290200100202).

Munn L. (2022), *Automation Is a Myth*, Stanford, Stanford University Press.

Natale S. (2022), *Macchine ingannevoli. Comunicazione, tecnologia, intelligenza artificiale*. Torino, Einaudi.

Norbert W. (1961), *Cybernetics; or, Control and Communication in the Animal and the Machine*, 2nd ed., New York, The MIT Press.

Paglen T. (2016), "Invisible Images (Your Pictures Are Looking at You)", *The New Inquiry*, 8 dicembre, 2016. Consultato il giorno 5 maggio da: <https://thenewinquiry.com/invisible-images-your-pictures-are-looking-at-you/>.

Paglen T., Downey A. (2020), "Algorithmic anxieties: Trevor Paglen in conversation with Anthony Downey", *Digi War*, 1, pp. 18-28, <https://doi.org/10.1057/s42984-020-00001-2>.

Paolucci C. (2010), *Strutturalismo e Interpretazione*, Milano, Bompiani.

Paolucci C. (2020), *Persona. Soggettività nel linguaggio e semiotica dell'enunciazione*. Milano, Bompiani.





- Paolucci C. (2021), *Cognitive semiotics. Integrating Signs, Minds, Meaning and Cognition*, New York, Springer.
- Pasquale F. (2015), *The Black Box Society. The Secret Algorithms That Control Money and Information*, Harvard University Press.
- Pasquinelli M. (2023), *The Eye of the Master: A Social History of Artificial Intelligence*, London & New York, Verso Books.
- Pasquinelli M., Joler V. (2021), "The Nooscope manifested: AI as instrument of knowledge extractivism", *AI & Society*, 36(4), pp. 1263-1280, <https://doi.org/10.1007/s00146-020-01097-6>
- Pavan M. (2022), "La creazione immaginaria. L'elemento immaginario di Cornelius Castoriadis." Rivista online *Fatamorgana*. Consultato il 27 aprile da: <https://www.fatamorganaweb.it/elemento-immaginario-di-cornelius-castoriadis/>
- Pedrazzi R. (2021), *Futuri possibili. Scenari d'arte e intelligenza artificiale*, Milano, Jaca Book.
- Patrick E. R. (2020), "Building the Black Box: Cyberneticians and Complex Systems", *Science, Technology, & Human Values*, 45(4), pp. 575-595, <https://doi.org/10.1177/0162243919881212>.
- Ponzo J. e Stano S. (2023), "Introduzione", in Ponzo J. e Stano S. (ed.) *Nuovi media. Nuovi miti*. Lexia 49, Roma, Aracne editore, pp. 11-20.
- Rancière J. (2008), *Le spectateur émancipé*, Paris, La Fabrique éditions.
- Roberge J., Seyfert R. (ed.) (2016), *Algorithmic Cultures. Essays on Meaning, Performance and New Technologies*, Abingdon & New York, Routledge.
- Ruiz C (2019), "Leading online database to remove 600,000 images after art project reveals its racist bias", *The Art Newspaper*. Consultato il 5 maggio da: <https://www.theartnewspaper.com/2019/09/23/leading-online-database-to-remove-600000-images-after-art-project-reveals-its-racist-bias>.
- Santangelo A. (2023), "Essere umani al tempo dell'intelligenza artificiale. Narrazioni a confronto.", in Leone M. e Santangelo A. (ed.), *Semiotica e intelligenza artificiale*, Lexia 48, Roma, Aracne editore, pp. 239-266.
- Stano S. (2023), "Ripensare l'ideologia: un'introduzione.", in Leone M. e Stano S. (ed.) *Ideologia*, Lexia 41-42, Roma, Aracne editore, pp. 9-18.
- Sterling T., and van den Berg S. (2023), "Ukraine war shows urgency of military AI, Palantir CEO says", *Reuters*, consultato il 25 aprile da:



<https://www.reuters.com/technology/ukraine-war-shows-urgency-military-ai-palantir-ceo-says-2023-02-15/>

Tilak S., Glassman M., Kuznetcova I., Pelfrey L. (2022), "Applications of cybernetics to psychological theory: Historical and conceptual explorations", *Theory & Psychology*, 32(2), pp. 298–325, <https://doi.org/10.1177/0959354321105380>.

Trione V. (2022), *Artivismo: arte, politica, impegno*, Torino, Einaudi.

Volli U. (2011), "L'immaginario delle origini", in Leone M. (ed.) *Immaginario*, Lexia 07-08, Roma, Aracne editore, pp. 31-62.

Von Hilgers P. (2011), "The History of the Black Box: The Clash of a Thing and its Concept", *Cultural Politics*, 7(1), pp. 41-58.

Voto C. (2022), "From archive to dataset. Visualizing the latency of big data", *Punctum. International Journal of Semiotics*, 08(01), pp. 47–62.

Wasielewski A. (2023), *Computational Formalism. Art History and Machine Learning*, Cambridge (MA), The MIT Press.

Zeilinger M. (2021), *Tactical entanglements. AI Art, Creative Agency, and the Limits of Intellectual Property*, Lüneburg, meson press.

Zylinska J. (2020), *AI Art: Machine Visions and Warped Dreams*, London, Open Humanities Press.

## Sitography

Future of Life Institute (2023) "Pause Giant AI Experiments: An Open Letter", *FLI Open Letters*, 23 marzo. Consultato il 5 maggio da: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/>.

Palantir, sito ufficiale, sezione "defense", consultato il 25 aprile da: <https://www.palantir.com/offerings/defense/>.

IDF, pagina web dedicata alla presentazione del software Gospel, consultato il 25 aprile da: <https://www.idf.il>.

Barbican, "Trevor Paglen: On 'From Apple to Anomaly'". Consultato il 5 maggio da: <https://sites.barbican.org.uk/trevorpaglen/>.

Paglen T., "Imagenet Roulette". Consultato il 5 maggio da:



Lorenzo Aimo, Luca Brunet  
*Immaginario Black Box*

<https://www.chiark.greenend.org.uk/~ijackson/2019/ImageNet-Roulette-cambridge-2017.html>.

Akten M., "Learning to See (series)". Consultato il 5 maggio da:  
<https://www.memo.tv/works/learning-to-see/>.

Barbican, "AI: More Than Human", Press Release. Consultato il 5 maggio da:  
<https://www.barbican.org.uk/sites/default/files/documents/2019-07/FINAL%20AI%20More%20than%20Human%20exhibition%20release%20280219.pdf>







A JOURNAL OF THE  
SOCIAL IMAGINARY

# *Oltre l'estetica McNuggets. Esplorazioni creative e interazioni tecno-sociali nella produzione artistica realizzata con l'IA*

Paola Panarese

paola.panarese@uniroma1.it  
CoRiS | Sapienza Università di Roma

Vittoria Azzarita

vittoria.azzarita@uniroma1.it  
CoRiS | Sapienza Università di Roma

Maddalena Carbonari

maddalena.carbonari@uniroma1.it  
CoRiS | Sapienza Università di Roma

## Abstract

*Beyond the McNuggets Aesthetics. Creative Explorations and Techno-Social Interactions in Artistic Production Made with AI*

In the contemporary cultural landscape, Artificial Intelligence (AI) emerges as a powerful accelerator of processes in a variety of fields, including the arts. So, the diffusion of AI in art-making raises crucial questions about the ability of AI to appreciate or produce art, the evolution of aesthetic diversity in the software age, and the future of artistic expression. On this basis, as part of the PRIN PNRR IMAGES project, this paper, through in-depth interviews with Italian artists who integrate AI into their creative routines, explores the dynamics of the relationship between AI and art, reflecting on changes in the processes of signification through artistic production. The results highlight that artists, aware of the transformations underway, endeavor through their works to challenge the commercial aesthetics of AI, which reproduces standardized and homogenizing social representations. Art, therefore, emerges as a space for experimentation and exploration of unconventional ideas, playing a significant role in new forms of relationships between humans and algorithmic systems.

## Keywords

Sistemi algoritmici | Arte | Interazioni tecno-sociali | Intelligenza artificiale | Pratiche creative

*Il progetto PRIN PNRR IMAGES è finanziato dall'Unione europea-Next Generation EU, Missione 4 Componente 2 CUP B53D23029990001*

N. 23 - Year XIII / July 2024 pp. 277-290  
[www.imagojournal.it](http://www.imagojournal.it)

ISSN: 2281-8138



Yet there is still one realm of human endeavor that most people believe the machines will never be able to touch. We have this extraordinary ability to imagine, to innovate and create. Our code, the creativity code, is one we have long felt that no programmer could ever crack. This is a code that we believe depends on being human.

Marcus du Sautoy (2019)



**N**el panorama culturale contemporaneo, l'Intelligenza Artificiale (IA) emerge come un potente acceleratore di processi che si verificano in un numero crescente di sfere dell'esistenza (Citron, Pasquale, 2014; O'Neil, 2017). Tra queste, uno spazio di interesse, in un contesto che sembra *certificare il trionfo degli immaginari* (Giacomantonio, 2023), è quello del connubio tra IA e arte. Un binomio che, alla luce di recenti *affordances* dell'IA generativa e persistenti *constraints* nella "creatività" dei sistemi algoritmici (Marcus, Davis, 2019), solleva domande cruciali sulla capacità dell'IA di apprezzare o produrre arte (Chatterjee, 2022), sull'evoluzione della diversità estetica nell'era dei software (Manovich, 2020) o sul futuro dell'espressione artistica (Calveri, Sacco, 2021).

Le attuali potenzialità consentono ai modelli di IA di creare nuovi prodotti e stili artistici o innovare quelli tradizionali (Arikan, Akam, 2022), con implicazioni importanti per l'arte visiva, considerata per millenni l'apice della creatività umana (Marcos, 2007; Mazzone, Elgammal, 2019): possono arricchire le possibilità espressive (Cetinic, She, 2022), per effetto delle invenzioni più iconiche dell'IA, quali *Neural Style Transfer* (NST) (Gatys et al. 2016) e *Generative Adversarial Networks* (GANs) (Elgammal et al., 2017; Goodfellow et al., 2014); possono orientare le scelte e modellare le estetiche delle immagini online (Manovich, 2020; Oh et al., 2018), sia tramite sistemi di raccomandazione che tracciano connessioni simboliche e amplificano le pratiche di consumo più diffuse, sia per mezzo di software che adeguano i contenuti a standard condivisi (Clark et. al., 2021).

Tuttavia, le forme più comuni di IA sono ancora limitate nella produzione artistica (Chatterjee, 2022; Linson, 2016) perché non possono astrarre concetti, comprendere metafore o provare sensazioni (Marcus, Davis, 2019); non possiedono autonomia di azione, nonostante alcune nozioni di agency vedano gli algoritmi come "agenti sociali", che mediano relazioni tra umano e non umano (Esposito, 2017); non generano immagini distintive, poiché di solito rappresentano "mashup" culturalmente riconoscibili e limitatamente originali (Manovich, 2019, 2022).

Ciononostante, il continuo sviluppo di macchine "esteticamente sensibili" rende tali sistemi parte integrante delle routine creative di un numero crescente di artisti, mettendo alla prova la definizione di bellezza e creatività e forse anche la comprensione della natura dell'arte (Chatterjee, 2022).

Se è vero che gli algoritmi erano impiegati nella creazione artistica già negli anni Sessanta, è vero anche che oggi l'IA ha potenzialità creative maggiori ed è incorporata in dispositivi e servizi utilizzati da miliardi di persone, configurandosi come un meccanismo di modellamento delle estetiche prevalenti, una "machine vision" che, nell'epoca della *riproducibilità algoritmica* e dell'*epidemia dell'immaginario* (Žižek, 1997), risente di un fenomeno simile a quella *produzione industriale delle differenze* che Baudrillard (1981) attribuiva alla pubblicità.

In questo macroambito tematico si colloca il progetto di ricerca PRIN PNRR IMAGES (Inclusive Machine learning system using Art and culture for tackling Gender and Ethnicity Stereotypes), promosso da Sapienza, Università di Roma, e dall'Istituto di Scienze e Tecnologie della Cognizione del CNR. Il progetto si prefigge di indagare l'inclusività e l'accessibilità dei sistemi di ML/IA utilizzati in dataset artistici e di usare le opere d'arte in percorsi formativi volti a promuovere l'inclusione e la consapevolezza delle diseguaglianze.

Come parte del lavoro di ricerca, si propone in questa sede un contributo che, a valle di una revisione sistematica della letteratura sulla relazione tra arte e sistemi algoritmici, esplora i punti di vista, le esperienze e le pratiche di artisti italiani che integrano tali sistemi nelle proprie routine creative.

Il contesto indagato è quello nazionale, piuttosto inesplorato nel momento in cui si scrive, in relazione a opportunità, condizionamenti, potenzialità e limiti, opacità e inclusività dei sistemi algoritmici nella creazione di opere d'arte, nonché nell'invenzione di nuove rappresentazioni immaginarie e finzioni culturali – immagini, segni, simboli, idee (Holmes, 2004) – da cui dipende la comprensione della società (Marzo, Mori, 2019).

Lo studio ha dunque l'ambizione di far luce sugli usi dell'IA in campo artistico con l'obiettivo specifico di approfondire la dinamica generativa e co-evolutiva del rapporto di interdipendenza tra tecnica e immaginazione, operazioni di desimbolizzazione e risimbolizzazione (Mouffe, 2008), per riflettere sulle trasformazioni in atto in quei processi di significazione (Castoriadis, 2022) che passano per la produzione artistica.



## 1. IA e arte: ambiti d'indagine e prospettive prevalenti in letteratura

Una mappatura sistematica della ricerca accademica su arte visiva e sistemi algoritmici<sup>1</sup> ha censito e analizzato 290 contributi internazionali sul tema, 40 dei quali considerati ad alta rilevanza rispetto agli obiettivi del lavoro. Tali studi indagano diversi aspetti della relazione tra arte e tecnologie algoritmiche da molteplici prospettive.

In linea con quanto rilevato da Cetinic e She (2022), prevalgono due ambiti tematici: il modo in cui l'IA viene utilizzata in grandi collezioni di opere digitalizzate e l'uso che se ne fa per generare nuove opere.

Il primo ambito include contributi che esaminano le tecniche di *computer vision* utilizzate da musei e gallerie d'arte, per sviluppare sistemi capaci di organizzare e analizzare grandi collezioni digitali di opere in modo efficiente (Balbi, Calise, 2023; Craig, 2021; Surapaneni et al., 2020). Tra questi, uno studio di Pereira e Moreschi (2021) sui modi "di vedere" delle principali IA commerciali di riconoscimento delle immagini, effettuato sulla collezione *Van Abbemuseum*, nei Paesi Bassi, rileva che i sistemi di IA tendono ad appiattire le opere d'arte, ignorandone il contesto o il significato, privilegiando alcuni valori e invisibilizzandone altri, mettendo così in discussione la loro configurazione come una magia "view from nowhere" (Haraway, 1988).

Sul fronte della creazione artistica, invece, prevale l'esplorazione dei limiti e, al contempo, degli aspetti più innovativi delle tecnologie algoritmiche, per comprendere se e in che modo tali strumenti possano eguagliare o addirittura superare *le meraviglie del nostro codice umano* (du Sautoy, 2019). A tale proposito, la maggior parte degli studi sembra caratterizzata più dal sollevare quesiti che dal fornire risposte. Tra questi, ricorrono interrogativi sulla capacità delle macchine di essere creative (Brinkmann et al., 2023; Chatterjee, 2022; McCormack, D'Inverno, 2013), sulla "natura artistica" delle opere create dall'IA e il loro posto nella storia delle arti visive (Hakopian, 2023; Hassine, Neeman, 2019), sull'impatto di tali tecnologie sul futuro dell'arte e della professione artistica (Caramiaux, Fdili Alaoui, 2022; Cetinic, She, 2022; Chatterjee, 2022; Hakopian, 2023; Hassine, Neeman, 2019), sulle percezioni dei pubblici nei confronti dell'arte generata da sistemi algoritmici (Horton et al., 2023; Lyu et al., 2022), oppure sul livello di autonomia dei computer nel prendere decisioni relative al processo creativo (Chatterjee, 2022; Wojciechowski, Korjonen-Kuusipuro, 2021).

Autori come Lyu e colleghi (2022), pur ribadendo che l'IA sia priva di una "coscienza umana" e quindi incapace di creare arte, mettono in luce come tali



<sup>1</sup>La *scoping review* della letteratura, condotta su Scopus e Web of Science, ha restituito 940 contributi. A seguito dell'eliminazione dei duplicati (265), dei "Proceedings" e dei contributi "Retracted" (22), degli articoli non in italiano, inglese, francese, spagnolo o portoghese (20), sono stati censiti e classificati, tramite una scheda di analisi, 633 record. Lo *screening* dei contributi sulla base di titolo e abstract ha portato a individuare 290 record pertinenti, ulteriormente analizzati, al fine di individuare la prospettiva disciplinare prevalente e la rilevanza ai fini della ricerca. Quelli a rilevanza alta, sottoposti a un'analisi più approfondita, sono 40.

tecnologie differiscano dagli strumenti tradizionali: l'imprevedibilità dei suoi output provoca un cambiamento nei modi in cui gli esseri umani controllano gli algoritmi, rendendoli una possibile fonte di ispirazione e favorendo i processi di collaborazione creativa tra artisti e sistemi computazionali. Rafforza tale esito uno studio di Caramiaux e Fdili Alaoui (2022) che, tramite interviste in profondità a cinque artisti di fama internazionale, mostra come l'IA venga usata sia come strumento sia come materiale di lavoro, ricercando l'imprevedibilità algoritmica anziché evitarla e apprezzando la ricchezza delle possibilità estetiche date dall'automazione. La ricerca mostra anche che gli artisti rimarkano la rilevanza dell'intervento umano necessario affinché un algoritmo di IA produca un'opera d'arte significativa, spesso oscurata da una narrazione che tende a descrivere tali opere come fatte "da" piuttosto che fatte "con" le macchine.

Sebbene da punti di vista diversi e da posizioni non sempre convergenti, gli studi raccolti nel censimento della letteratura concordano sui cambiamenti in atto in campo artistico e sulla difficoltà di fare previsioni sul prossimo futuro, considerata la rapida evoluzione dei sistemi computazionali. Alla luce di tale consapevolezza, il presente contributo si inserisce nell'attuale dibattito sull'influenza reciproca tra cultura e algoritmi. Offrendo un'istantanea del panorama artistico italiano che integra tali tecnologie nelle proprie prassi creative, intende esplorare le interazioni tecno-sociali abilitate attraverso l'arte e il modo in cui queste possono contribuire allo sviluppo della capacità immaginativa e narrativa.



## 2. Impianto e scelte di metodo

Per raggiungere gli obiettivi della ricerca, considerando gli esiti di una *scoping review* della letteratura scientifica (Peters et al., 2015), sono state realizzate 20 interviste in profondità ad artisti italiani che integrano l'IA nelle loro routine creative e artistiche, quali: Domenico Barra, Elena Bellantoni, Marco Cadioli, Mara Oscar Cassiani, Andrea Crespi, Francesco D'Abbraccio, Francesco D'Isa, Roberto Fassone, Irene Fenara, Kamilia Kard, Fabrizio Intonti, Mauro Martino, Elisa Giardina Papa, Chiara Passa, Nuvola Project, Oriana Persico, Mariagrazia Pontorno, Martin Romeo, Franz Rosati e Ultravioletto.

Gli intervistati sono stati selezionati sulla base di due criteri - l'utilizzo di sistemi algoritmici nella produzione artistica e la nazionalità italiana - e sono stati individuati a partire da un'esplorazione online e un confronto con storici dell'arte coinvolti nel progetto IMAGES.

L'intervista semi-strutturata in profondità è stata scelta come tecnica di raccolta dati perché coerente con gli obiettivi esplorativi del lavoro, oltre che utile per fornire una comprensione approfondita di atteggiamenti e comportamenti, attraverso le parole di testimoni privilegiati (Corbetta, 1999).

Le interviste, che hanno avuto una durata compresa tra 42 e 140 minuti, sono state condotte da remoto (su Google Meet e Zoom) o in presenza, sono state registrate, trascritte e analizzate.

La traccia è stata costruita – ispirandosi in parte al citato lavoro di Caramiaux e Fdili Alaoui (2022) – intorno a tre macroaree tematiche: il modo in cui l'IA è percepita e definita dagli artisti, le modalità con cui viene utilizzata nel percorso creativo e le implicazioni politiche e sociali connesse all'uso di tale tecnologia in campo artistico. A questi ambiti, ne sono stati aggiunti due: i cambiamenti in atto nel panorama artistico in relazione alla definizione di nuove estetiche e l'inclusività dei sistemi algoritmici.

Gli esiti delle interviste sono emersi da un'analisi tematica (Braun, Clarke, 2006) svolta da due diverse ricercatrici, che hanno classificato i dati con un approccio *bottom-up*, confrontando sia le dimensioni e il processo di codifica sia i percorsi interpretativi.

### 3. IA e routine creative: punti di vista e pratiche artistiche



282

In linea generale, gli artisti intervistati si caratterizzano per uno spiccato interesse verso le tecnologie digitali, che in molti casi precede l'avvento degli strumenti algoritmici. Il ricorso all'IA tende, infatti, a configurarsi come una naturale evoluzione di traiettorie espressive mosse dalla curiosità verso le opportunità offerte dalle tecnologie e, in alcuni casi, da una sensibilità nei confronti delle ricadute sociali connesse alla loro applicazione (anche) in ambito artistico.

I sistemi algoritmici sono integrati nelle routine produttive degli intervistati con diverse modalità e differenti gradi di interazione con altre forme espressive: se per alcuni rappresentano uno strumento da affiancare ad altri tool o un supporto per esplorare tecniche e linguaggi, per altri si configurano come il fulcro intorno al quale l'opera prende forma. Le tecnologie a cui gli artisti fanno ricorso spaziano da software e sistemi sviluppati da terzi (come ChatGPT, Midjourney, Stable Diffusion, ecc.) ad algoritmi creati ad hoc per la realizzazione di una specifica opera, grazie all'ausilio di sviluppatori informatici o in virtù di proprie competenze di programmazione.

Sebbene pratiche e prospettive non siano sempre coincidenti, emerge dalle interviste un unanime riconoscimento delle rilevanti potenzialità dei sistemi algoritmici in campo artistico. Si evidenzia, ad esempio, la loro capacità di stimolare non solo gli artisti, ma anche gli utenti, a esplorare modi di raccontare, linguaggi e percorsi creativi nuovi, imparando a dialogare e a relazionarsi con tali tecnologie.

A questo proposito, Mauro Martino per esempio afferma:

Le reti neurali hanno aperto un'area di democratizzazione molto forte di qualunque tipo di *expertise* [...] creando quell'*engagement* emozionale necessario all'individuo per essere creativo

Si evidenzia, inoltre, la facilitazione di alcuni processi produttivi che le tecnologie algoritmiche consentono, soprattutto in relazione alla risoluzione di problemi tecnici

e alla limitazione di compiti ripetitivi, consentendo agli artisti di dedicare maggior tempo alle fasi più creative dei loro progetti.

Viene poi sottolineato come questi strumenti siano in grado di mettere in comunicazione l'ambito artistico con quello scientifico, tramite collaborazioni fra creativi e ricercatori che alimentano un "esperimento collettivo di educazione collaborativa di questi agenti tecno-sociali" (Oriana Persico). In tale prospettiva, l'arte ha l'opportunità di recuperare un ruolo da protagonista nel discorso collettivo, contribuendo a rendere l'IA "un vantaggio per tanti" (Domenico Barra).

Il fronte delle criticità connesse all'uso dell'IA in campo artistico è ancora più compatto e fa riferimento soprattutto all'estetica dei prodotti dei sistemi algoritmici. In particolare, emerge l'idea che i modelli di IA possano alimentare un gusto uniforme, generare output caratterizzati da "un sapore standard" (Ultravioletto), riprodurre "stilemi già esistenti" (Nuvola Project) e "immagini tutte uguali che non portano a niente di nuovo" (Domenico Barra), andando ad "addormentare invece che a svegliare la creatività" (Domenico Barra) e incoraggiando "una semplificazione che riduce la capacità di immaginare" (Nuvola Project). In tal senso, viene sottolineato come uno dei principali problemi dell'utilizzo dell'IA in campo artistico derivi dal "fatto che molti utenti non hanno ancora maturato un proprio percorso creativo-artistico" (Mara Oscar Cassiani) e in mancanza di un bagaglio conoscitivo personale, il "loro immaginario rischia di diventare l'immaginario di qualcun altro" (Mara Oscar Cassiani).

Si cita anche il rischio che l'accesso di massa a questa tecnologia comporti un *inquinamento* della produzione creativa che potrebbe rendere estremamente difficile "distinguere [le opere] che hanno dietro un pensiero rispetto a una massa di contenuti simili tra loro" (Nuvola Project), permettendo a chiunque di improvvisarsi artista. Una criticità che secondo alcuni intervistati si lega alle conoscenze e competenze "tecniche" non solo degli artisti e degli utenti, ma anche delle professionalità che operano nel campo dell'arte contemporanea e a cui è demandata una funzione di "selezione", decretando "cosa meriti di essere esposto all'interno di un museo o di una galleria e cosa no" (Francesco D'Abbraccio). Secondo Francesco D'Abbraccio, ciò sembra derivare dall'esigenza di affrontare un fenomeno nuovo e dirompente "che suscita interesse e ha un impatto rilevante sulla società, senza aver avuto il tempo di coltivare le competenze intorno a questo tema".

Inoltre, si rileva una certa apprensione per la difficoltà delle IA generative di "creare qualcosa che sia brutto" (Domenico Barra), poiché le immagini prodotte tendono a replicare estetiche convenzionali tipiche del canone occidentale contemporaneo. È piuttosto diffusa tra gli intervistati, infatti, la convinzione che l'IA generi un immaginario "patinato, massificante, standard, neoliberalista, che sembra accontentare le esigenze delle agenzie pubblicitarie" (Marco Cadioli), riproponendo i canoni dell'estetica commerciale e svolgendo una funzione egemone nel "dominio del desiderio" (Ultravioletto).

Sono soprattutto le tecnologie "text-to-image" a produrre *estetiche mainstream*, realizzate a partire da "modelli pubblicitari della comunicazione visiva



contemporanea" (Fabrizio Intonti), capaci di soddisfare i gusti di audience ampie e diversificate. Così, per gli artisti di Ultravioletto,

Queste IA sono allenate per produrre delle immagini piacevoli, cioè come il gusto dei *McNuggets*, è impossibile che non ti piace, perché hanno quel sapore di glucosio che provoca piacere nel nostro cervello.

L'omologazione verso estetiche convenzionali è associata da alcuni intervistati al fatto che i programmi di IA oggi più conosciuti e usati siano sviluppati da società profit che agiscono secondo logiche e dinamiche tipiche della *platform society* (van Dijck et al., 2019). Una riflessione che chiama in causa i dati e il potere delle realtà che li controllano e gestiscono, con "una ricaduta sulle nostre vite che non possiamo al momento prevedere" (Mariagrazia Pontorno).

Anche in conseguenza di tale omologazione commercialmente orientata emerge una preoccupazione per le esclusioni delle marginalità insite in sistemi che incorporano e riproducono bias e pregiudizi (Crawford, 2021; Noble, 2018; O'Neil, 2017) difficili da contrastare, soprattutto se "non si conoscono i loro meccanismi di funzionamento" (Ultravioletto) e se non si ha la possibilità di intervenire nel processo di progettazione e sviluppo.

In ogni caso, il modo in cui tale estetica si ripercuote sul lavoro creativo varia da artista ad artista, tanto che alcuni affermano non solo di ricercare un'estetica "non umana" ma anche di essere interessati a mostrarla, a "lasciarla libera di apprendere ed esprimersi" (Oriana Persico).

Se è vero che l'estetica degli algoritmi si intreccia in vari modi con quella degli artisti, sembra altrettanto vero che questa possa influenzare i modelli condivisi della realtà immaginata, in un contesto generale in cui l'accesso a questi strumenti si sta allargando, coinvolgendo diversi gruppi di persone, "dal pensionato al professionista" (Domenico Barra). Per diversi intervistati, attraverso interazioni quotidiane e ripetute nel tempo, tali tecnologie partecipano alla costruzione sociale della realtà, anche in maniera "subliminale e invisibile" (Domenico Barra), contaminando il modo di percepire il mondo e le categorie e le forme simboliche con cui attribuiamo significato agli eventi sociali.

Tuttavia, nella prospettiva più diffusa tra gli artisti, l'arte non sembra condizionata drasticamente dalle estetiche omologate e omologanti, potenzialmente distorte e distorcenti, per diverse ragioni. La prima riguarda il fatto che l'IA è percepita come uno *strumento*, un "collaboratore disponibile e sempre reperibile" (Martin Romeo), in grado di aiutare "gli esseri umani non a essere super-umani, ma probabilmente a fare meglio quello che già facevano" (Nuvola Project). Tale strumento, dunque, non impatta sullo statuto dell'arte, né mette in discussione il ruolo dell'artista, che continua ad assolvere la fondamentale funzione di attribuzione di senso all'opera, al di là della mera produzione.

E sebbene alcuni intervistati individuino il rischio di un possibile "irrigidimento dei contesti artistici" (Franz Rosati), che dipende dalla tendenza dei sistemi algoritmici di portare l'artista a compiere scelte che non derivano dalla sua ispirazione, intesa



come “processo conviviale ed esperienziale” (Franz Rosati), quanto piuttosto dai suggerimenti della macchina che riduce l'importanza del vissuto dell'individuo e ne appiattisce creatività e immaginario espressivo, si riconosce che competenza ed estro possono trasformare i limiti dell'IA in un'occasione di sperimentazione. In tal senso,

L'originalità [...] non viene dallo strumento che ci incita a una omologazione, anche del meccanismo, perché tutti spingono gli stessi tasti, nella stessa maniera. L'originalità deve stare dietro, sul progetto, sul pensiero, sulla relazione delle persone coinvolte, perché è quella la fonte dell'unicità di un'opera (Ultravioletto).

La seconda ragione che sembra rafforzare la capacità dell'arte di offrire alla società “a chance to reflect collectively on the imaginary figures it depends upon for [...] its self-understanding” (Holmes, 2004: 549) risiede nel fatto che, quali “osservatori dei fenomeni sociali”, alcuni artisti si fanno interpreti del proprio tempo, creano contro-narrazioni e aprono a nuove rappresentazioni lontane dai processi di standardizzazione evidenziati. È questo, per esempio, il caso di Oriana Persico, che afferma:

In quale sistema socio-tecnico viviamo? In uno dove prevale l'addestramento, il *pre-training* di questi oggetti computazionali? Cosa racconta il modo in cui trattiamo gli agenti computazionali della società in cui viviamo? Un immaginario fatto di risorse estrattive dei dati e della computazione per raggiungere efficienza, performatività, perfezione

Va esattamente in questa direzione il progetto *IAQOS*, sviluppato dall'artista in collaborazione con Salvatore Iaconesi: una IA di quartiere *open source*, in cui il territorio si “prende cura dell'educazione di un agente computazionale”, sviluppando il concetto di IA di comunità.

La selezione e la gestione dei dati nei sistemi algoritmici sono al centro anche delle riflessioni di Nuvola Project che con i loro lavori intendono “aprire a sguardi, a percezioni e a modi di agire nella società diversi”.

Nello stesso solco si inserisce il lavoro di Domenico Barra, che usa “il *glitch* come un'opportunità di guardare al mondo da prospettive inedite, quelle più fragili e vulnerabili”, trasformando “l'imperfezione da limite a punto di partenza per parlare di quanto è omesso o dimenticato dalla narrazione della società”.

Un altro esempio è quello di Kamilia Kard, che con la performance partecipativa *Toxic Garden, Dance Dance Dance*, realizzata all'interno della piattaforma di gioco online Roblox, intende scardinare quei meccanismi che, negli spazi digitali guidati dagli algoritmi, portano a processi di omologazione o esclusione sociale.

Con un focus specifico su stereotipi e pregiudizi, poi, il progetto *I AM AI* di Fabrizio Intonti evidenzia i bias dei sistemi di IA generativa e ne decostruisce le rappresentazioni, mettendo in luce come questi sistemi tendano a riproporre in modo acritico “il senso comune senza scardinare le convinzioni di chi li usa”.



Similmente, Marco Cadioli, con *Subway Portraits in the Age of AI*, mostra la “semplicità e pervasività” di tecnologie algoritmiche come il *face detection*, in grado di “catturare anche le emozioni umane, fino a determinare una misura della felicità, con tutte le implicazioni che questo può comportare”.

Sebbene restituiscano una visione parziale e non esaustiva della ricca e variegata costellazione di esperienze e approcci emersi dal racconto degli artisti, gli esempi citati sono rappresentativi di quello che, secondo gli intervistati, è uno degli scopi, se non lo scopo principale, dell'arte: affrontare le tematiche della contemporaneità, del digitale, dell'impatto della tecnologia sulla società, attraverso “una lente che consente uno sguardo su questi fenomeni laterale e non condizionato dal mercato” (Marco Cadioli) e con “una coscienza diversa, libera da costrizioni, capace di mettere in luce anche i potenziali rischi” (Marco Cadioli).

#### 4. Sguardi laterali e pratiche di svelamento dell'estetica algoritmica

Le interviste in profondità ad artisti italiani che integrano l'IA nelle loro routine creative hanno permesso di raccogliere punti di vista esperti sulla relazione tra sistemi algoritmici e arte, le nozioni di estetica e autorialità, le dinamiche di interazione tra esseri umani e macchine e alcune dimensioni degli immaginari tecno-sociali, a cui la produzione artistica può apportare un contributo non marginale.

Gli esiti della ricerca restituiscono un quadro composito, articolato e in trasformazione, che tocca molti dei temi trattati dalla letteratura su arti visive e algoritmi, con particolare riferimento all'evoluzione della creatività nell'era dell'IA (Chatterjee, 2022), al concetto di “agency delle macchine” (Lyu et al., 2022), e alle ricadute di tali tecnologie sul futuro dell'arte e della professione artistica (Caramiaux; Fdili Alaoui, 2022; Cetinic, She, 2022; Chatterjee, 2022).

Diffondendosi in un ambito considerato tradizionalmente umano (Hertzmann, 2018), quale quello dell'arte e della creatività, che si muove all'interno di semiosfere culturali, circoli di esperienze e conoscenze socialmente e culturalmente situati, queste tecnologie pongono gli artisti di fronte a una serie di sfide, alcune inedite, altre non lontane da quelle lanciate in passato dalla fotografia o da altri strumenti tecnologici (D'Isa, 2024). Se da un lato l'IA amplia la gamma dei tool a disposizione, favorendo processi di collaborazione tra artisti e sistemi computazionali e tra arte e scienza, dall'altro impone l'esigenza di riaffermare la propria autorialità e la centralità di quel “lavoro umano” necessario per giungere alla creazione, ma anche e soprattutto alla significazione, di un'opera d'arte (Caramiaux, Fdili Alaoui, 2022).

In tale contesto, l'uso delle tecnologie algoritmiche sembra rafforzare le interconnessioni tra i tre sistemi principali che interagiscono nella creazione di un prodotto creativo (Csikszentmihalyi, 1999): il *dominio* o l'insieme di convenzioni, conoscenze e simboli che formano il contesto culturale di riferimento, l'*individuo* che porta novità nel dominio e il *campo* di esperti che riconosce e convalida l'originalità di un output.



Agendo all'interno di sistemi in cui ogni componente è un fattore necessario ma non sufficiente, di per sé, a produrre novità, gli artisti intervistati riconoscono la capacità dell'IA di arricchire le risorse del dominio, con una *machine vision* la cui imprevedibilità algoritmica alimenta il processo creativo e contribuisce a creare opere inattese, sebbene non sia percepita come una forza creatrice (Brinkmann et al., 2023; Chatterjee, 2022) e le sue opere siano considerate manchevoli di una natura artistica pienamente autonoma (Hakopian, 2023; Hassine, Neeman, 2019).

In uno scenario così caratterizzato, il ruolo dell'artista come l'unico in grado di attribuire significato all'opera appare agli intervistati ancor più rilevante, soprattutto in considerazione della diffusione di massa di tecnologie quali le IA generative "text-to-image", che rischia di favorire una sovrapproduzione di artefatti orientati da un'estetica commerciale, bella ed emozionale, ma stereotipata e manchevole di originalità (Manovich, 2023). Un'estetica che può plasmare percezioni, immaginari, condotte e pratiche, a vari livelli (Bucher, 2017), nonostante l'apparente neutralità e imperscrutabilità dei sistemi algoritmici (Airoldi, 2015), e che tende a consolidare immagini conformiste e omologanti, che rafforzano la "visione" dei gruppi dominanti, riflettono "ideologie e regole che vengono da un preciso luogo geografico" (Elisa Giardina Papa) e alimentano una concezione parziale del mondo (Noble, 2018).

Nell'era dei software, infatti, si assiste a una disseminazione senza precedenti della dimensione estetica e immaginativa: ogni giorno gli individui creano, condividono e interagiscono con un'enorme mole di nuovi artefatti digitali, capaci di condizionare in maniera diretta e indiretta sia le decisioni e i gusti estetici sia il modo di "classificare" e "vedere" il mondo (Manovich, 2020).

Ne deriva che, da un lato, l'IA si configura come uno strumento prezioso che consente di esplorare nuove forme espressive, ampliando i confini dell'immaginazione umana; dall'altro, appare come un pericolo per la diversità culturale, l'apertura e l'inclusione, soprattutto se usato da utenti non esperti e non bilanciato da rappresentazioni creative alternative. D'altronde, emerge piuttosto chiaramente che gli artisti si servono degli strumenti che il loro tempo offre, ma anche che possono incidere su tali strumenti, a conferma del rapporto di *mutual shaping* tra società e sistemi tecnologici (van Dijck, Poell 2013).

Ed è proprio all'interno di queste dinamiche di influenza reciproca, che l'arte può ampliare il proprio campo d'azione, intervenendo direttamente in una molteplicità di spazi sociali per dare vita a nuove finzioni culturali (Mouffe, 2008). Attraverso le loro opere, gli artisti hanno la possibilità di contestare e scuotere l'immaginario neoliberista della *platform society*, sia invitando a riflettere sui modi in cui le immagini, i simboli e le idee si riverberano sulla definizione della realtà sia ideando attivamente diverse realtà possibili tramite strategie e azioni di desimbolizzazione e risimbolizzazione (Holmes, 2004). In linea con quanto messo in luce da molti degli intervistati, un intervento artistico non dovrebbe limitarsi a ornare, abbellire o intrattenere, ma dovrebbe anche essere in grado di intervenire nella sfera pubblica, nello spazio dell'azione sociale (Scardi, 2011), facendo leva sul capitale simbolico di cui l'artista è portatore e sulla sua capacità di ridefinire il significato di determinati gesti, di intensificarne il senso e di conferire loro valore (Bargna, 2011).



In questa direzione, sono diversi gli intervistati che si dichiarano impegnati a evidenziare limiti e contraddizioni insite nella rappresentazione del mondo e nella produzione di conoscenza mediata da sistemi algoritmici (Hakopian 2023), nei molteplici ruoli di mediatori della tecnologia, innovatori culturali, custodi dell'originalità e autenticità dell'arte e autori di contro-rappresentazioni e contro-narrazioni anticonformiste.

Così, alcuni artisti si posizionano come attori che esplorano criticamente le possibilità delle tecnologie algoritmiche, ne sfidano i limiti e ne rivelano le implicazioni sociali e culturali. Soggetti che contribuiscono, con la pratica artistica, ad abilitare nuovi processi di significazione e mettere in discussione lo "sguardo binario, capitalista e orientato al valore economico", incorporato in tali sistemi (Pereira, Moreschi, 2021), fornendo un terreno fertile per la sperimentazione e l'esplorazione di idee e immaginari non convenzionali, dal sapore insolito, meno gradevole di quello dei *McNuggets*, ma capace di allertare i sensi.



## Bibliography

- Airoldi M. (2015), "You might also be interested in: recommender algorithms and social imaginary, the case of YouTube", *IM@GO*, 6: 132-150.
- Arikan E., Akam S. (2022), "Qualitative Study of Text-to-Image AI Generators and their Relationship with NFTs", *2022 IEEE International Conference on Computational Science and Computational Intelligence (CSCI)*, 240-245.
- Balbi C., Calise A. (2023), "The (theoretical) elephant in the room. Overlooked assumptions in computer vision analysis of art images", *Signata*, 14: 1-16.
- Bargna L. (2011), "Gli usi sociali e politici dell'arte contemporanea fra pratiche di partecipazione e di resistenza", *ANTROPOLOGIA*, XI, 13: 75-106.
- Baudrillard J. (1981), *Simulacres et simulation*, Paris, Éditions Galilée.
- Braun V., Clarke V. (2006), "Using thematic analysis in psychology", *Qualitative research in psychology*, 3, 2: 77-101.
- Brinkmann L., Baumann F., Bonnefon J.F., Derex M., Müller T.F., Nussberger A.M., Czaplicka A., Acerbi A., Griffiths T.L., Henrich J., Leibo J.Z., McElreath R., Oudeyer P.Y., Stray J., Rahwan I. (2023), "Machine culture", *Nature Human Behaviour*, 7: 1855-1868.
- Bucher T. (2017), "The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms, Information", *Communication & Society*, 20,1: 30-44.
- Calveri C., Sacco P.L. (2021), *La trasformazione digitale della cultura*, Milano, Editrice Bibliografica.
- Caramiaux B., Fdili Alaoui S. (2022), "Explorers of Unknown Planets': Practices and Politics of Artificial Intelligence in Visual Arts", *CSCW '22: ACM Conference On Computer-Supported Cooperative Work And Social Computing*, November 12-16, Taipei, TWN. ACM, New York, NY, USA, 1-24.
- Castoriadis C. (2022), *L'istituzione immaginaria della società*, Milano, Mimesis Edizioni.
- Cetinic E., She J. (2022), "Understanding and Creating Art with AI: Review and Outlook", *ACM Transactions on Multimedia Computing, Communications, and Applications*, 18, 2: 1-22.
- Chatterjee A. (2022), "Art in an age of artificial intelligence", *Frontiers in Psychology*, 13:1024449, 1-9.
- Citron D.K., Pasquale F.A. (2014), "The Scored Society: Due Process for Automated Predictions", *Washington Law Review*, 89, 8: 1-33.
- Clark E., August T., Serrano S., Haduong N., Gururangan S., Smith N.A. (2021), "All that's human is not gold: Evaluating human evaluation of generated text", arXiv preprint arXiv:2107.00061.





- Corbetta P. (1999), *Metodologia e tecniche della ricerca sociale*, Bologna, il Mulino.
- Craig J. (2021), "Computer Vision for Visual Arts Collections: Looking at Algorithmic Bias, Transparency, and Labor", *Art Documentation: Journal of the Art Libraries Society of North America*, 40, 2: 198-208.
- Crawford K. (2021), *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, New Haven/London, Yale University Press.
- Csikszentmihalyi M. (1999), "Implications of a systems perspective for the study of creativity", in Sternberg R. J. (Ed.), *Handbook of creativity*, Cambridge, Cambridge University Press, 313-335.
- D'Isa F. (2024), *La rivoluzione algoritmica delle immagini. Arte e intelligenza artificiale*, Roma, Luca Sossella Editore.
- du Sautoy M. (2019), *The Creativity Code: How AI Is Learning to Write, Paint and Think*, London, HarperCollins Publishers.
- Elgammal A., Liu B., Elhoseiny M., Mazzone M. (2017), "CAN: Creative adversarial networks, generating "art" by learning about styles and deviating from style norms", *Eighth International Conference on Computational Creativity (ICCC)*, Atlanta, 1-22.
- Esposito E. (2017), "Artificial communication? The production of contingency by algorithms", *Zeitschrift für Soziologie*, 46, 4: 249-265.
- Giacomantonio F. (2023), "Immaginari, cultura e istituzioni: attualità del pensiero di Cornelius Castoriadis", *Scenari*, Milano, Mimesis Edizioni.
- Gatys L. A., Ecker A. S., Bethge M. (2016), "Image style transfer using convolutional neural networks", *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2414-2423.
- Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. (2014), "Generative adversarial nets", *Advances in neural information processing systems*, 27: 2672-2680.
- Hakopian M. F. (2023), "Art histories from nowhere: on the coloniality of experiments in art and artificial intelligence", *AI & SOCIETY*, 39: 29-41.
- Haraway D. (1988), "Situated knowledges: the science question in feminism and the privilege of partial perspective", *Fem Stud*, 14, 3: 575.
- Hassine T., Neeman Z. (2019), "The zombification of Art History: How AI resurrects dead masters, and perpetuates historical biases", *Journal of Science and Technology of The Arts*, 11, 2: 28-35.
- Hertzmann A. (2018), "Can computers create art?", *Arts*, 7, 2: 1-25.
- Holmes B. (2004), "Artistic autonomy and the communication society", *Third Text*, 18, 6: 547-555.
- Horton C. B. Jr, White M. W., Iyengar S. S. (2023), "Bias against AI art can enhance perceptions of human creativity", *Scientific Reports*, 13: 19001.
- Linson A. (2016), "Machine Art or Machine Artists?: Dennett, Danto, and the Expressive Stance", in Müller V. C. (ed.), *Fundamental Issues of Artificial Intelligence (Synthese Library)*, Berlin, Springer.
- Lyu Y.R., Wang X.X., Lin R.T., Wu J. (2022), "Communication in Human-AI Co-Creation: Perceptual Analysis of Paintings Generated by Text-to-Image System", *Applied Science*, 12, 11312, 1-19.
- Manovich L. (2019), "Defining AI arts: Three proposals", *AI and dialog of cultures: exhibition catalog*, Saint-Petersburg, Hermitage Museum.
- Manovich L. (2020), *L'estetica dell'intelligenza artificiale. Modelli digitali e analitica culturale*, Roma, Luca Sossella Editore.
- Manovich L. (2022), "AI and Myths of Creativity", *Architectural Design*, 92, 3: 60-65.
- Manovich L. (2023), *Towards 'General Artistic Intelligence'?*, <https://www.artbasel.com/stories/lev-manovich>.
- Marcos A. (2007), "Digital Art: When artistic and cultural muse and computer technology merge", *IEEE Computer Graphics and Applications*.
- Marcus G., Davis E. (2019), *Rebooting AI: Building artificial intelligence we can trust*, New York, NY, Knopf Doubleday Publishing Group.
- Marzo P.L., Mori L. (2019), *Le vie sociali dell'immaginario. Per una sociologia del profondo*, Sesto San Giovanni (MI), Mimemis Edizioni.

- Mazzone M., Elgammal A. (2019), "Art, creativity, and the potential of artificial intelligence", *Arts*, 8, 26, 1-9.
- McCormack J., D'Inverno M. (2013), *Computers and creativity*, Berlin, Heidelberg, Springer.
- Mouffe C. (2008), "Art and Democracy. Art as an Agnostic Intervention in Public Space", *Open*, 14: 6-15.
- Noble S. (2018), *Algorithms of Oppression*, New York, New York University Press.
- Oh C., Song J., Choi J., Kim S., Lee S., Suh B. (2018), "I lead, you help but only with enough details: Understanding user experience of co-creation with artificial intelligence", *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1-13.
- O'Neil C. (2017), *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York, Crown.
- Pereira G., Moreschi B. (2021), "Artificial intelligence and institutional critique 2.0: unexpected ways of seeing with computer vision", *AI & SOCIETY*, 36, 1: 1201-1223.
- Peters M.D.J., Godfrey C.M., Khalil H., McInerney P., Parker D., Soares, C.B. (2015), "Guidance for conducting systematic scoping reviews", *International Journal of Evidence-Based Healthcare*, 13, 3: 141-146.
- Scardi G., a cura di, (2011), *Paesaggio con figura. Arte, sfera pubblica, trasformazione sociale*, Torino, Allemandi Editore.
- Surapaneni S., Syed S., Lee L.Y. (2020), "Exploring Themes and Bias in Art using Machine Learning Image Analysis", *2020 Systems and Information Engineering Design Symposium, SIEDS 2020*, 1-15.
- van Dijck J., Poell T. (2013), "Understanding social media logic", *Media and Communication*, 1, 1: 2-14.
- van Dijck J., Poell T., de Waal M. (2018), *The Platform Society: Public Values in a Connective World*, Oxford University Press Inc. Tr. It. *Platform society. Valori pubblici e società connessa*, Milano, Guerini, 2019.
- Wojciechowski A., Korjonen-Kuusipuro K. (2021), "Can Artificial Intelligence Become An Artist?", *Human Technology*, 17, 2: 118-125.
- Žižek S. (1997), *The Plague of Fantasies*. Tr. It. *L'epidemia dell'immaginario*, Roma, Meltemi, 2004.





A JOURNAL OF THE  
SOCIAL IMAGINARY

# AI's influence on the Creative and Cultural Industries

Ingrid Kofler

ingrid.kofler2@unibz.it

*Faculty of Design and Art | Free University of Bozen-Bolzano*

Mustapha El Moussaoui

Mustapha.ElMoussaoui@unibz.it

*Faculty of Design and Art | Free University of Bozen-Bolzano*

Romuald Jamet

Romuald.Jamet@inrs.ca

*Institut national de la recherche scientifique | Centre Urbanisation-Culture-Société,  
Quebec, Canada*



## Abstract

AI has significantly advanced, impacting various fields, particularly the Creative and Cultural Industries (CCI), prompting a reevaluation of traditional approaches and professional roles. This paper critically examines the evolving role of AI in the cultural and creative industries (CCI) by analyzing its impact on both professionals and society, using architecture and music as exemplary cases. The study challenges the traditional "myth" of creativity, historically viewed as an exclusively human ability, now being reshaped by AI technologies. By adopting a transdisciplinary approach, the paper highlights the need for a broader understanding of how AI influences the cultural sector, emphasizing the importance of considering both the economic and socio-technical imaginaries in this transformation.

## Keywords

Artificial Intelligence | Social Imaginary | CCI | Architecture | Music





Artificial Intelligence (AI) has made substantial advances in several fields, including the CCI (Creative and Cultural Industries) disrupting old approaches and employment responsibilities. In the creative economy, technology, particularly digital technology, has always had a significant role. Already the United Nations' Creative Economy Report of 2010 underscores technology as a major driver of the creative economy (United Nations, 2010). The concept of cultural and creative industries is contextual and varies geographically, influenced by local heritage and circumstances. It encompasses a fusion of "creative arts" and "cultural industries," evolving with advancements in technology and the emergence of new media technologies. This evolution has expanded the scope of creative industries to include a range of knowledge-based economic activities, necessitating innovative, multidisciplinary interactions (Hartley, 2005). Since the 1980s, the creative industries have emerged as an important dynamic sector within the global economy (Hartley, 2005; Collins, 2017). This trend is evident not only in Western Europe and the US, but also in Asia, particularly in China and South Korea (Collins, 2017). It seems, as culture and creativity have become key drivers of economic growth, as the European Agenda for Culture and the "Creative Europe" program recognize the importance of culturally driven development for sustainable and inclusive growth.

So far, creativity has increasingly been recognized as a crucial component of economic development, as well as a fundamental skill. This notion gained significant traction, especially following Richard Florida's book "The Rise of the Creative Class," (2002) which popularized the idea of a strong link between creativity and economic growth within the framework of capitalist logic. Consequently, the imaginary surrounding the cultural and creative sectors has become closely intertwined with the concepts of economic productivity and creativity. The evolution of imaginaries within cultural and creative industries has been a subject of extensive scholar inquiry, highlighting shifts in economic paradigms, technological advancements, and societal values. Campbell (2014) explores the emergence and dominance of the economic imaginary surrounding creativity, tracing its trajectory from a discourse centered on cultural industries to one emphasizing the creative economy.

This transition reflects broader policy shifts aiming to integrate culture into economic agendas, blurring the boundaries between cultural and creative domains.

The discourse surrounding creative industries is supported by statistical evidence, contributing to the stability of the economic imaginary. As the creative industries agenda evolved, it increasingly prioritized business development over cultural and social inclusion goals, aligning with neoliberal economic strategies, raising concerns about the commodification of culture (O'Connor, 2015).

This commodification brings to what already has been observed by Adorno (1964) and the Frankfurt School: the critique of the cultural industry (*Kulturindustrie*) underscores its capacity to produce a normative and heteronomous mass consciousness and perpetuate ideological conformity. By prioritizing commercial interests over intrinsic content, the industry homogenizes cultural products, stifling diversity and creativity. The cultural industry, through marketing strategies and

standardized distribution channels, reinforces social norms and restricts critical engagement, ultimately shaping individuals' attitudes, behaviors and imaginaries to align with preconceived ideologies. In this context, algorithmic technologies have gained new significance, not only in shaping individual subjectivities and their imaginaries through the hyperindividualization of algorithms (Roberge & Seyferth, 2016), but also in influencing the work of professionals within the creative sector.

Looking ahead, the integration of AI into cultural and creative industries is poised to further reshape social imaginaries. AI algorithms have the potential to optimize content creation, distribution, and consumption processes, leading to new modes of cultural production and consumption. However, concerns about algorithmic bias, filter bubbles, and the commodification of culture underscore the need for critical examination of AI's influence on social imaginaries. Social imaginaries are socially transmitted representational assemblages that interact with individuals' imaginaries and serve as mechanisms for meaning-making and world-shaping. The concept of the imaginary encompasses both the process of generating meanings and the resulting outcomes but also forging normative constitutions (Castoriadis, 1987).

Drawing on Manfred Steger's (2008) conceptualization, the social imaginary encompasses various layers of meaning, including ideas (expressed beliefs of individuals), ideologies (comprehensive belief systems representing patterned truths), and imaginaries (collective constructs framing diverse ideological manifestations). As such, the social imaginary is shaped not only by conceptual ideas but also by practices that reflect the meanings and traditions of the past (James, 2019). Thus, the social imaginary of the creative and cultural industries can be defined as the collective perceptions and conceptions that people hold regarding the sector, encompassing both conceptual ideas and practical engagements, both for professionals as also for individuals.

The investigation into the intricate relationship between AI integration and cultural production within the creative and cultural industries is essential for understanding their evolving impact on societal values and the formation of social imaginaries. As previously discussed, much of the existing literature on CCI focuses on the economic imaginary related to the cultural and creative sector. However, it often lacks a critical analysis of the developments over the past years and does not adequately consider how rapidly AI is transforming the sector. This transformation challenges the dominant imaginary surrounding the so-called 'myth' of creativity and the assumptions traditionally associated with the cultural and creative industries (Manovich 2022; Mazzone & Elgammal, 2019; Natale & Ballatore, 2020). Therefore, to effectively address imaginaries within the cultural and creative industries (CCI), it is essential to adopt a transdisciplinary perspective that goes beyond merely the economic or socio-technical imaginaries. In this paper we want to take two examples of the creative sector, taking them as "exemplary cases" (Kracauer, 1971) of how AI is impacting the creative sector, looking at two different perspectives: at one hand on the work of professionals and at the other the impact on individuals and society.





By delving into specific examples within these fields, we aim to elucidate how AI-driven technologies are transforming cultural production, consumption patterns, and the construction of meaning across diverse cultural landscapes in a transdisciplinary perspective. The paper is the result of the discussion between professionals working in music, architecture, design, research and social sciences permits therefore a unique and transdisciplinary understanding of the different impacts that AI has on CCI. The main contribution of this paper is lies therefore in its critical examination of the evolving role of AI within the cultural and creative industries (CCI) at the example of architecture and music challenging established (economic) imaginaries and assumptions surrounding creativity and aesthetics. By highlighting the rapid advancements in AI, the paper questions the traditional "myth" of creativity (Manovich, 2022) that has long been associated as a human ability, but is now changing. Moreover, using these two examples we show how technologies in the sector are nothing new, but what is new is its impact and possible future developments. This is important, because it emphasizes the need for a transdisciplinary approach to understanding these shifts, including both the effects at the side of the professionals, but also what happens to the individuals and the society. This perspective encourages a more comprehensive analysis of how AI is reshaping the cultural and creative sectors, offering new insights that could influence future research, policy, and practice within these industries. The social imaginaries surrounding AI have the potential to instill fear across society, including among creative professionals, while simultaneously encouraging a passive, standardized form of cultural consumption. It can be viewed as a modern iteration of Schumpeter's concept of creative destruction, underscoring the increasing subordination of cultural industries to data industries.

First, we will introduce the example of architecture: starting from the history of technology in architecture and how it has evolved over time, what are the main concerns and then showing that the difficulty is not how AI can become a "partner" in doing the creative work, but how professionals will use technology. This part of the paper examines the enduring role of architects in the age of AI, emphasizing that while AI advancements have transformed many aspects of design, the core elements requiring an architect's specialized understanding of aesthetics, space and societal needs remain crucial. Second, with the example of music we show how AI produce a hegemony of the sector and how algorithms are shaping and directly influencing taste, consumption and production of music.

## 1. Technologies in Architecture

The history of architecture is intricately linked to technological advancements, ranging from the creation of the compass and T-square to the evolution of CAD software and, more recently, the emergence of AI applications. Architects have consistently encountered the need to adapt and acquire fresh proficiencies with the

advent of each novel technology. Moreover, they have frequently confronted opposition and doubt from their peers within the architectural community.

The introduction of CAD tools in the 1980s brought about a significant transformation in the area of architecture (Gero, 1983). Architects transitioned from utilizing manual drawings and tangible models to employing digital design methodologies (Coyne, 1996). There was opposition to the implementation of CAD tools because people were worried about losing their jobs and the possibility of human creativity and talent becoming less valuable (Cooley, 1997). Research has demonstrated that utilising CAD tools can greatly enhance productivity and job satisfaction (Barfield, Shieldest, & Cooper, 1993). Computer-aided design (CAD) tool's influence on the process of creative problem solving in architectural design has been determined to have both advantageous and disadvantageous effects. On one hand, CAD tools can improve visualisation and speed up the process of drafting, while on the other, they carry the risk of limiting thinking and causing premature fixation on a particular design (Robertson and Radcliffe, 2009).

The advent of Computer-Aided Design (CAD) technologies into architectural domain brought about a substantial departure from traditional methodologies. This innovation greatly expanded the possibilities for architects to present innovative design concepts and enhance their efficiency. Nevertheless, this alteration also led to intricacies, particularly in the understanding and administration of spatial components. The conventional method of drafting, which depended on predetermined scales, provided a concrete understanding of size and measurements. However, the conceptualization of this element grew significantly with the development of CAD tools and their digital interface. The shift towards a more flexible and dynamic conception of space signifies a notable change in the thinking and methodology of architects in relation to the design process (Robertson and Radcliffe, 2009). The historical use of CAD required a reassessment of the comprehension of space, shifting from the limitations of tangible mediums to the boundless opportunities provided by digital realms. This move not only signified a substantial shift but also sparked an ongoing progression in architectural design, wherein the comprehension and depiction of space are constantly improving, inspired by continuous improvements in digital technology (Bhavnani, 1996).

The essential issue is not whether this represents a change or a continuance, but rather how architects can navigate and utilise these emerging models to leverage innovation while also maintaining a comprehensive grasp of spatial dynamics. The insights acquired from this period of technological implementation can provide valuable comprehension for the current integration of AI into the field of architecture.

As AI is advancing and becoming more widely used in the construction and design sector, there is also a growing concern in the industry about the future responsibilities of architects, or at least as architects know it. The core concern of this topic revolves around the potential for AI to supplant human architects or lessen the importance of their specific skill. The talents stated above include not only technical abilities such as modelling, drafting, and computational design, but also aesthetic



sensitivities such as the capacity to conceive and create visually pleasing and spatially balanced designs. The primary responsibility of an architect is to develop designs that make a deep connection with persons and effortlessly integrate form, function, and the surrounding environment. This requires a strong foundation in aesthetic skills. The concern is legitimate, as AI applications in architecture, such as generative design algorithms, have shown the potential to generate design solutions based on certain criteria and objectives. Although AI can enhance designs to make them more efficient, sustainable, or affordable, the intricate comprehension of creative worth, cultural importance, and emotional influence remains exclusively within the realm of human expertise (Zhou & Lee, 2024) (Mazzone & Elgammal, 2019).

Hence, the difficulty does not lay in substituting architects with AI, but rather in how architects may utilise AI as a tool to enhance their abilities, including aesthetic discernment. This is a mutually beneficial partnership in which AI manages the more numerical and data-oriented components of design, enabling architects to concentrate on the qualitative and imaginative aspects (Leach, 2022).

One fundamental understanding of this inquiry revolves around the notion of representation, an area where currently AI is its main play field in this domain (fig. 1).

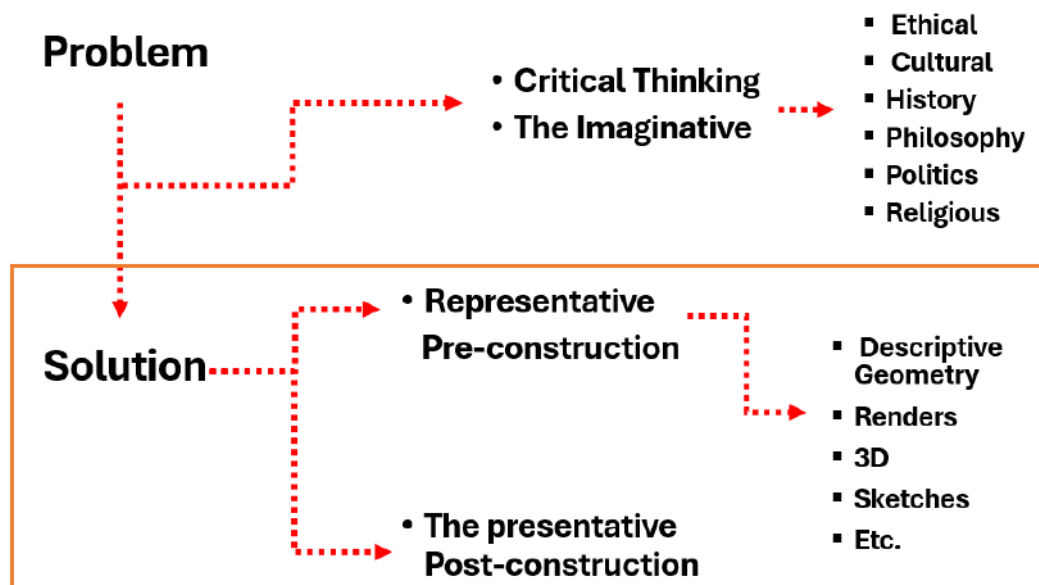


FIG. 1 – Simple problem-solving phases. Source: El Moussaoui (2024)

Architecture, at its core, is the pursuit of creative solutions to complex problems. At each stage of the design process, the architect must traverse diverse fields of knowledge to arrive at an outcome. This outcome, however, must be made manifest—it must be represented to exist beyond mere thought. Historically, representation has evolved through various mediums: from carvings on tablets and rocks, to sketches on paper, and later to digital models using CAD tools. These were

not the ideas themselves but vehicles to convey thought, whether fully formed or still emerging.

Representation methods and tools have always evolved with the development of civilizations, reflecting the advancements in communication. One of the key milestones in this evolution was the creation of descriptive geometry in the late 18th century by Gaspard Monge (Monge, 1795). His system allowed engineers and architects to project three-dimensional objects onto two-dimensional planes using mathematical precision, offering a shared "language" for representing complex structures (Barbin, 2019).

The next significant leap in representation came with the development of computer-aided design (CAD) tools in the latter half of the 20th century, as mentioned earlier. AutoCAD, introduced by Autodesk in 1982, became the predominant tool in architectural and engineering practices. It transformed the representation of designs from hand-drawn blueprints into digital models, significantly improving the accuracy, scalability, and flexibility of the design process.

AutoCAD introduced a new "language" of design, where architects could manipulate objects in a virtual space and immediately see the consequences of design decisions, thus streamlining the workflow and enhancing collaboration across disciplines. Studies have shown that AutoCAD increased productivity by up to 60% in certain industries by reducing manual errors and shortening design cycles

Today, we stand in the middle of major keystone at this point history, AI is playing a central role in the representational phase. The transformative impact of AI in architecture, particularly in representation, stems from the development of technologies such as Generative Adversarial Networks (GANs) in 2014 and the subsequent rise of Diffusion Models by 2022 (Kumar et al., 2024). These innovations have revolutionized how we visualize and externalize architectural thought, reshaping the very tools we use to express creative solutions.

Nevertheless, the advancement of representational techniques has greatly enhanced architecture and several associated fields. The incorporation of spatial changes into architectural techniques emphasises the continuous influence of computer technology in contemporary society. This highlights the crucial importance of projection in the discipline, a sentiment that is held by both theorists and practitioners.

The significance of this is further emphasised by the swift progress in projective geometric approaches propelled by AI. Projects such as NeRF showcase this progress by showcasing the capability to generate complete 3D models from a limited number of 2D images, hence uncovering previously undiscovered geometries. This transition from incomplete to comprehensive depictions represents a significant advancement in the objectives of descriptive geometry. Various AI models utilising distinct spatial representations are consistently advancing, leading to significant advancements in projective approaches that are applicable to architecture practice (Stanislas, 2023).



### **1.1 Magnitude of Change**

DamiLee, a social media blogger, undertook a straightforward experiment to demonstrate the alterations occurring in the field of architecture. Three human architects were compared to AI technologies in a competition to design a 280-square-meter home on a difficult sloped terrain. As the ideas were shown, each representing different viewpoints and nuances, the Archibeans Discord community was invited to evaluate them without knowing who created each design. The AI tool's concepts emerged as the winners in the voting process, surpassing the speed at which human equivalents could develop designs (DamiLee, 2023).

The AI's proficiency extended beyond just design, excelling in rendering and textual articulation tasks, underscoring its growing significance in architectural practices. However, it is clear that the spatial characteristics of human architects' designs are notably inferior to those produced by AI. This difference was especially evident when comparing the plans generated by AI with the spatial quality and intellectual depth found in architect-crafted designs.

Nevertheless, while examining the historical development of text-to-image models, the accuracy of these algorithms that rely on visual data has seen substantial improvement, enabling them to capture increasingly intricate and organised concepts. Advanced models, such as diffusers and LLMs, encode complex conceptualizations that go beyond the fundamental components observed in parametric modelling. State-of-the-art text-to-image models demonstrate this progress by transforming complex language prompts into top-notch images through the utilisation of references to styles, historical periods, and artistic movements. Furthermore, the pace of progress in these areas of research is amazing. When we enter identical instructions into these models, it becomes clear how rapidly these technologies have begun to imitate and improve upon key notions in our profession. A comparison between the GLIDE model (December 2021) and Midjourney V5 (March 2023) and Midjourney V6 (February 2024) demonstrates a notable improvement in the accuracy of outputs, such as floor layouts, in just eighteen months (Stanislas, 2023).





**FIG. 2** – *The magnitude of GANs capabilities in creating architectural plan, in a chronological order* Source: Stanislas (2023) and El Moussaoui (2024)

In the field of architecture, aesthetics and the primacy of the visual sense have been essential. The dominance of visual perception can be attributed to the Renaissance (Bowering, 2007). Globalisation has influenced architecture to prioritise vision above other senses, reflecting an ocularcentric approach. According to Pallasma (2005), architecture is increasingly moving towards being a visual art, focused on creating an impressive display rather than providing a physical and immersive encounter. Structures are frequently created with a primary focus on their visual aesthetics, while neglecting the importance of engaging all the senses and considering the overall experience. The inclination towards retinal architecture has wider societal implications, as the emphasis on form outweighs function significantly (El Moussaoui, 2020).

Given the discussions presented earlier, it is crucial to recognise the significant influence of AI applications, especially in the representational phases of architectural design. It is important to note that we are still in the early stages of comprehending the full potential of these tools. AI tools now handle these typically time-consuming processes. These tools have shown impressive ability to speed up the process of creating visual pieces, surpassing human efficiency even when assisted by non-AI technologies like Adobe Photoshop and rendering software. This move represents a crucial change in the workflow of architectural design, as jobs that are considered ordinary and onerous are assigned to AI, which will free up some time for architects, for an initial thought, this freed up time should be extra time to be spent on the critical thinking phase, the phase were designers/architects use their background knowledge to find a solution to the problem they are handling, as by now, the representational field of architecture can be done by AI much faster and easier.



With these progressions, the fundamental elements of architectural design that require an architect's distinct understanding have mostly remained unaffected by AI (up until this point), by fundamental elements of architectural design, i mean those phases in architectural design process that allows us to jump from the problem to the solution.

These components encompass the comprehension of space and the incorporation of varied scientific information to address societal demands. The renewed focus on the architect's function highlights the inherent significance of their contribution to the discipline. Architects are not only responsible for creating visual objects, but also for designing experiences that are closely connected to the philosophical, social, and political aspects of their work. Their specialised knowledge and skills are essential in effectively utilising AI systems, not just for automation, but as tools to solve problems, create meaningful solutions, and improve sustainable results. This viewpoint emphasises the significance of emphasising the experience elements of living, surpassing the constraints of a aesthetic approach to building.



## **2. AI in Music : the (re)emergence of *Kulturindustrie***

Music is a social and professional practice that requires no special academic or scientific knowledge but only specific skills (Becker & Faulkner, 2009). First and foremost, music and the music professions (musicians, sound engineers, tour managers, artistic directors, etc.) place originality, subjective and artistic expression above mere technicality (Perrenoud, 2007). Nevertheless, music, in all its aspects of production, dissemination and reception, has always been one of the liberal arts most affected by and sensitive to technological developments (Wikström, 2020).

Indeed, while music has always integrated the most recent technologies before the other arts, at the same time, these same technologies have often destabilized this ecosystem, which is particularly integrated into the market economy (Wikström, 2020). In this respect, a certain historicity exists in the music world regarding these "micro-revolutions". In this section, we'll show how the social imaginaries associated with AI can frighten the whole of society (including music professionals) and, at the same time, correspond to a form of passive, standardized consumption of cultural production. In this respect, AI - especially generative AI - can appear as a new iteration of the Schumpeterian concept of creative destruction and, in the case of AI, of a subservience of the cultural industries to the data industries.

### ***2.1 From "elevator music" to "fake Drake": acceptance and resistance in music production***

Western music is the result of the circulation of rationalities in the Middle Ages, which led to the codification of a social and scientific rationalization of both musical production and the instruments required for its production and dissemination (Weber, 1998, 1958). Thus, at least since Mozart and culminating with Wagner, the

European musical system has been entirely codified in both form and content (Elias, 1991). Aesthetic revolutions in music since the 18th century have been more the result of scientific treatises on music (Adorno, 1994; Weber, 1998, 1958), the social expression of invisibilized music (such as blues or folk music), or technological developments (electrification, digitization of musical instruments and recording media). However, the stability - and hegemony - of the European musical system has led to an industrialization of musical production, notably through what has come to be known as Muzak music, a first contemporary iteration of the notion of cultural industry developed by Adorno and Horkheimer (1947).

Since the 1920s, Muzak music has been based on the industrialization of music, both in terms of production and aesthetics. Described as "elevator music", Muzak music had a single function from the outset: to accommodate consumers with a repertoire based on the same aesthetic organizing principle i.e. rhythmic, harmonic and melodic simplification. As the means of mass broadcasting (radio, telephone cable) began to develop, the music industry took advantage of technological advances. While the musical structures and chord progressions needed to create "hits" were already known since Bach and Mozart, Muzak Inc increased the rationalization of the process: recording the same patterns on an (industrial) assembly line, with minimal variations, to produce a sound environment conducive to consumption, concentration or withdrawal (particularly in elevators) (Lanza, 2004).

Although fiercely criticized by critics and the music world alike, this operationalization of ancestral musical knowledge has nonetheless enabled careers to blossom.

Musicians playing for radio stations or recordings labels earned their living playing music for the first time, by ceding their performance rights. Playing this uninteresting and tasteless music allowed "non-classical" musicians to earn a living and free up their time to create and compose original works.

One of the first paradoxes of the cultural industries appeared in our view: the production of music with no artistic vocation potentially enabled the performers of this music to create radically new aesthetic and artistic content. Jazz, for example, could be considered as such: the "hits" mass-produced in New York - in particular by European Jews who had fled the pogroms and were paid by written title in a fordist perspective - could be appropriated and poached by African Americans in the southern United States - mainly in New Orleans. In this way, many musicians were able to "make a living" from their music, year in, year out: by ceding their (performance) rights, musicians gained aesthetic freedom.

Generative AI brings exactly the opposite to music production: if music's rules are well known since decades, the originality of composed works can be now calculated, replicated or modified. The result is a calculated (re)compositions with the same iteration with minimal variation. based on the idea that Western music is a closed system (limited number of notes, rhythms and melodies), generative AI could produce more (variations) of the same (Snickars, 2017). In this way, many composers demand a royalty when their compositions are included in databases used by Ais



(Robins-Early, 2024), in line with the lawsuit brought by the New York Times against OpenAI (chatgpt) for the use of their article (Le Monde, 27/12/2023).

In order to protect their works or artistic identities, many musicians and copyright holders struggle to protect - or sell - their catalogs (Jamet et Blum, 2022). We've seen (2023) Drake's representatives put streaming platforms on notice for allowing the broadcast of AI-generated music featuring the artist's voice, flow, intonations, lyrics, etc. (Daniel Tencer, *MBW*, 18/4/2023) while renowned artists such as Bruce Springsteen and Neil Young have sold their catalogs for hundreds of millions dollars to art/data-brokers such as Hypgnosis, who market these catalogs on the stock market (Jamet et Blum, 2022).

From a critical and adornian perspective, these elements are the both face of the same thing: an illustration of algorithmic capitalism (Durand-Folco et Martineau, 2023), these artists (especially the more established ones) derive maximum profit based on venture capital investment (data-broken) not necessarily linked to actual consumption, and on the other hand, demand a high level of rent for streaming consumption of their works (copyright, neighbors rights). So, for all their criticisms, they are the objective allies of the most avant-garde algorithmic capitalism. Which tends to disconnect consumption and production (of their products) to focus on stock market valuation through high-frequency trading (Pasquale, 2016) of their catalogs (which will be more or less relevant depending on the request of database needed by a music-generating AI). This positioning prevents emerging artists from gaining access to the market and listeners, because the "market" (i.e. streaming platforms) claims to have absorbed and to be the only representative of all "calculated" individual subjectivities and tastes (Introna, 2016). The impact of recommendation algorithms on the social imaginary of cultural and creative industries is examined by Airoidi (2015). These algorithms, operating on platforms like Netflix, Last.fm, and YouTube, exert considerable influence over users' browsing and consumption choices, shaping contemporary cultural consumption patterns.

The automated associations generated by recommendation algorithms contribute to the formation of shared semantic connections, but also lead to homophily, conservatism, and a reduced propensity for serendipity in cultural consumption.

## **2.2 AI, Music and broadcasting : the fusion of content and subjectivity**

Algorithmic prescriptions and AI-generative's contents precisely fulfill the platforms' aim to tune in (Schütz, 2007) and harmonize with these social activities.

The promise of music streaming platforms lies in the marketing argument that each user, in his or her singularity and habits, will have an interface and musical environment "tailor-made" for them. However, this environment is never truly "tailor-made". Indeed, given the sheer volume of content and users, platforms and their analytics departments (The Echo Nest at Spotify (Lunden, 2014), for example ultimately proceed on the basis of a set of "profile-types" by aggregating "comparable" user profiles (Roberge, Jamet, Rousseau, 2019). Ultimately, then, streaming platforms rely on categories of social tastes and uses of music that pre-



exist the platforms which are aggregated in music's database. Platforms have thus engaged in fierce competition to offer algorithmic devices that can recommend, anticipate and predict what listeners might like to hear, when and how, without the need for any interaction with the platform (Jamet, 2019). It is in this sense that we can speak of "algorithmic culture, : "Algorithms have developed and implanted their logic into the very fabric of all social processes, interactions and experiences that increasingly rely on computation to unfold: they now populate our daily lives, from sorting information (...) to predicting personal preferences and desires (...) to the point where no aspect of human experience escapes them" (Roberge and Seyfert, 2016: 1)

This is, in turn, directly correlated with musical experiences aimed at ensuring a certain emotional regulation of individuals. De Nora (2000) thus sees the desire of individuals to "sanctuarize" their intimacy and thereby create a musical halo protecting them "from the outside world" and its multiple solicitations. From a radically different perspective, Beer & Burrow (2013) emphasize the interactions between listening type and geospatial context, showing that the musical choices made by individuals are closely related to the listening environment and context.

In this way, music, via algorithmic recommendations, AI content and playlists (Hagen 2015; Hagen & Lüders, 2016), is used, among other things, to make emotional adjustments and achieve a socialized individuation of music. The aim is, on the one hand, to make available the same "singularized-normalized" prescriptive techniques and to classify users into socio-cultural profiles and types of use; and on the other, to provide an individualizing experience of music. This twofold objective aims to evade, from the user's ears and consciousness, the social roots of this type of recommendation, such as the definition of listening contexts, recommendations based on listener profile homology or socio-cultural profiling. Streaming and the technological devices that accompany it (from servers to headphones, playlists, etc.) are primarily designed to continuously broadcast musical "atmospheres" that can be interwoven with other places and activities. In the end, what is proposed is a new experience of the identical (Adorno, idem): without even prejudging the intrinsic quality of the musical content, the aim remains to offer a musical experience that doesn't stand out from the situation but sticks to its "design".

Deezer, for example, has for some years been developing an artificial intelligence (Flow) to detect the moods of its user-listeners based on their interactions on social networks, their geolocation data, the movements of their cell phones or even their listening times. The purpose of this artificial intelligence would be to predict what user-listeners would be looking for to accompany their personal moods.

Playlists for mood and emotional accompaniment fit in perfectly with the desire of the various generations of the music industry, from their very beginnings, to deliver musical content to every corner of our daily lives (elevators, supermarkets, waiting rooms, etc.). Streaming platforms are part and parcel of this history, but the technological devices they use are now primarily designed to address each individual in isolation, to "embuller" (Pariser, 2011) users not in their listening experience, but in their experiential flows (Schütz, 2007). Playlists typically aimed at public places



(bars, supermarkets, administrations, etc.) or activities (transport, sport, etc.), the target of streaming platforms becoming the permanent emotional accompaniment of their users, so to speak, "streaming their moods" (Jamet, 2019). Ultimately, however, the aim of these platforms remains to develop their business models by collecting, analyzing and reselling their users' data (Roberge and Seyfert, 2016; Introna, 2016).

All these elements thus directly affect the very experience of music as a social, subjective and aesthetic experience. Musical experiences are socially and subjectively grounded (Adorno, 1994; Bourdieu 1979), the result of social, political and cultural constructions of the "earing" (Pecqueux and Roueff, 2009) and differentiated ways of experiencing music (Hennion, 1981). Moreover, as already emphasized, these musical experiences only acquire individual and societal significance in a particular environment (Dewey, 2005) and according to particular listening contexts (De Nora, 2000, Beer & Burrow, 2013). What, then, do streaming and related technological devices imply for music consumption?

First of all, we can argue that the musical experience is significantly altered by the material conditions of its realization, such as cloud storage, connectivity and ubiquity, as it is above all individual portable devices, notably smartphones and other devices with screens and individualized interfaces, that are most used to listen to streaming music (IFPI, 2024). This material reality contributes to an apparently more individualized listening experience. But critical analysis forces us to see this promise differently: the promise of hyperindividualization is above all a social imaginary driven by platforms and the capitalism that underpins them. The cultural industry appears here as an attempt to merge content and container, to merge AI-generated music content, recommendations based on socio-emotional profiles and experiential flow.

### **3. Discussion: CCI as second nature: congruence and insubordination through creative experience**

The deployment of algorithmic technologies - of which machine learning and artificial intelligence are corollaries - is thus not only intended to offer a better "user experience", but also to become part of the experiential and - musical - intimacy of their user. Streaming platforms in a way aim to standardize musical tastes, although they may participate in doing so. On the contrary, they are the entrepreneurs, willingly or not, of a mass individualization and singularization of culture.

Notwithstanding the platforms' desire to approach their users' tastes and intimacy asymptotically via the analyses of their traces, we can observe a certain insubordination on the part of users to this technological pervasion of their intimacy. For many people, music remains a particularly personal element, and the very idea that algorithms could "calculate" the depths of their subjectivity remains a non-negligible obstacle. Thus, in the face of the experiential marketing of streaming platforms, we need to intensify future research into the avoidance tactics of user-



listeners (De Certeau, 1990) in the face of these commercial and pervasive strategies, with the aim of reappropriating their own subjectivities and ontological security (Giddens, 1984).

The deployment of algorithmic technologies, including machine learning and artificial intelligence, extends beyond enhancing user experience; it intrudes into the experiential and intimate realms of users, particularly in music. Streaming platforms aim to tailor musical tastes, but in doing so, they also contribute to the individualization and singularization of culture. However, despite these platforms' efforts to align with users' preferences through data analysis, there remains a resistance among users to this technological intrusion into their personal musical spaces. For many, music is deeply personal, and the notion of algorithms "calculating" the depths of their subjectivity is met with unease. Future research must explore how users resist and reappropriate their subjectivities against the pervasive strategies of experiential marketing (De Certeau, 1990), thereby securing their ontological security (Giddens, 1984).



#### 4. Discussion

The research question central to this paper examines AI's transformative impact on social imaginaries within creative and cultural industries (CCI), particularly in music and architecture. AI's increasing integration into these sectors reshapes creative, distributive, and consumptive processes, presenting professionals with both opportunities and significant challenges. However, the understanding of these changes requires more than just a focus on the technological aspects. AI profoundly impacts CCI by transforming both the visible and invisible aspects of society (Marzo & Mori, 2019).

It reshapes the social imaginary by altering how we interact with and understand the world through new narratives and meanings, driven by AI's integration into daily life.

AI influences tangible structures like smart cities and digital platforms, while also affecting the deeper, often invisible aspects of societal values and beliefs. This transformation underscores the need for a holistic understanding of AI's role in shaping cultural production and creative practices reflecting the interplay between visible changes and underlying social meanings. In the context of technological advancements, these imaginaries influence the design and acceptance of technological artifacts. They drive choices toward specific forms over others and shape processes of meaning attribution and recognition, but also of future projections as for socio-technical imaginaries (Napolitano, 2023). This dual relationship between technology and the imaginary is crucial: while the imaginary influences the form and meaning attributed to technology, technology itself participates in shaping the imaginary (Musso, 2019).

Understanding this relationship is essential in addressing how AI and other technologies are integrated into the CCI.

The myth surrounding AI's creative potential often overlooks the nuanced understanding that AI does not create in the same way humans do. Instead, AI reconfigures existing elements, perpetuating a cycle of standardized production rather than true innovation. This myth is sustained by recurring analogies and discursive shifts, which borrow concepts from other fields to describe AI technologies. Moreover, there is a rhetorical use of the future, suggesting that current limitations will soon be overcome. Controversies surrounding AI's capabilities are integral to the discourse of the AI myth and must be critically examined.

Technological artifacts are not neutral; they are social constructions imbued with specific values and worldviews. One of the critical tasks of research is to deconstruct what appears to be self-evident, revealing the underlying social, cultural, and political dimensions. In this sense, socio-technical imaginaries, which are the inextricable relations between technoscientific assemblages and sociopolitical order, are crucial. These imaginaries shape and are shaped by our capacity to imagine futures, which is a fundamental element of socio-political life.

In architecture, AI influences social imaginaries by redefining the practice and perception of the field. Architecture inherently involves problem-solving, where the architect uses critical thinking to address issues and represent solutions, thereby shaping social imaginaries. However, this process is more complex than just the reconciliation of technological advancement with enhanced "humanities" skills. The creation and perpetuation of social imaginaries in architecture require a more profound understanding of how these imaginaries are formed and how they interact with technological advancements. Similarly, in the music industry, AI reshapes social imaginaries around creation, dissemination, and reception, challenging traditional notions of authorship, authenticity, and cultural diversity. Musicians and industry professionals must navigate an environment where AI-driven algorithms significantly influence consumer preferences, shaping musical innovation and cultural expression.

To effectively counter AI's tendency toward standardization, professionals must reclaim their role in shaping these imaginaries by leveraging AI as a tool for creative augmentation, not replacement.

In conclusion, the integration of AI into CCI signifies a profound shift in societal perceptions and conceptions. The theory of social imaginaries provides a critical framework for understanding these changes. Professionals in CCI must critically assess AI's impact on social imaginaries, balancing technological opportunities with the preservation of human creativity's essence. This paper calls for a broader, theoretically grounded and transdisciplinary discussion on the intersection of AI and creativity, urging stakeholders to engage in dialogue that will harness AI's transformative potential while safeguarding the integrity of human expression. AI is not inherently problematic; it is a tool capable of performing tasks and calculations beyond human capacity.

However, its fetishization, driven by the social imaginaries associated with it, constructs its perceived power. By fostering critical reflection and dialogue,



stakeholders can harness AI's potential to catalyze genuine artistic revolution and preserve the essence of human expression in architecture and music.



## Bibliography

Adorno, T. W. (1964) "L'industrie culturelle", in *Communications*, 3, 12-18.  
<https://doi.org/10.3406/comm.1964.993>.

Adorno, T. W. (1994). *Introduction à la sociologie de la musique. Douze conférences théoriques*, Paris, Contrechamps.

Adorno, T. W. & Horkheimer M, (1947) *Dialektik der Aufklärung : Philosophische Fragmente*, Amsterdam, Querido.

Airoldi, M. (2015). "Potrebbe interessarti anche: recommender algorithms e immaginario, il caso YouTube", In *Im@go. A Journal of the Social Imaginary*, 6, 132-150.

ISO 690

Barfield W, Shieldst R, Cooper S. (1993). "A survey of computer-aided design: Implications for creativity, productivity, decision making, and job satisfaction". In *J Hum Factors Manuf*, 3(2), 153-167. Available from:  
<https://doi.org/10.1002/hfm.4530030205>.

Barbin, É. (2019). Monge's descriptive geometry: His lessons and the teachings given by Lacroix and Hachette. In É. Barbin, M. Menghini, & K. Volkert (Eds.), *Descriptive*



*geometry, the spread of a polytechnic art* (pp. 1-25). Springer, Cham.  
[https://doi.org/10.1007/978-3-030-14808-9\\_1](https://doi.org/10.1007/978-3-030-14808-9_1).

Beer, D., & Burrough, R. (2013). "Popular Culture, Digital Archives and the New Social Life of Data", in *Theory, Culture & Society*, 30(4), 47-71.

Benjamin, W., (1936), *L'œuvre d'art à l'époque de sa reproduction mécanisée*, Available from:<http://www.hypermedia.univparis8.fr/Groupe/documents/Benjamin/Ben3.html#ref>.

Bhavnani SK, Flemming U, Forsythe DE, Garrett JH, Shaw DS, Tsai A. (1996), "CAD usage in an architectural office: from observations to active assistance", in *Autom Constr.* 5(3):243-255. Available from: [https://doi.org/10.1016/0926-5805\(96\)00149-5](https://doi.org/10.1016/0926-5805(96)00149-5).

Bourdieu, P. (1979). *La distinction. Etude critique du jugement*, Paris, édition de Minuit.

Bowering J. (2007), "Sensory deprivation: globalisation and the phenomenology of landscape architecture", in *Globalisation and landscape architecture: Issues for education and practice*. 81-84.

Campbell, P. (2014). "Imaginary Success? -The Contentious Ascendance of Creativity", in *European planning studies*, 22(5), 995-1009.

Castoriadis, C. (1987). *The imaginary institution of society*. MIT Press.

Certeau M. de, (1990), *L'invention du quotidien. 1. Arts de faire*, Paris, Gallimard.

Chengyuan, L., Xie, H., Li, Q., Tang, Z., & Zheng, Y. (2024). "Generative AI for architectural design: A literature review", In *arXiv*.  
<https://doi.org/10.48550/arXiv.2404.01335>.

Connor, J. O', (2015), "Intermediaries and Imaginaries in the Cultural and Creative Industries." *Regional studies* (49)3, 374-387. Web.

Cooley MJE. (1977), "Impact of CAD on the designer and the design function", in *Comput-Aided Des*, 9(4), 238-242. Available from: [https://doi.org/10.1016/0010-4485\(77\)90004-5](https://doi.org/10.1016/0010-4485(77)90004-5).

Coyne RD, Sudweeks F, Haynes D. (1996) "Who needs the Internet? Computer-mediated communication in design firms", in *Environ Plan B: Plan Des.* ,23, 749-770.

DamiLee. (2023, June 14). *We tried to compete with AI... [AI vs. ARCHITECT]* [Video]. YouTube. <https://www.youtube.com/watch?v=N709ZrxoIP0&t=650s>.



DeNora, T. (2000). *Music in Everyday Life*, Cambridge (UK), Cambridge University Press.

Dewey, J. (2005). *L'art comme expérience*, Paris, Gallimard.

Durand-Folco, J, & Martineau, J, (2023), *Le capital algorithmique*, Montréal, Écosociété.

El Moussaoui M. (2020) "The Ocular-Centric Obsession Of Contemporary Societies", in *Civil Eng Archit.* 8(6),1290–1295.

Elias, N. (1991). *Mozart. Zur Soziologie eines Genies*, Francfort-am-Main, Suhrkamp.

Florida, R. L. (2002). *The rise of the creative class: and how it's transforming work, leisure, community and everyday life*. New York, NY, Basic Books.

Gero JS. (1983). "Computer-Aided Architectural Design—Past, Present And Future", in *Archit Sci Rev*, 26, 2-5.

Giddens, A. (1984). *The Constitution of Society: Outline of the Theory of Structuration*. Cambridge (MA), Polity Press.

Hagen, A. N. (2015). "The Playlist Experience: Personal Playlists in Music Streaming Services", *Popular Music and Society*, 38(5), 625-645.

Hagen, A. N., & Lüders, M. (2016). "Social streaming? Navigating music as personal and social", in *Convergence: The International Journal of Research into New Media Technologies*, 23(6), 643-659.

Hanrahan, N. W. (2016). "La découverte musicale en régime numérique", in Le Guern A. (dir.), *Où va la musique ?*, Paris, Presse des Mines.

Hanrahan, N. W. (2018). "Hearing the Contradictions: Aesthetic Experience, Music and Digitization", in *Cultural Sociology*, 12(3), 289-302.

Hennion, A. (1981). *Les professionnels du disque. Une sociologie des variétés*, Paris: A.-M. Métailié.

IFPI (2024), *Global Music Report 2024*, Accessible from : [https://ifpi-website-cms.s3.eu-west-2.amazonaws.com/IFPI\\_GMR\\_2024\\_State\\_of\\_the\\_Industry\\_db92a1c9c1.pdf](https://ifpi-website-cms.s3.eu-west-2.amazonaws.com/IFPI_GMR_2024_State_of_the_Industry_db92a1c9c1.pdf).

Introna, L. D. (2016). "The algorithmic choreography of the impressionable subject", in In Roberge, J. & Seyfert, R. (Eds.), *Algorithmic Culture* (pp. 26-51). New York City: Routledge.

James, P. (2019). "The social imaginary in theory and practice." In C. Hudson & E. K.



Wilson (Eds.), *Revisiting the Global Imaginary: Theories, Ideologies, Subjectivities: Essays in Honor of Manfred Steger* (pp. 33-47). [https://doi.org/10.1007/978-3-030-14911-6\\_3](https://doi.org/10.1007/978-3-030-14911-6_3).

Jamet, R. (2019), "Streaming the mood: dispositifs de recommandation algorithmique et expériences sociétales de la musique", in *Milieu(x)*, 5, 44-49.

Jamet, R & Blum, G (2022), "Comment spotify a transformé nos artistes en aspirateur", *Le Devoir*, Accessible from : <https://www.ledevoir.com/opinion/idees/669714/idees-comment-spotify-a-change-les-artistes-en-aspirateurs>.

Jamet, R. et Roberge, J., (2020). "La musique québécoise est-elle compatible avec le streaming? Usages et représentations de la musique francophone sur les plateformes de streaming au Québec", in *TIC et Société*, 14 (1-2), 221-246.

Kracauer S., (1971), *Die Angestellten*, Frankfurt, Suhrkamp.

Kumar, A., Tanwar, S., Singh, P. K., Ganzha, M., & Epiphaniou, G. (2024). "Navigating the realm of generative models: GANs, diffusion, limitations, and future prospects—A review", In S. Tanwar, P. K. Singh, M. Ganzha, & G. Epiphaniou (Eds.), \*Proceedings of Fifth International Conference on Computing, Communications, and Cyber-Security. IC4S 2023\* (Vol. 991, pp. 329-343). Springer, Singapore. [https://doi.org/10.1007/978-981-97-2550-2\\_23](https://doi.org/10.1007/978-981-97-2550-2_23).

Lanza, J. (2004). *Elevator Music: A Surreal History of Muzak, Easy-listening, and Other Moodsong*, Chicago (MI): University of Michigan Press.

Le Monde/AFP (2023), Le « New York Times » poursuit en justice Microsoft et OpenAI, créateur de ChatGPT, pour violation de droits d'auteur, Le Monde, 27/12/2023, [https://www.lemonde.fr/pixels/article/2023/12/27/le-new-york-times-poursuit-en-justice-microsoft-et-openai-createur-de-chatgpt-pour-violation-de-droits-d-auteur\\_6207946\\_4408996](https://www.lemonde.fr/pixels/article/2023/12/27/le-new-york-times-poursuit-en-justice-microsoft-et-openai-createur-de-chatgpt-pour-violation-de-droits-d-auteur_6207946_4408996).

Leach N. (2022), "Architecture in the Age of Artificial Intelligence", in *Bloomsbury Publishing*; p. 3-14.

Lunden, I. (2014), "Spotify Acquired Music Tech Company The Echo Nest In A \$100M Deal", *Techcrunch*, 7/3/2014, <https://techcrunch.com/2014/03/07/spotify-echo-nect-100m/>.

Manovich, L. (2022), "AI and Myths of Creativity", in *Archit. Design*, 92: 60-65. <https://doi.org/10.1002/ad.2814>.



Marzo, P. L., & Mori, L. (2019), *Le vie sociali dell'immaginario: Per una sociologia del profondo*, Milano, Mimesis.

Mazzone M., Elgammal A. (2019), "Art, Creativity, and the Potential of Artificial Intelligence", in *Arts*; 8(1):26. <https://doi.org/10.3390/arts8010026>.

Musso, M.G. (2019), "Immaginario, tecnica e mutamento sociale", in Marzo P. L. e Mori L., *Le vie sociali dell'immaginario. Per una sociologia del profondo*, Mimesis, Milano.

Napolitano, D. (2023), "The Shaping of a Standard Voice: Sonic and Sociotechnical Imaginaries in Smart Speakers", in *Im@go. A Journal of the Social Imaginary*, (21), 177-196.

Natale, S., & Ballatore, A. (2020), "Imagining the thinking machine: Technological myths and the rise of artificial intelligence", in *Convergence*, 26(1), 3-18. <https://doi.org/10.1177/1354856517715164>.

Oxman R. (2006), "Theory and design in the first digital age", in *Des Stud.*,27(3), 229-265.

Pallasmaa J. (2005), *The Eyes Of The Skin. Architecture And The Senses*. Chichester, Wiley.

Pariser, E. (2011), *The Filter Bubble: What the Internet Is Hiding from You*, NYC (NY), Penguin Press.

Pasquale, F (2016), *The Black Box Society. The Secret Algorithms That Control Money and Information*, Cambridge (MA), Harvard University Press.

Pecqueux, A. et Roueff, O. (2009), *L'écologie sociale de l'oreille, enquête sur les expériences musicales*, Paris, EHESS.

Perrenoud, M. (2007), *Les Musicos. Enquête sur les musiciens ordinaires*. Paris, La Découverte.

Roberge, J. et Seyfert, R. (2016), *Algorithmic cultures : Essays on Meaning, Performance and New Technologies*, NYC (NY), Routledge.

Roberge, J., Jamet, R. & Rousseau, A. (2019). *L'impact social des algorithmes de recommandation sur la curation des contenus musicaux francophones au Québec*, Québec, INRS.

Robertson BF, Radcliffe DF. (2009). "Impact of CAD tools on creative problem solving in engineering design", in *Comput-Aided Des*, 41(3), 136-146. Available from:



<https://doi.org/10.1016/j.cad.2008.06.007>.

Robins-Early, N. (2024), "Billie Eilish, Nicki Minaj, Stevie Wonder and more musicians demand protection against AI", in *The Guardian*, 2/4/2024, <https://www.theguardian.com/technology/2024/apr/02/musicians-demand-protection-against-ai>.

Rouvroy, A. & Berns, T. (2013). "Gouvernementalité algorithmique et perspectives d'émancipation: Le disparate comme condition d'individuation par la relation ?", *Réseaux*, 177, 163-196. <https://doi.org/10.3917/res.177.0163>.

Schütz, A. (2007). *Ecrits sur la musique. 1924-1956*, Paris, Edition MF.

Snickars, P. (2017). "More of the Same – On Spotify Radio", in *Culture Unbound*, 9(2), 184-211.

Stanislas C. (2023), "When Form Follows Meanings: AI's Semantic Turn in Architecture", in *Architect*. 2023 June 12. Available from: <https://architect.com/features/article/150350201/when-form-follows-meanings-ai-s-semantic-turn-in-architecture>.

Tencer, D. (2023), "AI 'fake Drake' track deleted on spotify, youtube, tiktok after Universal Music Group copyright claim", in *Music Business Worldwide*, 18/4/2023, <https://www.musicbusinessworldwide.com/ai-fake-drake-track-deleted-on-spotify-youtube-tiktok-millions-of-plays/>.

Weber M. (1998, 1958). *Sociologie de la musique. Les fondements rationnels de la musique*, Paris, A-M Métaillé.

Wikström, P. (2020). *The Music Industry: Music in the cloud*, Polity.

Zhou, E., & Lee, D. (2024). Generative artificial intelligence, human creativity, and art. *PNAS Nexus*, 3(3), pgae052. <https://doi.org/10.1093/pnasnexus/pgae052>.





*Off-Topic*





# Manosfera e pick-up artists


Giovanna Vingelli

[giovanna.vingelli@unical.it](mailto:giovanna.vingelli@unical.it)

*Dipartimento di Scienze politiche sociali | Università della Calabria*

## Abstract

### *Manosphere and Pick-up artists*




This article analyses pick-up artist (PUA) communities within the manosphere. Pick-up artists (PUAs) are communities of self-proclaimed seduction experts or wannabes who use a set of routines, techniques and 'scripts' that men must follow to achieve sexual success with a woman. After defining the key terms, the article describes the ideologies, beliefs and discourses that animate PUA groups, including a comparison with other groups in the manosphere (incels). The reconstruction of academic contributions from recent years, especially from the Anglophone world, provides a chiaroscuro analysis. The changes in PUA narratives and practices, especially in the European context, call for new theoretical and empirical analyses capable of intercepting the new forms of masculinity that are thickening on the border between neoliberalism and patriarchal backlash.

## Keywords

Manosphere | Antifeminism | Pick-up artists | Incel

You need cooling / Baby, I'm not fooling / I'm  
gonna send you back to schooling

Led Zeppelin, Whole Lotta Love (1969)



L'articolo analizza le comunità dei pick-up artist (PUA) all'interno dello spazio della manosfera. I pick-up artist sono comunità di esperti di seduzione, autodefiniti o aspiranti tali, che utilizzano una serie di routine, tecniche e "copioni" che gli uomini devono seguire per ottenere il successo sessuale con una donna. Dopo la definizione dei termini principali, saranno descritte le ideologie, le credenze, i discorsi che animano i gruppi PUA, anche attraverso un confronto con altri gruppi della manosfera (incel). La ricostruzione dei contributi accademici negli ultimi anni, in particolare di area anglofona, restituisce un'analisi in chiaroscuro. Le modifiche nella narrazione e nelle pratiche dei PUA, in particolare in ambito europeo, richiede analisi teoriche ed empiriche rinnovate, capaci di intercettare le nuove forme di mascolinità che si addensano al confine fra neoliberalismo e backlash patriarcale.

## 1. Antifemminismo e manosfera

L'antifemminismo è una corrente eterogenea, attraversata da differenti ideologie, e presente in diversi ambiti. La copertura mediatica dei movimenti antifemministi ha messo in luce comunità e attività specifiche (la manosfera), che riflettono idee misogine e antifemministe, spesso collegate a movimenti suprematisti, nativisti, alla destra radicale o all'*alternative right*.<sup>1</sup> La proliferazione di tecnologie digitali ha contribuito a dare forma a nuovi tipi di pratiche, e l'opposizione al femminismo è diventata più vocale, transnazionale e meglio organizzata, riunendo attori consolidati e nuovi: attivisti maschilisti, movimenti anti-gender ed eteroattivisti, gruppi religiosi, attori conservatori e gruppi nazionalisti e populistici di destra (Roggeband, Krizsán, 2020). Il termine manosfera è oggi utilizzato nella letteratura per indicare differenti comunità online di uomini che si riuniscono intorno all'opposizione al femminismo: dagli attivisti per i diritti degli uomini (*Men's Rights Activists* - MRAs), i *Men Going Their Own Way* (MGTOW), i *Pick Up Artists* (PUA), gli incels (*involuntary celibates*), i movimenti per i diritti dei padri e così via. Lunghi dall'essere omogenea, si tratta una galassia virtuale di spazi in cui si esaltano la riscoperta e la supremazia della mascolinità (egemone) e la subordinazione delle donne, presentando le conquiste del movimento femminista come tentativo di erosione dei diritti degli uomini (Dordoni, Magaraggia, 2021). Uno degli elementi unificatori è il riferimento alla teoria Red Pill

<sup>1</sup> Il Southern Poverty Law Center (SPLC) ha recentemente incluso le comunità suprematiste maschili tra i gruppi d'odio, evidenziando i legami tra la supremazia bianca e la supremazia maschile, entrambe guidate dalla paura e dalla rabbia per la perdita dello status di maschio bianco. Cfr. "Supremazia maschile", SPLC (<https://www.splcenter.org/fighting-hate/extremist-files/ideology/male-supremacy>)



(TRP), la “pillola rossa” che - come nel film Matrix - permette di risvegliarsi, “aprire gli occhi” per vedere il mondo nella sua vera forma. Ging (2019) definisce la TRP come una visione del mondo che bilancia ideologia ed emozioni allo scopo di generare consenso e appartenenza tra gli elementi divergenti della manosfera, fornendo un sistema di credenze condivise dalla comunità e generando una *networked misogyny* (Banet-Weiser, Miltner, 2015).<sup>2</sup> Allontanandosi dai tradizionali discorsi misogini, oggi le organizzazioni della manosfera interpretano una nuova retorica, che si focalizza sulla presunta discriminazione sociale e istituzionale a danno degli uomini, attivata dalle pratiche, le politiche e le narrazioni femministe. In questa narrazione, esiste una crisi contemporanea della mascolinità che deriva dal fatto che gli uomini sono esplicitamente oppressi, emarginati e discriminati; gli uomini sono in crisi e soffrono a causa delle donne in generale e delle femministe in particolare, e la soluzione ai loro problemi risiede nel contenere l'influenza del femminismo e nel rivalutare la mascolinità. A condividere questa visione del mondo ci sono anche i pick-up artist, che possono essere considerati una “industria della seduzione”; non rappresentano le ideologie più estreme della manosfera, ma i loro discorsi e l'immaginario sulle donne sono in sintonia con essa: le donne sono universalmente ritratte come inferiori agli uomini e come oggetti destinati a riceverne le attenzioni; la manipolazione e la coercizione sono presentate come modi adeguati di interagire con le donne; i discorsi antifemministi abbondano; e le questioni di violenza sessuale sono negate, minimizzate o apertamente approvate (Gotell, Dutton, 2016).

## 2. I pick-up artist

Ogni gruppo della manosfera presenta ideologie specifiche, a volte in conflitto fra loro: i MRAs credono che gli uomini siano discriminati in una società ginocentrica (Vingelli 2019); gli Incel attribuiscono la responsabilità di non riuscire a trovare partner romantici alle donne e alla società (Tranchese, Sugiura, 2021); i MGTOW credono che le donne siano così tossiche da dover essere evitate il più possibile (Lin, 2017). I *pick-up artists* (PUA) sono uomini<sup>3</sup> che si dedicano a migliorare le proprie capacità con il sesso opposto e che praticano strategie di seduzione rapida delle donne, influenzati dalla convinzione che l'applicazione di specifiche routine e *script* siano la chiave per avere successo in questa impresa (Dayter, Rüdiger, 2022). “Artisti del rimorchio”, sono una comunità costruita intorno alla definizione di “gioco”, un

<sup>2</sup> La manosfera si interseca anche con l'alt-right, abbracciando il nazionalismo bianco, l'antisemitismo e l'anti-islamismo. Non sorprende quindi che per giovani uomini immersi nella cultura di internet i gruppi della manosfera sia spesso un trampolino di lancio verso le convinzioni suprematiste bianche (Marwick & Caplan 2018; Mamié et al. 2021), mentre lo spostamento verso una narrazione antifemminista nella destra alternativa può essere correlato allo sviluppo di legami con la manosfera (Lyons 2017). La politica patriarcale (Lyons 2017) ha guadagnato sempre più spazio nei movimenti dell'alt right e del nazionalismo bianco e offre una pericolosa coalizione tra la politica di destra (estremista) e le visioni del mondo misogine, integrando la convinzione che gli uomini siano vittime di ingiustizie da parte delle femministe e il razzismo inverso (i bianchi oppressi nella società contemporanea).

<sup>3</sup> Meno comunemente donne (FPUA)

insieme di tecniche, strategie e mentalità che possono aiutare gli uomini (eterosessuali) a sedurre le donne attraverso un misto di adulazione, manipolazione psicologica e coercizione. Anche se questo approccio può sembrare estraneo alla manosfera, il "gioco" spesso implica l'oggettivazione delle donne e la promozione di tecniche di molestia, come il *negging*.<sup>4</sup> Una strategia per assicurarsi l'interesse delle donne è infatti la manipolazione psicologica attraverso l'insulto subdolo: questo farebbe apparire l'uomo più sicuro di sé e contemporaneamente abbasserebbe l'autostima della donna, rendendola più accondiscendente. Gli uomini inoltre non dovrebbero rassicurare le donne sulla loro fedeltà, e non dovrebbero mai perdere l'occasione di farle ingelosire:

Women will never admit this but jealousy excites them. The thought of you turning on another woman will arouse her sexually. No girl wants a man that no other woman wants. The partner who harnesses the gale storm of jealousy controls the direction of the relationship<sup>5</sup>



L'industria dei PUA comprende oggi centinaia di siti web e organizzazioni in tutto il mondo, che offrono costosi seminari e campi di addestramento. Divenuti comunità discorsive e di pratiche - in cui i membri condividono le loro esperienze attraverso un gergo di gruppo basato sulla terminologia militare, sportiva e commerciale - come altri gruppi della manosfera si nutrono di teorie pseudoscientifiche, supportati da dati e ricerche metodologicamente carenti. Diversamente, tuttavia, dagli Incel e dai MGTOW, i pick-up artist mirano a raggiungere il successo sessuale in un gioco di conquista guidato da guru che offrono specifiche (in)formazioni su come risultare vincenti (Scotto di Carlo, 2023). A differenza degli Incel, che spiegano la difficoltà di intraprendere relazioni sentimentali e sessuali con il determinismo biologico e il fatalismo, i PUA sostengono che l'arte della seduzione può essere "appresa", migliorando il proprio "gioco interiore" (personalità e sicurezza) e il "gioco esteriore" (aspetto fisico).

Le origini dei gruppi PUA contemporanei risalgono agli anni Settanta del XX secolo, con la pubblicazione di libri che promuovono tecniche di programmazione neuro-linguistica o la "seduzione veloce", come ad esempio *How to Pick Up Girls!* di Erik Webber (1970). La community dei PUA si è sviluppata a partire dal 1992, quando Ross Jeffries - a cui si ispira il personaggio interpretato da Tom Cruise nel film *Magnolia*<sup>6</sup> - pubblica il libro *How to Get the Women You Desire into Bed*. La vera

<sup>4</sup> Il *negging* prevede insulti sottili volti a minare l'autostima, così che le vittime possano sentirsi più vulnerabile e quindi più facile da sedurre. Si tratta di una manipolazione ambigua in cui non ci si espone mai del tutto ma si instilla il dubbio nella persona corteggiata che si sentirà criticata, denigrata e svaloriata.

<sup>5</sup> "The Sixteen Commandments of Poon," *Chateau Heartiste*, May 19, 2019, <https://heartiste.org/the-sixteen-commandments-of-poon/>.

<sup>6</sup> La figura del pickup artist/seduttore seriale è negli ultimi anni oggetto di rappresentazioni in film e serie tv. Nella commedia "The Pick-Up Artist" (1987) di Molly Ringwald, Robert Downey Junior interpreta Casanova ed utilizza una serie di battute e routine strettamente programmate. Più recente è il film "Magnolia" (1999) con Tom Cruise che interpreta Frank T. J. Mackey, personaggio che predica il suo marchio di seduzione particolarmente tossico ("rispetta il cazzo, doma la fica") nel corso di seguitissimi



esplosione del fenomeno si ha nel 2005 con la pubblicazione di *The Game. Penetrating the Secret Society of Pick-Up Artists* di Neil Strauss, testo ormai considerato il riferimento classico dell'“arte di rimorchiare”. Il libro è la narrazione del viaggio che l'autore conduce per diventare un “guru del pick-up” (PUG): il presupposto è che, con l'adeguata conoscenza dei meccanismi che stanno alla base dei rapporti umani, ogni uomo possa elevare il proprio status e sedurre qualsiasi donna desideri.

Attraverso un mix di tecniche motivazionali, formule comportamentali sbrigative e superficiali nozioni di psicologia evolutiva che danno una parvenza scientifica al tutto, il testo fornisce una rassicurante spiegazione per ogni insuccesso e un altrettanto confortante soluzione. Dopo il 2005, i PUA acquistano visibilità, condividendo e monetizzando le tecniche di seduzione attraverso workshop, libri e sessioni di coaching, con un giro di affari che vale circa 100 milioni di dollari all'anno (O'Neill, 2018). Nascono negli USA “The Style Life Academy” di Neil Strauss, “The Venusian Arts” di Mystery e in Europa PUATraining di Richard La Ruina (2006), che mettono a disposizione degli aspiranti PUA corsi base e intensivi, centinaia di articoli sui blog, mailing list, gruppi Facebook e Instagram privati, incontri one-to-one e *bootcamp* con i guru del pick up dove viene sviluppata la “scienza della seduzione”, svelata ai cosiddetti “**sfigati ordinari (o medi) frustrati**” (AFC, *Average frustrated chump*) attraverso un caleidoscopico mix di **tecniche motivazionali** e psicologia, sociologia, fisiologia e **neurolinguistica che hanno l'obiettivo di far capitolare ogni donna**.

I PUG contribuiscono abitualmente a rubriche, scrivono articoli spesso pubblicati in note riviste come *Cosmopolitan* e *Men's Health*, e somministrano consigli ad un ampio pubblico tramite popolari siti online come AskMen dove si presentano come “esperti di appuntamenti” e “lifestyle coaches”.<sup>7</sup>

La pedagogia PUA descrive un “mercato sessuale” in cui ogni persona ha un valore di mercato (SMV) determinato da caratteristiche quali aspetto fisico, forma fisica, età, ricchezza e classe sociale. I forum PUA sostengono che questo sistema è stato creato (e attivamente sostenuto) dal femminismo, e controllato da donne che cercano uomini con un SMV più alto del loro (“iper-gamia femminile”). I PUA suggeriscono agli

---

seminari pubblici. Anche alcune serie tv come “How I Met Your Mother” (2005-14) rappresentano al loro interno personaggi che utilizzano la teoria della seduzione per trasformare le loro amicizie in incontri sessuali; in particolare, il personaggio di Barney Stinson (Neil Patrick Harris) fa del sedurre le donne una vera e propria ossessione che lo porta ad inventare strategie sempre più elaborate. Le tecniche messe in atto sono poi raccolte nel “Playbook”, libro scritto dal personaggio che raccoglie tutte le sue tecniche di seduzione, rendendo ancora più evidente la connessione con il mondo dei pick-up artist.

<sup>7</sup> A seguito di una serie di scandali e controversie, i pick-up artist sono stati ampiamente trattati dai media, ma allo stesso tempo sono diventati oggetto di analisi, non solo accademica. È del 2019 il documentario *The Pickup Game* ([www.thepickupgame-film.com](http://www.thepickupgame-film.com)), diretto da Matthew O'Connor e Barnaby O'Connor. Gran parte dell'attenzione degli studiosi nei confronti dei PUA è stata rivolta dal punto di vista della sociologia (O'Neill, 2015, 2017, 2018; Schuurmans, Monaghan, 2015), e dagli studi di genere e femministi (Denes, 2011; Ging, 2019). Gli artisti pick-up non sono un fenomeno solo anglofono e sono state identificate e descritte forme localizzate di comunità, ad esempio, nei contesti tedesco (Schutzbach, 2018), cinese (Guo, Yin 2021), italiano (Scotto di Carlo, 2023). Infine, i PUA hanno ricevuto attenzione da parte dei/le linguisti/e (Hambling-Jones, Merrison, 2012), spesso sotto l'ombrello della manosfera e in contrasto con altri gruppi come gli MGTOW, incels, ecc. (Krendel, 2020; Dayter, Rüdiger 2016, 2019, 2020; 2023; Rüdiger, Dayter, 2017, 2020).



uomini di migliorare il loro SMV imparando il “gioco”, guadagnando più soldi e/o migliorando il loro aspetto fisico. In questo scenario, le donne sono descritte come superficiali, manipolabili e non meritevoli di rispetto o empatia, e gli uomini come vittime di un sistema femminista ingiusto. Per definizione, gli artisti del pick-up non rifiutano questo sistema, ma cercano piuttosto di manipolarlo per trarne vantaggio e, in virtù di ciò, di ascendere e diventare essi stessi degli uomini alpha.

Le strategie utilizzate dai pick-up artist – presentate come scientificamente affidabili – si rifanno anche alla cosiddetta **programmazione neurolinguistica (NLP)**. Per migliorare le proprie tattiche di seduzione, le tecniche dei PUA intervengono su due livelli, sul **miglioramento dell’individuo** (“cura il tuo look”, “scegli un taglio di capelli”, “diventa loquace” e suggerimenti motivazionali: “dimostra energia maschile”) e sulle **tattiche di approccio** (spiegazioni pseudoscientifiche sull’utilizzo del linguaggio del corpo e del contatto visivo fino a vere e proprie **“aperture e routine”** attraverso cui avvicinare il “target”). Le donne sono infatti solo bersagli, oggetti (sessuali) da conquistare attraverso strategie studiate per farle capitolare.

Le community utilizzano un lessico specifico, attraverso cui le persone vengono etichettate, insieme alla descrizione di **tecniche, strategie, situazioni. Vengono classificati** i vari maschi alpha e beta ma soprattutto i bersagli: quelli ideali sono definiti **HB, Hot Babe** (solitamente classificate con un voto da 6 a 10; es. HB8), mentre le UG, (Ugly girl) ricevono **voti da 1 a 5**. Il manuale di seduzione *The Mystery Method: How to Get Beautiful Women into Bed* (2007) scritto da un pick-up artist americano noto come Mystery (Erik von Markovik) identifica nelle interazioni eterosessuali tre fasi distinte: Attrazione, Comfort e fiducia, Seduzione, elaborando la “regola delle sette ore”, basata sull’idea che le donne debbano trascorrere sette ore con un uomo prima di fare sesso con lui (avendo tipicamente bisogno di più tempo e intimità per sviluppare la stessa quantità di passione degli uomini). Queste guide di seduzione tendono a privilegiare le risposte biologiche come verità assoluta (Denes, 2011). L’effetto finale è che le differenze supposte tra uomini e donne, e le disuguaglianze che ne derivano, vengono naturalizzate come una questione di “eredità genetica” (Cameron, 2009: 175), privilegiando sistematicamente le prospettive e le esperienze degli uomini rispetto a quelle delle donne. In questa direzione, si può individuare un’affinità tra gli argomenti dei PUA e una più complessa “industria della seduzione”, sostenuta ed aiutata da certa psicologia evolutiva che ne fornisce un alibi accademico, che afferma di aver scoperto la “verità” in merito alla sessualità femminile, ma ignora le esperienze ed il punto di vista delle donne stesse (O’Neill, 2018).<sup>8</sup> Denes (2011) analizza in particolare il testo di Mystery,

<sup>8</sup> Ad esempio, Oesch e Miklousic (2012) sostengono che, malgrado la sua etichetta provocatoria, l’industria della seduzione è fondata su ricerche empiriche solide ed esperienze di relazione e corteggiamento innovative, ad esempio sfruttando gli spunti evoluti di ciò che le donne trovano generalmente attraente negli uomini. Secondo gli autori, molte affermazioni dei siti PUA presentano un buon grado di ricerca psicologica, e che questa conoscenza potrebbe essere di grande aiuto per le coppie, riducendo conflitto e frustrazione e innescando relazioni più soddisfacenti.



in cui l'autore insegna agli uomini a interpretare l'eccitazione fisica delle donne come consenso, ignorando qualsiasi comunicazione verbale contraria.

Altri riferimenti alla cultura dello stupro sono presenti nel manuale autopubblicato di Ken Hoinsky, *Above the Game: A Guide to Getting Awesome with Women* (2013). Nel suo libro, Hoinsky sostiene che "all the greatest seducers in history could not keep their hands off women. They aggressively escalated physically with every woman they were flirting with." L'autore si spinge oltre, incoraggiando i suoi lettori a forzare l'interazione sessuale:

Physically pick her up and sit her on your lap. Don't ask for permission. Be dominant. Force her to rebuff your advances [...] Pull out your cock and put her hand on it. Remember, she is letting you do this because you have established yourself as a LEADER. Don't ask for permission, GRAB HER HAND, and put it right on your dick.



L'insieme di conoscenze e pratiche dei PUA sono organizzate intorno alla convinzione che la capacità di incontrare e attrarre le donne sia un'abilità che gli uomini eterosessuali possono coltivare attraverso la formazione pratica e lo sviluppo personale, prevedendo altresì l'utilizzo di una serie di strategie: uscire per incontrare il maggior numero di donne possibile (*sarging*) e la reiterazione di specifici approcci per superare le resistenze, eventualmente anche dopo un primo rifiuto. I PUA sostengono infatti che il "no" di una donna è spesso solo una "resistenza simbolica" (Osman, 2013) volta a testare l'autocontrollo di un uomo o a difendersi dall'essere etichettata come "donna facile". L'addestramento dei PUA si basa così sulla strumentalizzazione e l'oggettivazione delle donne, che diventano veicoli per il successo e la realizzazione maschile. In questo orizzonte, le donne devono essere controllate e la loro autostima deve essere erosa, in modo da aumentare al contrario l'autostima maschile:

Confidence here means procuring trust and taking one into confidence, the core technique of getting the con game to stick—in this case, conning her into having sex with you. In this way, the PUA industry is neoliberal misogyny at its peak—a self-help industry that encourages masculine entrepreneurship for self-confidence, relies extensively on experts as pedagogical guides, and expects women to enact social reproduction (Bratich, Banet-Weise, 2019: 5013).

Rosalind Gill (2009) identifica diversi modelli di rappresentazione o repertori discorsivi dei PUA. Il primo di questi pattern è l'"imprenditorialità intima", basata sull'idea che il sesso e le relazioni debbano essere meticolosamente pianificati, continuamente valutati e attivamente gestiti. In secondo luogo, i repertori dei PUA si collocano in una vera e propria "men-ology", dove viene spiegato come sono gli uomini e vengono date alle donne istruzioni dettagliate su come compiacerli: le donne sono quindi incoraggiate a ripensare il modo in cui vedono il proprio corpo, i propri desideri, le relazioni e le pratiche sessuali. O'Neill (2018) riconosce infine le comunità della seduzione come siti di "intimità mediata", che elaborano competenze

attraverso rappresentazioni mediatiche che ci circondano e attraversano le nostre vite nella forma di consigli relazionali e sessuali.

### 3. Fra virtuale al reale. PUA e Incel

Nel 2014 Julien Blanc (autoproclamatosi pick-up guru) si vede revocato il visto per il Regno Unito e l'Australia dopo aver pubblicato un video (introdotto dall'eloquente hashtag #ChokingGirlsAroundtheWorld) in cui incoraggiava gli uomini ad afferrare le donne senza il loro consenso e a forzare la testa verso il loro inguine. Precedentemente, Blanc aveva commentato su Facebook un'infografica sul ciclo della violenza domestica dicendo che si trattava in pratica di una "lista" di cose da fare per conquistare una donna. In molti altri paesi, inclusa l'Italia, sono state lanciate petizioni per impedire i suoi workshop itineranti, che includono tattiche che prevedono la manipolazione e la violenza domestica.<sup>9</sup>

Altri PUG sono stati criticati per il loro comportamento sessista e talvolta molesto (O'Neill, 2018). Ad esempio, Daryush Valizadeh—meglio noto come Roosh V, è un blogger che ha ripetutamente sostenuto il comportamento sessuale predatorio e violento nei suoi libri, sul suo blog e sulla sua popolare piattaforma "Return of Kings".

Nella sua narrazione le donne sono intellettualmente inferiori agli uomini, necessarie solo per il piacere sessuale e la riproduzione, e non dovrebbero godere del diritto di voto. Nel suo Manifesto della neomascolinità (2015a) sostiene:

A woman's value significantly depends on her fertility and beauty. A man's value significantly depends on his resources, intellect, and character [...] We've arrived at the point where men are so consistently oppressed from birth to adulthood that any notion of female oppression can be immediately laughed at on its face [...] A boy born today will be institutionally, systematically, and deliberately oppressed by all facets of society while a girl born today will be given undeniable preferential treatment over him.

Dopo aver pubblicato guide di viaggi in cui offriva consigli su come orientarsi nella cultura e nella vita notturna di alcune destinazioni turistiche per avere rapporti sessuali con le donne, Roosh V si è poi allontanato dalle comunità PUA e ha scritto principalmente di mascolinità e maschilismo – chiudendo il sito *Return of the Kings* nel 2023, dopo essersi convertito al cristianesimo<sup>10</sup> -, sostenendo un modello di

<sup>9</sup> La campagna online e l'hashtag #TakeDownJulienBlanc, diventati presto virali, hanno suggerito alla rivista Time di definire Julian Blanc, nel 2014, "l'uomo più odiato al mondo" (Gibson, 2014). In seguito, Blanc è apparso in un video della CNN scusandosi per i contenuti pubblicati, affermando che gran parte di ciò che era stato considerato offensivo era un "orribile, orribile tentativo di umorismo" ed era stato "estrapolato dal contesto" (Lana Massi 2014, 02:53). Blanc continua a lavorare come executive coach, in maniera sempre più orientata all'autorealizzazione e all'autosviluppo (Blanc, 2017), rinunciando a gran parte dei contenuti basati sulla seduzione che venivano comunicati in precedenza nei suoi prodotti.

<sup>10</sup> Il suo sito web conteneva una serie di articoli sessisti e antifemministi, con titoli come "La Svezia sta costringendo i suoi giovani ragazzi all'omosessualità" (Roark, 2018), "Il movimento #MeToo ha rovinato la



mascolinità polarizzante che si allinea al concetto di maschilità egemonica di Connell (2005). In risposta agli omicidi di Isla Vista, dove Elliot Rodger ha ucciso sei persone (*infra*), Roosh V ha sostenuto che il pick-up e l'allineamento con le sue convinzioni neomaschiliste avrebbero impedito gli omicidi e facilitato l'accettazione sociale e sessuale dell'omicida (Dewey, 2014b). Nel 2015 Roosh V ha scatenato una serie di polemiche dopo aver pubblicato un articolo in cui ipotizzava la legalizzazione dello stupro come strumento di prevenzione. Legalizzando lo stupro, una donna sarebbe effettivamente responsabilizzata a proteggere il proprio corpo e a garantire la propria sicurezza in ogni momento: "a girl would absolutely not enter a private room with a man she doesn't know or trust unless she is absolutely sure she is ready to sleep with him" (Valizadeh, 2015b).<sup>11</sup>

Autrici come Sady Doyle sottolineano i pericolosi collegamenti fra i pick-up artists e gli incel, nonostante le narrazioni e le strategie differenti che caratterizzano i due movimenti. Gli incel sarebbero, in questa prospettiva, dei PUA disillusi: nonostante le indicazioni e le strategie apprese, incapaci di accedere ai "corpi femminili", e quindi "pronti a provare mezzi più violenti" (Doyle, 2018). Come è accaduto per gli attivisti per i diritti degli uomini, anche la parabola degli incel ha seguito una traiettoria non lineare. Sebbene il termine sia ora associato a giovani uomini, nei primi anni del XXI secolo non faceva riferimento a un genere specifico. Il termine è stato creato per la prima negli anni Novanta, in concomitanza con lo sviluppo di Internet e dei primi forum online, da una studentessa canadese bisessuale, nota sul web con il soprannome di «Alana»; sebbene la comunità associata al termine fosse, quasi per definizione, solitaria e insoddisfatta dal punto di vista sentimentale, inizialmente non era collegata all'odio e a minacce violente (Zimmer, 2018).

Il postulato alla base del movimento incel (e della teoria Red Pill) è quello dell'"iper-gamia", ovvero l'idea che la maggior parte delle donne è attratta da un numero limitato di uomini, detti Chad, che monopolizzerebbero gli incontri sessuali. La maggiore selettività femminile indirizza l'80% delle donne verso gli uomini Chad (appena il 20% degli uomini), lasciando così appena una minuscola possibilità all'80% di uomini non-Chad di trovare una partner.<sup>12</sup> L'argomento apparentemente paradossale è che la liberazione sessuale e la promiscuità femminile hanno portato a un numero crescente di uomini che non accedono a prestazioni sessuali, perché le donne sono naturalmente attratte dagli uomini dominanti.

---

mia vita sessuale" (William, 2018) e "27 ragazze attraenti che sono diventate brutte pazze a causa del femminismo" (Valizadeh, 2017).

<sup>11</sup> Bersaglio di campagne che affermavano che lui e i suoi seguaci promuovevano attivamente la cultura dello stupro, Roosh V è stato costretto a cancellare varie apparizioni in tutto il mondo a causa dell'impossibilità di garantire che gli incontri programmati potessero essere mantenuti privati e al sicuro da interferenze esterne. Roosh V ha in seguito dichiarato che il suo articolo che suggeriva la legalizzazione dello stupro era un'opera di satira e che lui è rigorosamente antistupro e anti-violenza contro le donne.

<sup>12</sup> Secondo la Teoria Red Pill, nel definire chi sia un Chad, si farebbe riferimento alla cosiddetta Teoria LMS, legata a 3 fattori principali: Look (L, Aspetto), Money (M, Denaro), e Status (S). (DeCook, 2021). Una versione più estrema della Red Pill, detta "Black Pill" (o "Pillola Nera") afferma che dalla condizione di incel non si possa uscire, e che l'unica cosa possibile da fare sia "lay down and rot", ovvero "sdraiarsi e marcire".





Queste argomentazioni sulla natura delle donne sono simili a quelle utilizzate dai pick-up artists: la differenza è che gli incel rifiutano l'argomentazione dei PUA secondo cui il carisma (l'abilità di gestire il "gioco") possa aumentare il loro status abbastanza da raggiungere il successo sessuale. Gli incel disprezzano invece il piccolo numero di Chad che nel mondo godono di un accesso illimitato alle "Stacy" promiscue - un gruppo, quest'ultimo, che considerano contemporaneamente troppo "troie" per essere degne di rispetto e troppo esigenti quando si tratta di partner sessuali<sup>13</sup>. Nei forum incel, la rabbia nei confronti della blogosfera PUA non è così comune come quella nei confronti delle donne come categoria, ma è un tema costante. Il forum incel *PUAhaters*, ad esempio, definiva la cultura PUA come un gioco di truffe effettuato da artisti della truffa, identificando l'industria della seduzione come "the scams, deception, and misleading marketing techniques used by dating gurus and the seduction community to deceive men and profit from" (Greig, 2014: 36). Ben presto, tuttavia, l'ostilità nei confronti delle donne che negano un presunto diritto all'intimità diventa più forte del sentimento di frustrazione nei confronti dei PUA, e il sito - chiuso nel 2014 - riapre con un nuovo nome: *Sluthate*. Perdendo fiducia nel gioco, se ne conserva tuttavia la fantasia patriarcale. Sebbene gli incel esistano come opposti quasi polari della comunità dei pick-up artists, "the result is the same: they view women's thoughts and feelings as irrelevant. They just want a vaguely woman-shaped thing for them to use" (Ricard, 2028: 3).

Nel 2009 George Sodini - che frequentava seminari di seduzione ed era un avido consumatore della pedagogia PUA - spara e uccide tre donne in una palestra in Pennsylvania; il suo blog racconta le sue giustificazioni per l'attacco: la mancanza di relazioni sessuali e romantiche, la rabbia verso le ragazze e le donne sessualmente attive e il rifiuto da parte di "30 milioni" di donne single (Roth, 2009). Il PUA Roissy scrisse sul suo blog che se Sodini avesse solo imparato il "gioco", le vite delle donne sarebbero state risparmiate (North, 2009). Due anni dopo, nel 2011, un altro PUA, Allen Robert Reyes uccide una donna che aveva rifiutato le sue avances. Nel 2014, il ventiduenne Elliot Rodger uccide sei persone (ferendone altre quattordici) a Isla Vista, in California.<sup>14</sup> Prima del massacro, Rodger aveva inviato via e-mail il suo

<sup>13</sup> Le donne sono spesso definite con un linguaggio avvilente, ad esempio "femoidi", abbreviazione di umanoidi femminili, è un termine comune nei forum incel, che indica che il non considerare le donne pienamente umane.

<sup>14</sup> Negli ultimi 10 anni, le uccisioni di massa "giustificate" dal rifiuto sessuale delle donne nei confronti degli uomini sono aumentate drammaticamente negli USA e in Canada (Bratich, Banet\_Weiser 2019). Anche in Europa la frequenza dei forum incel è in forte aumento (RAN, 2021), e alcuni incidenti mostrano legami con le ideologie di estrema destra, mescolando temi razzisti e convinzioni misogine. È il caso del 27enne Stephan Balliet che, nel 2018, si è diretto alla sinagoga del quartiere Paulusviertel di Halle (Germania) nel giorno sacro ebraico dello Yom Kippur, intenzionato a sparare alle persone in preghiera e uccidendo due persone. Sia nel video registrato in livestream durante l'attacco che nel suo manifesto, oltre alle teorie cospirative antisemite e al suprematismo bianco, Balliet ha usato termini di autodisprezzo, come "perdente", e ha mostrato forti opinioni antifemministe (Koehler, 2019). L'anno successivo, il quarantatreenne Tobias Rathjen ha compiuto due sparatorie in un bar di Hanau e in un bar di Kesselstadt, che hanno provocato la morte di nove persone e, una volta tornato a casa, ha sparato alla madre e poi si è ucciso. Il suo manifesto online conteneva opinioni razziste e xenofobe, oltre alla frustrazione per non aver mai avuto una relazione sentimentale con una donna (Jasser et al., 2020). Il 12 agosto 2021, il ventiduenne

"manifesto" di 107.000 parole a trentaquattro persone, dichiarandosi un incel, ma anche un "artista del pick-up frustrato", giustificando le sue azioni come una ritorsione contro le donne che lo rifiutavano. Avido lettore del forum PUAhate, il sito gli aveva confermato "many of the theories [he] had about how wicked and degenerate women really are, and [...] how bleak and cruel the world is due to the evilness of women". Nel suo manifesto, intitolato "La guerra alle donne", nel suo video finale, e nei post finali su PUAhate.com, sinteticamente esprime il passaggio da PUA a incel: "Se non posso avervi, ragazze, vi distruggerò" ("If we can't solve our problems, we must DESTROY our problems [...] One day incels will realize their true strength and numbers, and will overthrow this oppressive feminist system" (Glasstetter, 2014).<sup>15</sup>

#### 4. I PUA fra presente e futuro. Appunti per future ricerche "sul campo"



Nei primi anni del XXI secolo era più facile identificare chi fosse un PUA, mentre oggi la popolarità della comunità è molto più difficile da valutare e misurare, date le differenze di stile, di coinvolgimento del pubblico e di opinioni di autori e blogger.

Molti autori riferibili alla galassia PUA hanno iniziato discutendo di "tecniche di rimorchio", ma oggi gli spazi virtuali coprono una serie di argomenti, dall'educazione dei genitori alla politica. L'essentialismo biologico continua, tuttavia, ad essere presente negli spazi PUA - influenzato dalla percezione di un sesso biologico binario - nell'affermazione che l'ipergamia sia qualcosa di innato nelle donne, mentre gli uomini sono genericamente vittime di una gerarchia femminista.

Le pillole nere e rosse, oltre al determinismo, sono alcune delle traiettorie unificanti della manosfera. La comprensione della Red Pill theory ne è il cuore, un requisito se si vuole partecipare ai suoi spazi. Diventare *redpilled* richiede che un individuo accetti la fredda verità della misandria nella società occidentale; ogni uomo che aderisce a questa ideologia può scegliere di agire sulla base di questa conoscenza, aderendo a comunità come incel, MGTOW, PUA o gruppi affiliati agli attivisti per i diritti degli uomini. I gruppi PUA sostengono azioni di auto-miglioramento, ma promuovono anche una particolare ideologia antifemminista, in una sinergia fra neoliberalismo e patriarcato. Jack Bratich e Sarah Banet-Weiser descrivono il neoliberalismo all'intersezione tra l'orientamento imprenditoriale (autostima, individualizzazione, autogestione) e la fiducia nelle competenze (*self-help*, agenzie di formazione, figure pedagogiche), e i PUA sembrano oggi emergere nello

---

Jake Davison ha ucciso cinque persone, tra cui la madre e un bambino di tre anni, e poi si è suicidato a Plymouth (Regno Unito). A quanto pare, Davison era coinvolto nelle comunità Incel e aveva postato messaggi misogini online (The Week staff, 2021).

<sup>15</sup> Il massacro di Isla Vista del 2014 ha segnato la rapida crescita della mobilitazione misogina online, con un aumento vertiginoso dei numeri di iscritti ai forum suprematisti maschili. L'attacco ha preceduto di qualche mese il noto incidente del #Gamergate, una campagna di molestie che ha preso di mira sviluppatrici e recensore di videogiochi donne e femministe (Dewey, 2014a). Il #Gamergate ha portato all'attenzione dei media mainstream la crescita della mobilitazione misogina e razzista online, che nel 2016 avrebbe interagito con i movimenti alt-right, anche nel sostegno all'elezione di Donald Trump (Lees, 2016).

scenario della razionalità tecnologica capitalista e delle manifestazioni culturali della società dei consumi. Già O'Neill (2015) analizzava come la cultura dei campi di addestramento PUA fornisse agli uomini un modo per dare un senso alle forme sempre più commercializzate della sessualità (maschile e femminile) nella società.

Il suo resoconto etnografico sfida l'idea che i PUA siano "patetici, patologici o perversi (*an army of sleazebags, saddos and weirdos*)", ma mostra piuttosto come essi forniscano agli uomini comuni un contesto per diventare più assertivi, "not so much a deviation or departure from current social conventions as an extension and acceleration of existing cultural norms" (O'Neill, 2015:10). La pratica sociale e discorsiva dei PUA è influenzata dalle razionalità consumistiche, dagli ideali egemonici di bellezza femminile, e le stesse manifestazioni dell'intimità diventano prodotto discorsivo delle ideologie consumistiche della società occidentale tardo-moderna, in cui le razionalità neoliberali si estendono alla vita di coppia degli individui. L'uso delle tecnologie di seduzione promette inoltre ai pick-up artist una mobilità individuale verso l'alto nella gerarchia delle mascolinità, in una prassi omosociale che si confronta continuamente con lo sguardo maschile egemonico.

Il PUA esiste come soggetto maschile subordinato e quindi tecnologizza la seduzione per ottenere i dividendi sociali e sessuali di cui gode il soggetto maschile egemonico.<sup>16</sup> In questo scenario, negli ultimi anni i gruppi PUA (o che gravitano in generale nell'abito dell'"industria della seduzione") prediligono una narrazione che assegna importanza al "gioco interiore" *l'inner game* (sviluppo personale, auto-miglioramento), piuttosto che al "gioco esteriore" *l'outer game* (trucchi e tecniche). Questa modifica nella narrazione e nelle pratiche richiede analisi teoriche ed empiriche rinnovate, capaci di intercettare le nuove forme di mascolinità che si addensano al confine fra neoliberalismo e backlash patriarcale, in cui strutture gerarchiche del capitalismo e l'eteronormatività si rafforzano a vicenda. In particolare, è interessante analizzare come forme di mascolinità ibrida (Demetriou, 2001) emergano nella manosfera, in cui la mascolinità egemonica può prendere in prestito elementi o caratteristiche da altre mascolinità non egemoniche per continuare a garantire l'egemonia in un panorama in evoluzione.

Infine, ricerche recenti mostrano come oggi convivano due discorsi ideologici divergenti che promuovono entrambi il diritto degli uomini al potere sociale (Vingelli, 2019). Da un lato, la creazione di contenuti che utilizzano temi di aggressione esplicita nei confronti delle donne e alla riduzione delle donne a merci sessuali. Dall'altro lato, posizioni più sociopolitiche, che hanno l'obiettivo di ricollocare il



<sup>16</sup> Corey Wayne, nel suo libro *How To Be a 3% Man: Winning the Heart of the Woman of Your Dreams* (2014a), utilizza lo strumento delle sessioni di coaching che rispondono a ogni tipo di domanda, dal sedurre le donne e organizzare incontri occasionali, fino alle relazioni a lungo termine, al riaccendere l'attrazione con un ex o all'affrontare le pressioni della vita in generale, come pianificare i figli o gestire un periodo di disoccupazione. Essere nel 3% degli uomini si riferisce a quel modello egemonico che comprende e incarna ciò che serve per essere attraenti: "A man who is unapologetically himself. A man who knows what he wants, why he wants it and is determined to get it despite the risks and potential for failure, a man who marches to the beat of his own drum, a man who is a gentleman, strong, charming, chivalrous, humorous, determined and respectful; a man who will chose walking away verses staying in a relationship that requires him to sacrifice his hopes, goals and dreams to make a woman happy" (Wayne, 2014b)

discorso sulle nuove maschilità in contrapposizione alla misandria percepita della società contemporanea. Come altri gruppi della manosfera, anche i PUA hanno avviato strategie di professionalizzazione e di costruzione di un sapere esperto, allontanandosi dalle narrazioni più esplicitamente misogine e sessiste del passato.

In particolare, la manosfera contemporanea utilizza strategie discorsive e immaginari rinnovati, interpretando un sessismo *evidence-based* (Rothermel, 2023). In base alla quale teorie e metodi (statistiche, studi, notizie ecc.) di presentazione della conoscenza sono 'mimati' e stravolti per sostenere e legittimare opinioni essenzializzanti e polarizzanti sulle relazioni di genere nella società. Questa torsione è un aspetto fondamentale per le mobilitazioni contemporanee, perché fornisce un senso di autorità e forgia un'identità collettiva, inquadrata come "alternativa" alla conoscenza tradizionale, fungendo altresì da connettore tra la manosfera, il mainstream conservatore, discorsi di estrema destra e cospirazionisti.

In questo senso, la manosfera si pone come spazio distinto e "alternativo" per la produzione di conoscenza di genere, affiancandosi alle tradizionali comunità epistemiche - la chiesa, la comunità scientifica e i media - che producono conoscenza sul modo in cui il genere e le relazioni di genere sono definite, comprese e dotate di significato (Verloo, 2018). In questa logica, gli uomini sono presentati come le vittime finali dell'attuale clima sociale, mentre altri gruppi (in particolare le donne) sono considerati vincitrici inequivocabili delle attuali relazioni di genere. Questa narrazione vittimistica è stata identificata come un aspetto comune dei discorsi della destra, utilizzata per giustificare l'odio contro alcuni gruppi della società (Wilson, 2022).

La presentazione del sapere di genere "alternativo" così creato, in contrapposizione al sapere "mainstream" della società, rafforza l'identità collettiva di vittime e crea ulteriore legittimità per la creazione di spazi alternativi digitali di saperi di genere nella manosfera. Ad oggi, quasi tutte le comunità della manosfera hanno costruito una sorta di glossario o wiki, che imita i modi tradizionali di presentare la conoscenza e delinea i principi fondamentali della loro ideologia, fornendo un sistema di riferimento "alternativo" chiuso. Lo studio di questi archivi può consentire di analizzare l'epistemologia sociale dei gruppi, le loro fantasie condivise, le strategie discorsive e gli immaginari. In questi archivi, l'immagine delle gerarchie di genere tra uomini e donne è codificata in sistemi linguistici specifici che, sebbene differiscano a seconda del sottogruppo, si riferiscono l'uno all'altro in pratiche discorsive che riaffermano l'identità di gruppo e contemporaneamente l'appartenenza alla manosfera.



## Bibliography

- Banet-Weiser S., Miltner K. M. (2015), "#MasculinitySoFragile: culture, structure, and networked misogyny", *Feminist Media Studies*, 16(1): 171–174.
- Bratich J. Banet-Weiser S. (2019), "From pick-up artists to incels: con(fidence) games, networked misogyny, and the failure of neoliberalism", *International Journal of Communication*, 13: 5003-5027.
- Cameron D. (2009), "Sex/Gender, Language and the New Biologism", *Applied Linguistics*, 31(2): 173–192.
- Connell R.W. (2005), *Masculinities*, Cambridge, Polity Press (2<sup>nd</sup> edition).
- Dayter D., Rüdiger S. (2014). "Speak Your Mind, but Watch Your Mouth: Complaints in CouchSurfing References", in Bedijs K., Held G., Maaß C. (eds.), *Face Work and Social Media*, Zürich/Berlin: LIT, 193–212.
- Dayter D., Rüdiger S. (2016), "Reporting from the Field: The Narrative Reconstruction of Experience in Pick-Up Artist Online Communities", *Open Linguistics*, 2 (1): 337–351.
- Dayter D., Rüdiger S. (2019), "In Other Words: 'The Language of Attraction' Used by Pick-Up Artists", *English Today*, 35 (2): 13–19.
- Dayter D., Rüdiger S. (2020), "Talking About Women: Elicitation, Manual Tagging, and Semantic Tagging in a Study of Pick-Up Artists' Referential Strategies", in Dayter D., Rüdiger S. (eds), *Corpus Approaches to Social Media*, Amsterdam/Philadelphia, John Benjamins, 63–86.
- Dayter D., Rüdiger S. (2022), *The Language of Pick-Up Artists: Online Discourses of the Seduction Industry*, New York and Abingdon, Routledge.
- DeCook J.R. (2021), "Castration, the Archive, and the Incel Wiki", *Psychoanalysis, Culture & Society*, 26, 234–243.
- Demetriou D. Z. (2001), "Connell's Concept of Hegemonic Masculinity: A Critique", *Theory and Society*, 30 (3), 349.
- Denes A. (2011), "Biology as consent: Problematizing the scientific approach to seducing women's bodies", *Women's Studies International Forum*, 34(5): 411–419.
- Dewey C., (2014a), "The only guide to Gamergate you will ever need to read," *Washington Post*, 14 ottobre (<https://www.washingtonpost.com/news/the->



[intersect/wp/2014/10/14/the-only-guide-to-gamergate-you-will-ever-need-to-read/](https://www.washingtonpost.com/news/the-intersect/wp/2014/10/14/the-only-guide-to-gamergate-you-will-ever-need-to-read/))

Dewey C (2014b), 'Inside the 'manosfera' that inspired Santa Barbara shooter Elliot Rodger', *Washington Post*, 27 maggio, ([https://www.washingtonpost.com/news/the-intersect/wp/2014/05/27/inside-the-manosphere-that-inspired-santa-barbara-shooter-elliott-rodger/?noredirect=on&utm\\_term=.45117fb7100d](https://www.washingtonpost.com/news/the-intersect/wp/2014/05/27/inside-the-manosphere-that-inspired-santa-barbara-shooter-elliott-rodger/?noredirect=on&utm_term=.45117fb7100d)).

Dordoni A. Magaraggia S. (2021), "Modelli di mascolinità nei gruppi online Incel e Red Pill: narrazione vittimistica di sé, deumanizzazione e violenza contro le donne", *AG-ABOUT GENDER*, 10(19): 35-67.

Doyle S. (2018), "The deadly incel movement's absurd pop culture roots", *Medium*, 6 giugno, (<https://gen.medium.com/the-deadly-incel-movements-absurd-pop-culture-roots-e5bef93df2f5>).

Gibson M. (2014), "Is this the most hated man in the world?", *Time*, 12 novembre, (<http://time.com/3578387/julien-blanc-feminism-real-social-dynamics/>).

Gill R. (2009), "Mediated intimacy and postfeminism: a discourse analytic examination of sex and relationships advice in a women's magazine", *Discourse & Communication*, 3(4): 345-369.

Ging D. (2019), "Alphas, Betas, and Incels: Theorizing the Masculinities of the Manosphere", *Men and Masculinities*, 22 (4): 638-657.

Glasstetter J. (2014), "Shooting Suspect Elliot Rodger's Misogynistic Posts Point to Motive", *Southern Poverty Law Center* (<https://www.splcenter.org/hatewatch/2014/05/23/shooting-suspect-elliott-rodgers-misogynistic-posts-point-motive>).

Górska A. Kulicka K., Jemielniak D. (2022), "Men not going their own way: a thick big data analysis of #MGTOW and #Feminism Tweets", *Feminist Media Studies*, 23 (8): 3774-3792.

Greig A. (2014), "I will slaughter every single blonde s\*\*\* I see", *Daily Mail*, 25 maggio, <https://www.dailymail.co.uk/news/article-2638049/7-dead-drive-shooting-near-UC-Santa-Barbara.html>.

Guo C, Yin H. (2021), "Localized PUA Culture and Gender Politics in Contemporary China", *Journal of Gender Studies*, 30 (5): 609-620.

Hambling-Jones O., Merrison A. J. (2012), "Inequity in the Pursuit of Intimacy: An



Analysis of British Pick-Up Artist Interactions", *Journal of Pragmatics*, 44 (9): 1115–1127.

Hoinsky K. (2013), *Above the game: a guide to getting awesome with women*, CreateSpace Independent Publishing Platform

Jasser G., Kelly M., Rothermel A. (2020), *Male supremacism and the Hanau terrorist attack: between online misogyny and far-right violence*, The Hague: International Centre for Counter-Terrorism (ICCT) (<https://icct.nl/publication/male-supremacism-and-the-hanau-terrorist-attack-between-online-misogyny-and-far-right-violence/>).

Koehler D. (2019), "The Halle, Germany, Synagogue Attack and the Evolution of the Far-Right Terror Threat", *CTC Sentinel* [online], 12(11): 14–21 (<https://ctc.usma.edu/halle-germany-synagogue-attack-evolution-far-right-terror-threat/>).

Krendel A. (2020), "The Men and Women, Guys and Girls of the 'Mano-sphere': A Corpus-Assisted Discourse Approach", *Discourse & Society*, 31 (6): 607–630.

Lana Massi (2014), CNN interviews Julien Blanc (17 novembre), online video (<https://www.youtube.com/watch?v=wPT5dFHK3wo>).

Lees M. (2016), "What Gamergate should have taught us about the 'alt-right'", *The Guardian*, 1 dicembre (<https://www.theguardian.com/technology/2016/dec/01/gamergate-alt-right-hate-trump>).

Lin J. L. (2017), "Antifeminism online. MGTOW (men going their own way)", in Frömring U.U, Fox S., Köhn S., Terry M. (eds.), *Digital environments. Ethnographic perspectives across global online and offline spaces*, Bielefeld: Edition Medienwissenschaft, 77-96.

Lyons M. N. (2017), *Ctrl-Alt-Delete: The Origins and Ideology of the Alternative Right*, Somerville, MA: Political Research Associates ([www.politicalresearch.org/2017/01/20/ctrl-alt-delete-report-on-the-alternative-right/](http://www.politicalresearch.org/2017/01/20/ctrl-alt-delete-report-on-the-alternative-right/)).

Mamié R., Horta Ribeiro M., West R. (2021), "Are Anti-Feminist Communities Gateways to the Far Right? Evidence from Reddit and YouTube", arXiv:2102.12837v2.

Marwick A. E., Caplan R. (2018), "Drinking Male Tears: Language, the Manosphere, and Networked Harassment", *Feminist Media Studies*, 18 (4): 543–559.



Mystery, Odom C. (2007), *The Mystery Method: How to Get Beautiful Women Into Bed*. New York: St. Martin's Press.

North A. (2009), "Gunman murders gym-going women; Misogynists approve", *Jezebel*, 5 agosto, (<https://jezebel.com/5330560/gunman-murders-gym-going-women-misogynists-approve>).

O'Neill R. (2015), "The Work of Seduction: Intimacy and Subjectivity in the London 'Seduction Community'", *Sociological Research Online*, 20 (4): 172–185.

O'Neill R. (2017), "The Aesthetics of Sexual Discontent: Notes from the London 'Seduction Community'", in Elias A. S., Gill R., Scharff C. (eds), *Aesthetic Labour. Dynamics of Virtual Work*, London: Palgrave Macmillan, 333–349.

O'Neill R. (2018), *Seduction: Men, masculinity and mediated intimacy*, Cambridge: Polity Press.

Oesch N., Miklousic I. (2012), "The Dating Mind: Evolutionary Psychology and the Emerging Science of Human Courtship", *Evolutionary Psychology*, 10 (5): 427-440.

Osman S. L. (2013), "Token resistance to sex scale", in Fisher T., Davis C., Yarber W. (eds.), *Handbook of sexuality-related measures*, London, Routledge.

RAN (Radicalisation Awareness Network) (2021), *The Incel Phenomenon: Exploring Internal and External Issues Around Involuntary Celibates* ([https://home-affairs.ec.europa.eu/system/files/2021-08/ran\\_cn\\_incele\\_phenomenon\\_20210803\\_en.pdf](https://home-affairs.ec.europa.eu/system/files/2021-08/ran_cn_incele_phenomenon_20210803_en.pdf)).

Ricard A. (2018), "How sad boys get angry", *Medium*, 28 maggio, (<https://medium.com/@alexjricard/how-sad-boys-get-angry-8c4a2d781936>).

Roark H. (2018), 'Sweden is forcing its young boys into homosexuality', *Return of Kings*, 16 aprile, (<http://www.returnofkings.com/168668/sweden-is-forcing-its-young-boys-into-homosexuality>) (archiviato dall'originale il 18 ottobre 2018).

Roth L. M. (2009), "Woman-hating and the LA Fitness massacre", *Huffpost*, 7 settembre, ([https://www.huffpost.com/entry/woman-hating-and-the-la-f\\_b\\_254020](https://www.huffpost.com/entry/woman-hating-and-the-la-f_b_254020))

Rothermel A. K. (2023), "The role of evidence-based misogyny in antifeminist online communities of the 'manosfera'", *Big Data & Society*, 10 (1).

Rüdiger S., Dayter D. (2017), "The Ethics of Researching Unlikeable Subjects:



Language in an Online Community," *Applied Linguistics Review*, 8 (2/3): 251–269.

Rüdiger S., Dayter D. (2020), "Manbragging online: Self-praise on pick-up artists' forums", *Journal of Pragmatics*, 161(1): 16-27.

Scotto di Carlo G. (2023), "No Does Not Always Mean No": The Discursive Representation of Female Sexual Rejection as "Last Minute Resistance in Pick-Up Artist Communities", *International Journal of Communication*, 17: 5108-5127.

Schutzbach F. (2018), "Dominante Männlichkeit und neoreaktionäre Weltanschauungen in der Pick-Up-Artist-Szene", *Feministische Studien*, 36 (2): 305–321.

Schuurmans J., Monaghan L. F. (2015), "The Casanova-Myth: Legend and Anxiety in the Seduction Community", *Sociological Research Online*, 20 (1): 94–107.

Strauss N. (2005), *The game: Penetrating the secret society of pickup artists*, New York, ReganBooks.

The Week staff (2021), "Terror in Plymouth: what motivated Jake Davison?", *The Week* [online], 19 agosto, (<https://www.theweek.co.uk/news/uk-news/953864/terror-in-plymouth-what-motivated-jake-davison>).

Tranchese A., Sugiura L. (2021), "I don't hate all women, just those stuck-up bitches": How Incels and mainstream pornography speak the same extreme language of misogyny", *Violence Against Women*, 27(14): 2709-2734.

Valizadeh D. (2015a) "What Is Neomascularity?", *rooshv.com*, 6 maggio, archiviato dall'originale il 23 marzo 2022.

Valizadeh D. (2015b), "How to stop rape", *Return of Kings*, 16 febbraio, (<http://www.rooshv.com/how-to-stop-rape>), archiviato dall'originale il 18 aprile 2022.

Valizadeh D. (2017), "27 attractive girls who became ugly freaks because of feminism", *Return of Kings*, 14 febbraio (<http://www.returnofkings.com/114644/27-attractive-girls-who-became-ugly-freaks-because-of-feminism>), archiviato dall'originale il 23 marzo 2022.

Verloo M. (2018), "Gender knowledge, and opposition to the feminist project: Extreme-right populist parties in the Netherlands", *Politics and Governance*, 6 (3), 20–30.

Vingelli G. (2019), "Antifemminismo online. I Men's Rights activists in Italia", *Im@go. A*



Giovanna Vingelli  
*Manosfera e pick-up artists*

*journal of the social imaginary*, 12: 219-247.

Wayne C. (2014a), *How to be a 3% man: Winning the heart of the woman of your dreams*, Self-published eBook (<https://understandingrelationships.com/>).

Wayne C. (2014b), *What women really want*, online video, (<https://www.youtube.com/watch?v=RH-9ZVE6pj8>).

William J. (2018), "The #metoo movement has ruined my sex life", *Return of Kings*, 27 marzo (<http://www.returnofkings.com/162403/the-metoo-movement-has-ruined-my-sex-life>), archiviato online il 18 aprile 2022.

Wilson C. (2022), "Nostalgia, entitlement and victimhood: The synergy of white genocide and misogyny", *Terrorism and Political Violence*, 34 (8), 1810-1825.

Zimmer B. (2018), "How 'Incel' Got Hijacked", *Politico*, 8 maggio, ([www.politico.com/magazine/story/2018/05/08/intel-involuntary-celibate-movement-218324/](http://www.politico.com/magazine/story/2018/05/08/intel-involuntary-celibate-movement-218324/)).







A JOURNAL OF THE  
SOCIAL IMAGINARY

# Siamo l'esercito dei "bravi ragazzi". La politica identitaria della manosphere

Manolo Farci

manolo.farci@uniurb.it

Università degli Studi di Urbino



## Abstract

*We are the army of "nice guys". The identity politics of manosphere*

Historically, white men rarely questioned their masculinity, which enjoyed privileged invisibility. However, contemporary identity politics increasingly scrutinize masculinity, leading to conflicting reactions. Men feel targeted and demonized, yet also perceive invisibility as a disadvantage in a diverse, visibility-focused society. The rise of the manosphere reflects men's attempts to negotiate their identity within this politicized landscape. Through figures like the *nice guy*, men resist feminist politicization of masculinity while grappling with being both invisible victims and vilified embodiments of the *status quo*.

## Keywords

Masculinity | Nice Guy | Identity Politics | Social Network | Manosphere



## 1. Introduzione

In prospettiva storica, i ragazzi bianchi non hanno mai dovuto pensare a sé stessi come maschi. Questo non perché tali aspetti della loro identità fossero irrilevanti, ma poiché la maschilità ha sempre goduto di una prossimità privilegiata all'universale e al normativo, che l'ha resa invisibile a qualsiasi tipo di investigazione sulla sua natura. Che cosa significa essere uomo è una domanda che sarebbe stata impensabile solo qualche decennio fa: gli uomini sapevano talmente bene chi fossero che nessuno pensava di interrogarsi sull'identità maschile (Badinter, 1992).

Oggi, tuttavia, le cose stanno cambiando. La maschilità viene sempre più vista attraverso la lente delle politiche identitarie, considerata come incarnazione di una particolarità che casualmente coincide con il sé normativo e, dunque, non marcato. Questo meccanismo di identificazione provoca reazioni contrastanti negli uomini. Da un lato, avendo da sempre goduto del privilegio dell'invisibilità, molti si sentono incredibilmente danneggiati da una politica identitaria che li addita e li demonizza *essenzializzandoli*. Bollati con il pennello omogeneizzante della categorizzazione di genere e razziale, gli uomini stanno vivendo sulla propria pelle ciò che le donne e le altre etnie hanno sperimentato a lungo: essere costretti a significare il particolare e il materiale. Buona parte delle lamentele maschili di oggi, pertanto, nascono proprio dal tentativo di sfuggire alla politicizzazione delle proprie identità. Dall'altro lato, tuttavia, in un'epoca caratterizzata da una crescente consapevolezza e valorizzazione delle diversità (di genere, razza, orientamento sessuale, ecc.), quella stessa invisibilità, che per tanto tempo ha rappresentato un privilegio maschile, può essere adesso percepita come uno svantaggio e un peso. Gli uomini possono sentirsi esclusi o trascurati, proprio poiché avvertono di non avere una differenza visibile da evidenziare.

L'ipotesi di questo lavoro è che il recente fenomeno della manosphere – una variegata galassia di siti web, blog, comunità online che mettono al centro delle loro discussioni il ruolo dell'uomo nella società contemporanea e la difesa dei suoi interessi (Nagle, 2017) – rappresenti un tentativo complesso e spesso contraddittorio da parte degli uomini di negoziare la propria posizione all'interno del campo delle politiche identitarie contemporanee. Gli utenti che frequentano questi spazi sembrano dedicare tutti i loro sforzi a respingere la politicizzazione femminista della maschilità e a liberarsi dal fardello di un'identità connotata (spesso in termini di potere e privilegio). Tuttavia, nel far ciò, non possono fare a meno di adottare quelle stesse strategie e logiche che li collocano al centro delle battaglie culturali contro cui si oppongono, sfruttando tutto l'armamentario simbolico che deriva dall'occupare la posizione di soggetto in crisi. In altre parole, da un lato, questi uomini denunciano di essere stati *invisibili* dalla visibilità degli altri (in particolare donne e altre minoranze) e rivendicano una propria esperienza maschile specifica, spesso usando il linguaggio della vittima. Dall'altro, lamentano di essere stati resi *visibili* come la personificazione tanto diffamata dello *status quo* e, per questo, disconoscono la maschilità come quell'incarnazione di genere che condizionerebbe la loro identità. La figura del "bravo ragazzo" rappresenta un modo per venire a patti con tale



contraddizione. Come vedremo, essa permette ai ragazzi della manosphere di posizionarsi sia dentro che fuori dalle attuali lotte sulle politiche identitarie (Robinson, 2000).

## 2. Dal maschio in crisi al "bravo ragazzo"

La percezione di una crisi della maschilità non è un fenomeno nuovo. Essa rappresenta una condizione ricorrente nella storia (Kimmel, 1987) tanto da aver permeato e modellato le rappresentazioni degli uomini nella cultura popolare occidentale di questi ultimi decenni (Beynon, 2001). Tale percezione non solo si riverbera attraverso le epoche, ma si adatta anche alle mutevoli dinamiche sociali e culturali del tempo in cui si manifesta. Oggi il *cliché* del maschio in crisi popola gli spazi della manosphere, sebbene in forme differenti rispetto al passato recente. Nei decenni precedenti, la retorica della crisi maschile era impiegata dai tradizionali movimenti per i diritti degli uomini, composti prevalentemente da individui "tipicamente tra i quaranta e i cinquant'anni, spesso divorziati o separati e quasi sempre eterosessuali" (Flood, 2004: 263), per promuovere rivendicazioni di carattere politico o istituzionale. Attualmente, la maschilità in crisi è diventata un concetto in grado di mobilitare soprattutto i ragazzi e sfruttare le loro "ansie per il cambiamento degli standard di consenso e delle norme sessuali e di genere" (Gotell, Dutton, 2016: 76), avvenuti in questi ultimi anni.

Negli spazi della manosphere si assiste, così, ad un proliferare di discussioni su quali strategie e meccanismi nascosti guiderebbero i comportamenti femminili in ambito di appuntamenti, relazioni e attrazione verso l'altro sesso. Tali discorsi, spesso velati da un sottile strato di misoginia, si basano solitamente su un'interpretazione capziosa delle ricerche evoluzionistiche riguardanti le strategie naturali di accoppiamento femminile (Bachaud, Johns, 2023). Secondo questa lettura, le donne (così come altri mammiferi femminili) non sono uniformemente timide e passive come si potrebbe immaginare, ma possono, talvolta, essere altrettanto sessualmente assertive quanto i maschi. In particolare, sarebbero portate ad adottare nell'interesse dei propri geni una "doppia strategia di accoppiamento". Una prima strategia coinvolgerebbe la scelta di un partner stabile che possa fornire supporto e risorse a lungo termine, sia per la donna stessa che per la prole. Questo partner verrebbe scelto anche in base alla sua capacità di essere un buon genitore e un compagno affidabile. Una seconda strategia si concentrerebbe sull'ottenimento di geni di alta qualità per i propri figli. Questo avverrebbe attraverso relazioni extraconiugali o occasionali con uomini percepiti come geneticamente superiori o più attraenti. L'idea è che, mentre il partner stabile fornisce le risorse, il partner occasionale contribuisce con geni migliori, aumentando così la *fitness* biologica della prole. Un famoso meme che circola nella manosphere - "Alpha Fux, Beta Bux" - riassume bene questa visione: le donne tendono a rimanere in relazioni con uomini



poco attraenti ma stabili e premurosi che provvedono a loro (*beta*), mentre li tradiscono con maschi più attraenti (*alfa*).

L'ipotesi della doppia strategia di accoppiamento femminile informa gran parte dei ragionamenti che circolano nelle diverse comunità della manosphere: dagli artisti del rimorchio (O' Neill, 2018) alla *Red Pill* sino agli Incel. Ogni fenomeno sociale viene *ossessivamente* riportato a tale matrice concettuale. Il femminismo e l'antifemminismo, ad esempio, non sarebbero altro che due strategie riproduttive diametralmente contrapposte. Il femminismo è considerato un movimento che supporta le donne nel perseguire i propri obiettivi e impulsi geneticamente determinati, permettendogli di selezionare liberamente i partner, cambiarli quando desiderato, e ottenere risorse e vantaggi migliori. L'antifemminismo, d'altro canto, rappresenta la risposta maschile speculare: un tentativo di ribaltare le attuali regole del gioco, percepite come ingiustamente favorevoli alle donne, restituendo agli uomini eterosessuali maggiore controllo e potere nei rapporti sessuali e sentimentali con l'altro sesso (Van Valkenburgh, 2018).

Benché le analisi sull'accoppiamento femminile condivise nella manosphere provengano spesso da fonti scientifiche attendibili, esse vengono generalmente utilizzate in maniera superficiale, per fornire interpretazioni riduttive e deterministiche del comportamento umano, alterando sostanzialmente le intenzioni originarie delle teorie a cui fanno riferimento (Baselice, 2022). Uno dei principali limiti di queste interpretazioni è la tendenza a trascurare la *natura ipotetica* di tali scoperte. Ad esempio, la teoria della doppia strategia di accoppiamento femminile non viene considerata come un'ipotesi scientifica da verificare e discutere, ma presentata come un fatto accertato e incontestabile. Manca, inoltre, qualsiasi considerazione del contesto storico in cui queste teorie sono state sviluppate. Nella ricerca accademica, le spiegazioni evolutive si basano sull'osservazione di comportamenti aggregati sviluppati nel corso dei millenni. Questa prospettiva a lungo termine è essenziale per comprendere le radici biologiche delle condotte contemporanee. Al contrario, nella manosphere, le teorie evolutive vengono sfruttate per interpretare episodi o aneddoti tratti dalla vita quotidiana, senza la profondità temporale necessaria per discussioni evolutive significative. Questo porta a una sovrapposizione tra cause ultime e cause prossime. In psicologia evolutiva, le cause ultime spiegano il motivo per cui un comportamento si è sviluppato nel corso dell'evoluzione e analizzano le ragioni biologiche a lungo termine, come la sopravvivenza e la riproduzione, che hanno reso quel comportamento *vantaggioso* per gli individui. Le cause prossime, invece, si riferiscono ai meccanismi immediati che scatenano una specifica condotta, e possono includere fattori fisiologici, neurologici, psicologici e sociali che influenzano direttamente come e quando tale condotta si manifesta. In sintesi, le cause ultime spiegano il "perché" evolutivo di un particolare modo di agire, mentre le cause prossime chiariscono il "come" immediato. Nella manosphere, questa distinzione è spesso trascurata o mal interpretata: fenomeni sociali e comportamentali complessi sono ridotti a una singola causa *ultima* evolutiva, come i fattori genetici ereditari che guiderebbero le strategie riproduttive di uomini e donne. Educazione, norme sociali e circostanze personali non hanno alcuna



influenza significativa nello spiegare o modificare tali comportamenti. Uomini e donne agiscono in modi specifici perché "così sono stati programmati dalla natura".

Esiste, tuttavia, un limite ancora più problematico nel modo in cui la psicologia evolutivista viene applicata all'interno della manosphere: la tendenza a focalizzarsi quasi esclusivamente sui modi di agire femminili, trascurando le basi evolutive della sessualità maschile. Questa omissione conduce ad una rappresentazione distorta e caricaturale dei meccanismi dell'accoppiamento umano secondo i quali sarebbe *esclusiva responsabilità* della donna valutare i propri potenziali partner, basandosi su quegli attributi fisici e sociali (spesso riassunti nella formula *Look, Money, Status*), che le permetterebbero di massimizzare la qualità genetica e le risorse disponibili per la propria prole. Questa prospettiva unilaterale non solo ignora le complesse pressioni personali, sociali e culturali che entrano in gioco nella decisione femminile, ma omette qualsiasi considerazione sulle intenzioni del potenziale compagno di intraprendere una relazione. Concentrandosi unicamente sul comportamento femminile, e omettendo qualsiasi riferimento alla sessualità maschile (verso la quale la psicologia evolutiva non è certamente maggiormente lusinghiera), la manosphere finisce per legittimare un racconto della realtà di chiara impronta conservatrice e sessista, dove le donne sono ritratte come animali strategici, guidati da intenti moralmente deprecabili. Non è raro che i comportamenti femminili vengano paragonati a quelli delle scimmie: così come queste ultime praticano il *branch-swinging*, ovvero non lasciano un ramo senza prima afferrarne un altro, allo stesso modo si sostiene che le donne non lascino una relazione finché non hanno già consolidato un nuovo rapporto. Le donne considerate attraenti raramente rimangono single; essendo per natura *iper-gamiche*, cercano sempre di migliorare la loro situazione sentimentale passando a un uomo che percepiscono come di maggior valore, sia esso economico, sociale o fisico. L'antico doppio standard sessuale si riproduce in queste discussioni apparentemente scientifiche: per gli uomini, l'inclinazione all'accoppiamento a breve termine e alle relazioni extraconiugali è considerata così evidente e non problematica da non richiedere alcun commento o spiegazione, mentre per le donne rappresenta una questione talmente rilevante da giustificare condanne severe.

Infuse di emotività e giudizi morali, le teorie evolutive vengono mobilitate con uno scopo ben preciso: aggiornare l'immagine del maschio in crisi alla figura del "bravo ragazzo", decisamente più in linea con le esigenze e i valori delle nuove generazioni che frequentano questi spazi (Farci 2022). Lo stereotipo del *nice guy* deriva dalla convinzione diffusa secondo la quale, nelle relazioni romantiche, "i bravi ragazzi finiscono ultimi". Secondo tale credenza, le donne tenderebbero a manifestare atteggiamenti e comportamenti contraddittori nella scelta dei loro partner sentimentali. Nonostante spesso si descrivano alla ricerca di uomini gentili, sensibili ed emotivamente aperti, quando realmente messe di fronte a una preferenza tra un "bravo ragazzo" e un uomo scortese, insensibile e chiuso, tipicamente etichettato come "macho" o "stronzo", esse rifiuterebbero sistematicamente il "bravo ragazzo", preferendo il suo rivale più virile. Come ha sostenuto Warren Farrell (1986), le donne "prediligono gli eroi pur dichiarando di desiderare la vulnerabilità" (p. 351).





In realtà, l'idea che "i bravi ragazzi arrivano ultimi" non è supportata dai dati di ricerca. Uno studio influente condotto in tal senso (Buss, Barnes 1986) ha dimostrato che qualità come gentilezza, comprensione e onestà sono molto apprezzate nella scelta di un partner, sia da uomini che da donne. Altri studi hanno confermato questa conclusione, sottolineando come tali doti siano particolarmente importanti nelle preferenze romantiche, soprattutto tra le donne (Buss, Angleitner, 1989; Goodwin, 1990; Hatfield, Sprecher, 1995, Doosje et al., 1999). Ulteriori ricerche hanno rivelato che uomini che mostrano tratti comunemente considerati più "femminili", come la sensibilità e il garbo, sono spesso preferiti per relazioni serie e a lungo termine (Desrochers, 1995; Regan, Berscheid, 1997; Regan, 1998). In linea generale, qualsiasi spiegazione univariata dell'attrattività non è mai sufficiente: non esiste una risposta semplice alla domanda se i "bravi ragazzi" arrivino ultimi, in quanto la scelta di un partner può dipendere da numerosi fattori, come le caratteristiche personali degli individui e il contesto della relazione. Per esempio, uno studio (Herold, Milhausen, 1999) ha trovato che le donne che attribuiscono meno importanza al sesso e che hanno avuto meno partner sessuali in passato tendono a preferire ragazzi gentili, sensibili e rispettosi. Un'altra ricerca (Urbaniak, Kilmann, 2003) ha confermato che i "bravi ragazzi" finiscono per avere meno partner sessuali, ma sono spesso scelti per relazioni serie e a lungo termine. Al contrario, i ragazzi che mostrano comportamenti dominanti aumentano la propria attrattività sessuale agli occhi delle donne (Ahmetoglu Swami, 2012), anche se sono percepiti come meno simpatici e non desiderabili come coniugi (Sadalla et al., 1987). Secondo un'altra ricerca, la dominanza maschile da sola non aumenta l'attrazione delle donne: l'interesse femminile per gli uomini cresce solo quando tale dominanza è accompagnata da comportamenti prosociali, cooperativi e altruistici (Jensen-Campbell et al., 1995). Infine, un'ulteriore indagine (McDaniel, 2005) ha confermato che molte giovani donne riconoscono e apprezzano le qualità desiderabili di un "bravo ragazzo". Tuttavia, instaurare una relazione seria con un *nice guy* richiederebbe un'attenta valutazione delle sue caratteristiche, la capacità di apprezzarne i tratti positivi e una solida fiducia nella propria abilità di giudizio. Per tutte quelle ragazze che non sono pronte ad impegnarsi in un rapporto duraturo, ma vogliono semplicemente divertirsi con incontri meno vincolanti, questo processo di riflessione approfondita può risultare impegnativo. Spesso, può mancare loro il tempo, la pazienza o la predisposizione a dedicarsi a un approccio così meticoloso nella scelta del partner.

Un altro aspetto che complica la narrazione del "bravo ragazzo" riguarda la definizione fumosa che si porta dietro questo termine. Potrebbe essere che le donne che rifiutano gli uomini con il pretesto che siano "troppo gentili" possano in realtà usare questa espressione come un eufemismo, consciamente o inconsciamente, per "troppo noiosi". Sebbene essere gentili rappresenti effettivamente un fattore molto apprezzato nelle decisioni delle donne, certamente non costituisce l'unico criterio di valutazione. È probabile che le donne rifiutino anche uomini gentili se questi non soddisfano ulteriori aspettative, come avere interessi simili o possedere altre qualità personali eccitanti o interessanti. Ma questa condotta non deriva da una presunta

ipergamia femminile (la propensione delle donne a cercare uomini di status superiore). In realtà, è normale che le persone tendano a cercare partner che siano simili a loro stessi in termini di qualità e caratteristiche desiderabili, come il livello di intelligenza, l'aspetto fisico, la personalità e lo status socioeconomico. Tale comportamento è spiegato dal desiderio di stabilità e compatibilità a lungo termine nella relazione, dove coppie simili hanno maggiori probabilità di condividere valori, obiettivi e stili di vita compatibili (Gangestad, Simpson, 2000). Gli individui preferiscono essere in una relazione a lungo termine con altri che hanno un valore di accoppiamento approssimativamente uguale al proprio. La sola gentilezza potrebbe non essere sufficiente in un contesto di relazioni romantiche, laddove tale gentilezza non riuscirebbe a compensare altre carenze. In questo senso, l'affermazione che "i bravi ragazzi arrivano ultimi" potrebbe effettivamente essere un modo improprio di dire che "i ragazzi noiosi e per me poco interessanti arrivano ultimi, indipendentemente dalla loro gentilezza".

A questo si aggiunga un ulteriore elemento: spesso la cortesia dei bravi ragazzi può, in realtà, essere avvertita come poco autentica, in quanto nasconderebbe aspettative non espresse e bisogni repressi. Molti *nice guy* interagiscono con le donne come se fossero guidati da una sorta di "contratto occulto" (Glover, 2000), ossia una serie di credenze e comportamenti disfunzionali basati sulla convinzione errata che, se saranno "bravi", generosi e faranno favori alle ragazze, verranno amati, otterranno ciò che desiderano e avranno una vita senza problemi. Quando i *nice guy* si rendono conto che il loro supposto diritto all'affetto non viene rispettato, possono provare una serie di emozioni negative: ansia, dubbio verso sé stessi, ma soprattutto rabbia e ostilità nei confronti delle donne, percepite come responsabili della loro mancanza di gratificazione. In questo contesto, persino la misoginia può essere vista quale risposta legittima a una situazione percepita come ingiusta nei loro confronti.



### 3. Le ferite del "bravo ragazzo"

Nelle discussioni sul bravo ragazzo all'interno della manosphere, le spiegazioni evolutive vengono impiegate per promuovere una visione monolitica e inflessibile del "comportamento biologico" femminile, ma raramente sono applicate alla condotta degli uomini. Questo meccanismo argomentativo permette al maschile di liberarsi dal fardello di un'identità connotata, di *invisibilizzarsi* dietro il linguaggio "oggettivo e disincarnato" della scienza. Purtroppo, come abbiamo già detto, al giorno d'oggi la maschilità non può porsi completamente al di fuori delle dinamiche della politica identitaria. Anche gli uomini sentono la necessità di dotarsi di una propria visibilità incarnata. Per ottenere questo riconoscimento, devono mostrare di essere anche loro vittime di qualche forma di discriminazione o ingiustizia per ottenere l'autorità culturale solitamente conferita a coloro autenticamente *disempowered*. La figura del "bravo ragazzo" diventa uno strumento chiave in questo processo. I "bravi ragazzi" possono identificarsi come una categoria specifica basata sul genere, che si mobilita intorno a una discriminazione percepita. Tuttavia, per

rendere questa mobilitazione efficace, non basta utilizzare le spiegazioni semplicistiche delle teorie evolutive, che spesso riducono il comportamento umano a meccanismi biologici rigidi. È necessario costruire una narrazione più convincente e valida riguardo alle difficoltà che i "bravi ragazzi" affrontano e alle ingiustizie che percepiscono nella vita quotidiana.

Tale narrazione è quella *liberazionista*. Negli spazi della manosphere, i ragazzi si ritrovano a discutere delle loro esperienze e difficoltà che affrontano come maschi spesso recuperando linguaggio e terminologie dei *vecchi* movimenti di liberazione maschile degli anni Settanta (White, 2019). Questi movimenti erano guidati da un principio fondamentale: gli uomini sono danneggiati dai ruoli di genere tanto quanto le donne e, nonostante possano detenere più potere istituzionale, soffrono nel cercare di aderire a standard maschili imposti dalla società, che limitano la loro libertà personale e psicologica (Pleck, Sawyer, 1981). Gli uomini godono certamente di prestigio sociale e di molte forme di privilegio in virtù del loro essere maschi, "ma il modo in cui abbiamo strutturato quel mondo di potere causa loro dolore, isolamento e alienazione" (Kaufman, 1994: 142). Dal momento che i ruoli di genere sono costruiti e appresi attraverso la socializzazione, i liberazionisti auspicano che gli uomini possano sfidare e decostruire le norme sociali che perpetuano i miti della maschilità tradizionale. Indubbiamente, l'idea che il genere maschile sia un ruolo socialmente appreso suona progressista, capace di rivolgersi efficacemente ad un pubblico ben specifico: giovani uomini (ma anche donne), istruiti e ben educati che non aspirano certamente ad apparire come rozzi misogini tradizionalisti e conservatori (Messner, 2016). Paradossalmente, sembra approdare alle medesime conclusioni delle visioni femministe: anche gli uomini devono *liberarsi* dai preconcetti sociali che associano la maschilità alla forza fisica, all'invulnerabilità, al successo economico e all'incapacità di esprimere apertamente le proprie emozioni (David, Brannon 1976). Tuttavia, c'è una chiara differenza di intenti tra le due prospettive. Mentre per le femministe i ruoli tradizionali devono essere rifiutati perché garantiscono un potere maschile sulle donne, al contrario, per i ragazzi della manosphere essi vanno dismessi perché danno alle donne maggiore controllo sugli uomini, costringendoli a aderire a ruoli di genere per loro psicologicamente letali (Clatterbaugh, 1997). È qui che l'ideologia liberazionista riesce a conciliarsi bene con le teorie evolutive. Preferendo *naturalmente* i maschi sicuri e dominanti, le donne sarebbero complici nell'alimentare un senso di inadeguatezza e frustrazione nei "bravi ragazzi". Così, anche se la società dice loro che possono essere gentili, vulnerabili e meno *tossici*, i ragazzi della manosphere sanno che, in realtà, le aspettative femminili continuano a preferirli forti, aggressivi e audaci (Kimmel, 2008). Per quanto essi dimostrino di saper prendere le distanze da un'idea che percepiscono anacronistica di maschio duro e puro, sono consapevoli che questa devirilizzazione non corrisponde a quello che i meccanismi evolutivi della selettività femminile istintivamente cercherebbero in loro.

La figura del *nice guy* rifiuta quindi la maschilità tradizionale, per recuperarla, in realtà, come *postura vittimizzante* attraverso cui affrontare il sentimento di sconfitta nei confronti del potere di scelta sessuale delle donne (Messner, 2000). In tal senso,



sfrutta la medesima idea di *maschilità fallita* che caratterizza altre figure radicate nell'immaginario delle culture digitali, come il nerd o il geek (Farci, Scarcelli, 2022).

Difatti, il discorso del "bravo ragazzo" può essere visto come un riadattamento del tropo del "nerd sessualmente inattivo", che da sempre rappresenta l'archetipo dell'appassionato medio di cultura informatica (Salter, Blodgett, 2017). Nella cultura occidentale, i nerd e i geek sono solitamente descritti come i classici *maschi beta*: ragazzi fisicamente deboli, poco attraenti e mal vestiti; incapaci di avere una relazione reale; che si rivolgono agli ambienti disincarnati dei sistemi informatici, fatti di regole proprie chiare e oggettive a cui obbedire, per fuggire alle richieste imprevedibili e disordinate del mondo fisico, soprattutto quelle che riguardano la sessualità (Kendall, 2002). Al pari di questi, il *nice guy* si percepisce come un *outsider*, escluso dai privilegi associati ai tradizionali marcatori di successo della maschilità, non abbastanza duro o fisicamente prestante per essere uno di quei *maschi alpha* potenti che, nei suoi racconti, attirerebbero di più le ragazze.

Non potendo più fare affidamento sulle marcature normative della maschilità egemone, la figura del "bravo ragazzo" può rappresentarsi nel modo più completo e convincente esclusivamente assumendo lo status di vittima (Banet-Weiser, Miltner 2016). La rappresentazione del "maschio bianco come vittima" (Savran 1998) non è certamente una novità dei media digitali. Essa rappresenta una delle più pervasive figure dell'immaginario collettivo occidentale che è andata emergendo in risposta ai cambiamenti socioculturali del Secondo dopoguerra. Sebbene i primi semi di rabbia e insicurezza tra i maschi bianchi vengono gettati negli anni Cinquanta, è soprattutto negli anni Settanta che il tema del maschio bianco *in crisi* diventerà egemonico nella cultura americana e produrrà diverse varianti: il *new men* sensibile che prova ad adattarsi a nuove definizioni di maschilità, il maschio in cerca della sua essenza archetipica, l'*angry white male* che reagisce con rabbia e ostilità alla perdita dei propri privilegi. Benché profondamente differenti, tutte queste interpretazioni del maschile sembrano condividere un elemento comune: si basano su ciò che Freud ha definito come "sadosomochismo riflessivo", una condizione in cui l'ego viene astutamente diviso tra una metà masochistica che non si vergogna a mostrarsi soggiogata, sotto assedio, persino abusata e martirizzata, e una parte sadica che si delizia nell'esibire la propria capacità di sopportare le ferite, la propria implacabile durezza, in definitiva la sua virilità.

Al pari del soggetto freudiano, il *nice guy* non si vergogna di esibire le sue ferite psicologiche e sociali. Mostra con orgoglio la propria identità dolorosamente incarnata, che attinge pesantemente al potere morale della sofferenza, e sfrutta lo stesso vocabolario di dolore e urgenza che caratterizza i gruppi minoritari e oppressi.

Negli spazi della manosphere abbondano racconti di adolescenti che dicono di essere stati "trattati male" dalle donne, "ridicolizzati" o considerati "deboli", usati solo come amici su cui piangere, senza che nessuno si preoccupi di quello che loro stessi desiderano o sentono. Al pari delle narrazioni semplificate della psicologia evolucionista, questi racconti tendono alla semplificazione, alla totalizzazione, alla chiusura, si rifiutano di introdurre al loro interno qualsiasi elemento di contraddizione, complessità, ambiguità. Gli altri utenti leggono queste storie prive di



qualsiasi chiaroscuro, si infuriano anche essi, fino a sentire il battito cardiaco accelerare e le guance arrossarsi, mentre picchiano furiosamente sui tasti della tastiera per ottenere il sostegno o la conferma delle loro opinioni da parte di altri commentatori antifemministi che provano reazioni simili.

Non è importante che questo risentimento sia collettivamente strutturato, né diretto sempre in risposta a una persona specifica. Esso è efficace perché riesce a mobilitare il linguaggio delle emozioni e a definire un racconto della crisi da cui i "bravi ragazzi" possano emergere come vittime della violenza perpetrata dalle donne e dalla tirannia femminista. In questo contesto, le manifestazioni emotive diventano forme di rivendicazione straordinariamente potenti proprio perché non possono essere negate: i "bravi ragazzi" non hanno l'onere di dimostrare la veridicità delle loro sofferenze, poiché il peso dei loro sentimenti prevale sull'attendibilità dei fatti che li hanno fatti sentire in quel modo. Ogni loro affermazione diventa dunque incontestabile proprio perché gli affetti non possono essere analizzati in modo razionale, ma solo vissuti: "le mie esperienze sono le mie esperienze e non puoi negarle perché io credo che siano vere" (Allan, 2016: 7). Come spiega ContraPoints, famosa YouTuber americana con un dottorato in filosofia, molti utenti della manosphere condividono un'epistemologia masochistica: credono che tutto ciò che li fa soffrire meriti di essere considerato vero. Non importa, ad esempio, che i dati italiani dimostrino che gli uomini conducano ancora una vita sessualmente più attiva delle donne (Spallaci, 2019). Frequentando gli spazi della manosphere italiana ci si ritrova catapultati in un incubo distopico che ricorda i romanzi di Michel Houellebecq, dove il sesso è onnipresente, ma i maschi sono condannati a una pauperizzazione sessuale forzata.

Se i reclami di vittimismo e svantaggio maschile sono sempre stati parte integrante dei movimenti maschili sin dalla loro nascita (Menziès 2008), mai come adesso i linguaggi dell'esperienza vissuta e della sofferenza personale vengono sfruttati per rispondere alle proprie ansie di consolidamento identitario. In questo contesto, il ruolo dei media digitali assume un'importanza centrale. Le emozioni condivise nella manosphere si muovono come pratiche affettive *discorsivamente* costruite. Esse non solo accrescono il proprio impatto emotivo man mano che la loro diffusione aumenta, ma permettono di materializzare nuove forme di appartenenza collettiva attorno a specifiche risonanze emotive le quali, a loro volta, servono a rafforzare specifici punti di vista (de Boise, 2017). Esse definiscono, così, quelle che potremmo chiamare come "comunità affettive" (Farci, 2019) o comunità *senza* comunità (Dean, 2010), ossia formazioni che, mancando dei prerequisiti delle tradizionali comunità online – come la presenza di un unico luogo in cui radunarsi o l'appartenenza continuativa nel tempo – si definiscono esclusivamente a partire da una forma di coinvolgimento e intimità affettiva tra soggetti e un *altro immaginato*. Questo altro immaginato non solo diventa l'oggetto peculiare degli investimenti affettivi, ma attiva quel movimento di reclutamento che produce il soggetto emotivo stesso, nel nostro caso il *nice guy*. Lungi dall'essere semplicemente sfoghi di rabbia o frustrazione motivati da sentimenti di "diritto leso" (Kimmel, 2017), gli appelli alle emozioni diventano, così, lo strumento chiave per rafforzare la distinzione tra un noi



– i “bravi ragazzi”, e un loro – le donne, dove i loro vengono portati nella sfera di esistenza del noi esclusivamente come una minaccia, mentre il noi si costituisce proprio a partire da questo sentimento di comunanza per ciò che sente minacciato (Ahmed, 2004).

Sfruttando il linguaggio dell'esperienza e della sofferenza personale, il *nice guy* scalza le femministe dalla loro posizione di autorità personale e politica per assumere lo status moralmente inespugnabile di vittima. In un tempo in cui tutte le identità sono in crisi, essere vittima dà luogo a un supplemento di sé: “essere vittime dà prestigio, impone ascolto, promette e promuove riconoscimento, attiva un potente generatore di identità, diritto, autostima” (Giglioli 2014: 6). La posizione vittimaria non risponde di nulla, non ha bisogno di giustificarsi. Esiste esclusivamente come funzione di ciò che ha subito, di ciò che ha perduto, di ciò che le è stato tolto. Questo la immunizza da qualsiasi critica, le garantisce innocenza al di là di ogni ragionevole dubbio: una vittima non ha debiti, solo crediti.



#### 4. Conclusione

La figura del *nice guy* dimostra la relazione complicata che la maschilità contemporanea ha con una politica identitaria finora praticata prevalentemente contro di essa. Da una parte, la retorica del “bravo ragazzo” garantisce un modo per svincolarsi da quella marcatura di potere e privilegio che sembra associare i giovani d'oggi ad una singola categoria oppressiva (i famigerati maschi bianchi etero *basic*).

Per recuperare questa sorta di invisibilità di genere, i *nice guy* applicano, in modo capzioso, il linguaggio apparentemente neutro e disinteressato della psicologia evolutivista ai comportamenti femminili, trascurando invece qualsiasi accenno alle radici biologiche della loro stessa sessualità. Questo atteggiamento trasforma l'appello ai fatti della scienza in una forma di *disincorporazione* che permette ai maschi di svincolarsi da quella legge del corpo che viene rigorosamente invocata quando si parla di donne. La maschilità può così recuperare il proprio privilegio come simbolo di una verità universalista e disincarnata e farsi immune alla marcatura che accompagna il discorso sulle identità di genere.

Dall'altra parte, il *nice guy* agisce in modo totalmente opposto. Mette in scena una crisi in cui il trauma sociale inflitto all'autorità culturale della maschilità viene dislocato proprio su un piano più personale, sessuale e corporeo. Il “bravo ragazzo” esibisce una maschilità *debole* in balia del femminile, fisicamente ferita ed emotivamente traumatizzata. Offre agli adolescenti un modo per identificarsi anche essi come categoria specifica e particolare: in un mondo di definizioni razziali e di genere, non siamo anche noi una razza e un genere? Come il soggetto emarginato o minoritario che è prima bollato dalla cultura dominante, e poi si appropria di quella marcatura per i suoi scopi, il *nice guy* sperimenta anche lui un etichettamento imposto da altri (la maschilità come privilegio) e poi impara a ristabilire la sua identità come marcato. Utilizzando la retorica liberazionista, i bravi ragazzi possono, così, rappresentarsi come “vittime virtuali in cerca di uguaglianza” (Schmitz, Kazyak,

2016), che si appropriano del linguaggio tipicamente associato al femminismo per attribuire legittimità sociale alle loro esperienze di vulnerabilità e marginalizzazione sociale.

Questo meccanismo di appropriazione tattica non significa, ovviamente, che le ferite evidenziate da questi ragazzi non siano reali. Certi gruppi di attivisti per i diritti maschili hanno probabilmente ragione nel sostenere che gli uomini affrontano problemi che vanno oltre quelli psicologici o emotivi ("agli uomini non è concesso piangere") e includono cose come morire in guerra, abusare di alcol e droghe, avere tassi più alti di suicidio. Non è solo una specifica ideologia della maschilità tossica a influire sulla vita degli uomini – spingendoli, ad esempio, a lavorare duramente e ad assumersi rischi pericolosi, anche se ciò può danneggiare la loro salute (Botto et al., 2022). È anche la loro posizione nella società, sia essa nel mondo del lavoro, nella sfera pubblica o all'interno delle dinamiche familiari a produrre effetti materiali nelle loro vite, che sono *sistematicamente oppressivi*, anche se manca un beneficiario evidente di tale oppressione (New, 2001). Indefinitiva, non tutto ciò che viene discusso in questi spazi manca di fondamento. Eppure, questi ragionamenti trascurano spesso un elemento essenziale: i costi di essere un uomo sono legati al potere istituzionale degli uomini (Messner, 1998) e possono essere superati solo attraverso una trasformazione radicale dell'ordine di genere in alleanza con le donne e non in opposizione ad esse. I *nice guy* rifiutano, invece, qualsiasi riflessione sul radicamento sociale e istituzionale del potere maschile. Preferiscono concentrarsi esclusivamente sulle ferite psicologiche e sociali patite dal cosiddetto "sesso sacrificabile" (Farrell, 1993). Sostituiscono ogni tipo di analisi strutturale, storica e ideologica su potere e il privilegio con un confronto costante su chi tra i due sessi soffre di più o se la passi peggio. Questa mancanza di comprensione del quadro più ampio delle disuguaglianze di genere li porta spesso a confondere la linea di demarcazione tra questioni personali e politiche. In molti casi, ciò che appartiene alla sfera "personale", come le esperienze individuali di rifiuto o insuccesso in ambito relazionale, viene reinterpreto come una questione "politica". Allo stesso modo, ciò che è "politico", come ad esempio la critica femminista del privilegio maschile, viene trasformato in un fatto "personale", percepito come un attacco diretto all'identità e al valore individuale degli uomini. Di conseguenza, concetti come oppressione e discriminazione si riducono spesso a mere categorie retoriche, usate principalmente per alimentare dibattiti infuocati su quale genere subirebbe davvero maggiori disparità sociali e posizionare, ancora una volta, gli uomini *contro* le donne.

Alla luce di quanto detto, risulta difficile immaginare che la figura del *nice guy* possa contribuire ad un avanzamento in termini di parità di genere. Molti "bravi ragazzi" che frequentano la mansphere disprezzano più o meno velatamente le donne, credono che le violenze maschili siano un crimine inventato, che le false accuse di stupro abbiano raggiunto proporzioni epidemiche, o che il femminismo rappresenti un'ideologia d'odio simile al nazismo. Molte ricerche hanno dimostrato come la semplice esposizione ripetuta a questo tipo di argomentazioni, anche quando avviene in modo passivo o occasionale, rischia di pregiudicare la salute mentale e fisica dei ragazzi (Tietjen, Tirkkonen, 2023), compromettere



significativamente le loro relazioni interpersonali (Costello et al., 2023), spingerli ad abbracciare posizioni misogine e sessiste (Farci, Righetti, 2019; Preston et al., 2021), o adottare visioni del mondo reazionarie, facendole passare non solo come plausibili, ma addirittura come preferibili (Hawley, 2017; Wendling, 2018).

Nonostante ciò, la figura del *nice guy* può offrire un'importante chiave di lettura per capire come la maschilità definisca sé stessa in epoca di politiche identitarie.

Come sappiamo, gli uomini non fondono volentieri le loro identità individuali in un'identità di gruppo se non attorno a percepite perdite di potere. Nel caso del *nice guy* questa perdita di potere è misurata sempre in relazione al presunto vantaggio delle donne. Il potere femminile viene sfruttato per costruire l'immagine di una maschilità pericolosamente sotto assedio: una sorta di trucco retorico attraverso cui gli uomini, presentandosi come vittime di una presunta oppressione femminile, cercano di consolidare la loro autorità morale e giustificare il mantenimento del sistema patriarcale (Savran, 1998). Benché il potere sessuale femminile sia tatticamente sfruttato in chiave vittimistica, è pur vero che è proprio in rapporto a questo potere che molti ragazzi misurano la propria esperienza maschile, presunto o reale che sia. Poiché la maschilità è una costruzione sociale, è improbabile che possa essere creata e mantenuta esclusivamente dagli uomini. Le espressioni della maschilità sono modellate *anche* dai desideri e dalle attese, anch'esse socialmente costruite, delle donne. Sebbene non si possa dire che le donne controllino direttamente gli uomini, il loro ascendente sul comportamento maschile è comunque significativo. È dimostrato che le donne non sono sempre *spettatrici passivi* nel consolidamento della maschilità egemonica (Connell 1995): in molti casi, le loro aspettative sociali possono confermare quegli stessi ideali di forza e sicurezza che gli uomini cercano di avvalorare nei contesti dominati da altri uomini (Talbot, Quayle, 2010). In definitiva, oltre a temere mancanza di rispetto nei contesti omosociali (come evidenziato dalle teorie della maschilità egemonica di Connell o Kimmel), gli uomini hanno paura di non essere considerati abbastanza forti e attraenti dalle donne (Arrowsmith, 2015). Le donne rimangono, dunque, una presenza simbolica costante, anche all'interno di contesti esclusivamente maschili, che siano l'esercito, una serata tra ragazzi, o le comunità della manosphere

Nonostante il bisogno degli uomini eterosessuali di impressionare le donne e di essere da loro convalidati sia cosa nota (anche se qualsiasi *nice guy* negherebbe di avere bisogno di tale convalida), questo aspetto del maschile non viene sempre pienamente riconosciuto nella nostra società. Non sorprende allora che tanti discorsi che circolano nella manosphere sembrino, in realtà, guidati proprio dalla necessità psicologica di colmare questo vuoto: sconfessare il potere femminile, riconoscendone al contempo gli effetti deleteri. Potremmo dire che dietro ogni commento misogino e sessista, a margine di qualsiasi analisi critica nei confronti delle politiche di genere progressiste, al di là di qualsivoglia rivendicazione che circola in questi spazi, c'è sempre una *mossa difensiva* contro il potere femminile percepito nelle relazioni sentimentali e sessuali.



Affrontare questo aspetto dell'esperienza maschile potrebbe essere un modo per fornire agli uomini una *marcatatura identitaria* che tenga conto delle loro insicurezze relazionali e sentimentali, evitando al contempo che tali insicurezze vengano utilizzate come scuse per perpetuare comportamenti sessisti e manipolatori.

Sappiamo bene che applicare un'ottica esclusivamente psicologica e terapeutica alle questioni di genere – come, del resto, avviene spesso negli spazi della manosphere, significa precludersi strategie politiche capaci di sfidare i supporti istituzionali del dominio maschile (McMahon, 1993; Farci, Ricci 2021). Ma è anche vero che lo psicologico fa parte dell'esperienza quotidiana e non può essere ignorato così facilmente; è interconnesso con il contesto strutturale più ampio e non può essere trascurato.

In definitiva, non possiamo valutare la portata delle motivazioni offensive (sessiste, insensibili, misogine) degli adolescenti senza prima riconoscere le loro *motivazioni difensive* (soprattutto perché il comportamento risultante potrebbe essere simile). Per questo, serve una narrazione della vergogna e delle preoccupazioni maschili che investighi le ansie psicologiche degli uomini in risposta al loro bisogno di riconoscimento sociale e alla loro percezione dello sguardo femminile. Dobbiamo fare più domande su come tanti uomini sperimentano il potere delle donne, piuttosto che definire tale potere in relazione ai loro bisogni.

Solo così potremmo sviluppare un contesto che permetta ai ragazzi di riflettere serenamente sulle proprie vulnerabilità, invece di spingerli a nascondere tali insicurezze dietro affermazioni lamentose e vittimiste che finiscono solo per colpevolizzare il sesso opposto. Senza prestare ascolto ai sentimenti di sfiducia, senso di colpa, invidia e soggezione che gli uomini possono provare nei confronti del femminile, difficilmente potremmo comprendere e prevenire il percorso che trasforma adolescenti insicuri in un esercito di "bravi ragazzi" in guerra contro le donne.



## Bibliography

Allan J.A. (2015), "Phallic Affect, or Why Men's Rights Activists Have Feelings", *Men and Masculinities*, vol.19, n.(1): 22-41

Ahmed S. (2004), *The cultural politics of emotion*, Edinburgh, Edinburgh University Press.

Ahmetoglu G. Swami, V. (2012), "Do women prefer "nice guys"? The effect of male dominance behavior on women's ratings of sexual attractiveness", *Social Behavior and Personality*, vol.40, n.(4): 667-672.

Arrowsmith A. (2015), *Rethinking Misogyny: Men's Perceptions of Female Power in Dating Relationships*, Burlington, VT, Ashgate.

Bachaud L., Johns S. E. (2023), "The use and misuse of evolutionary psychology in online manosphere communities: The case of female mating strategies", *Evolutionary Human Sciences*, vol.5, n.28.

Badinter E. (1992), *XY. De l'identité masculine*, Paris, Editions Odile Jacob. Trad. It. *XY. L'identità maschile*, Longanesi, Milano, 1993.

Banet-Weiser S., Miltner K. M. (2016), "#MasculinitySoFragile: Culture, structure, and networked misogyny", *Feminist Media Studies*, vol.16, n.(1): 171-174.

Baselice K. A. (2023), "Analyzing Incels through the lens of evolutionary psychology", *Culture and Evolution*, vol. 20, n.1: 42-58.

Beynon J. (2001), *Masculinities and Culture*, Philadelphia, Pa, Open University Press.

Botto M., Filippi D., Ciccone S., Farci M., Virtù, L. (2022), "Maschilità tossica: potenzialità e limiti di un concetto diffuso", *AG About Gender-International Journal of Gender Studies*, vol.11, n.(21): 346-372.

Buss D. M., Angleitner A. (1989), "Mate selection preferences in Germany and the United States", *Personality and Individual Differences*, vol. 10, n.(12): 1269-1280,

Buss D. M., Barnes M. (1986), "Preferences in human mate selection", *Journal of Personality and Social Psychology*, vol.50, n.(3): 559-570.

Clatterbaugh K. (1996), *Contemporary Perspectives On Masculinity: Men, Women, And Politics In Modern Society*, Boulder-Colo, Westview Press.

Connell R. W. (1995), *Masculinities*, Stanford, Stanford University Press.



Manolo Farci  
*Siamo l'esercito dei "bravi ragazzi"*

Costello W., Rolon V., Thomas A. G., Schmitt, D. P. (2023), "The Mating Psychology of Incels (Involuntary Celibates): Misfortunes, Misperceptions, and Misrepresentations", *The Journal of Sex Research*, vol.0, n.(0): 1-12.

David D.S., Brannon, (1976), *The Forty-Nine Percent Majority: The Male Sex Role*, Reading MA, Addison Wesley Publishing Company.

De Boise S. (2017), "The personal is political...just not always progressive: Affective interruptions and their promise for CSMM", *NORMA*, 13(3 4): 158-174.

Dean J. (2010), *Blog Theory: Feedback and Capture in the Circuits of Drive*, Cambridge, U.K, Polity Press.

Desrochers S. (1995), "What types of men are most attractive and most repulsive to women?" *Sex Roles*, 32: 375-391.

Doosje B., Rojahn K., Fischer A. (1999), "Partner Preferences as a Function of Gender, Age, Political Orientation and Level of Education". *Sex Roles* 40: 45-60

Farci M. (2019), "La cultura digitale affettiva: Folle, pubblici e comunità", *Sociologia. Rivista quadrimestrale di Scienze Storiche e Sociali*, 53, 3: 7-14.

Farci M. (2022), "No Country for Men: Negotiating Men's Rights Activism in Digital Spaces", in Krijnen T., Nixon P.G., Ravenscroft M. D., Scarcelli C. M. (eds.), *Identities and Intimacies on Social Media*, London, Routledge, 134-149.

Farci M., Righetti, N. (2019). "Italian men's rights activism and online backlash against feminism", *Rassegna Italiana di Sociologia*, vol. 60, n.(4): 765-781.

Farci M., Scarcelli C. M. (2022), *Media digitali, genere e sessualità*, Milano, Mondadori Università.

Farci M., Ricci, O. (2021), "Io onestamente oggi non vorrei nascere maschio. Il caso Marco Crepaldi", *AG About Gender - International Journal of Gender Studies*, vol.10, n.(19): 167-198.

Farrell W. (1986), *Why men are the way they are*, New York, McGraw-Hill.

Farrell W. (1993), *The Myth of Male Power: Why Men Are the Disposable Sex*, London, Random House.



Manolo Farci  
*Siamo l'esercito dei "bravi ragazzi"*

Flood M. (2004), "Backlash: Angry Men's Movements", In Rossi S. E. (ed.), *The Battle and Backlash Rage On: Why Feminism Cannot Be Obsolete*, Philadelphia: PA, Xlibris Press, 261-278.

Gangestad S. W., Simpson, J. A. (2000), "The evolution of human mating: Trade-offs and strategic pluralism", *Behavioral and Brain Sciences*, vol.23, n.(4): 573-644.

Giglioli D. (2014), *Critica della vittima. Un esperimento con l'etica*, Milano, Nottetempo.

Glover R. A. (2003), *No More Mr Nice Guy: A Proven Plan for Getting What You Want in Love, Sex, and Life*, Philadelphia: PA, Running Press Adult.

Goodwin R. (1990), "Dating agency members: Are they different?", *Journal of Social and Personal Relationships*, n.7: 423-430.

Gotell L., Dutton E. (2016), "Sexual Violence in the 'Manosphere': Antifeminist Men's Rights Discourses on Rape", *International Journal for Crime, Justice and Social Democracy*, vol.5, n.(65): 65-80.

Halpin M., Richard N., Preston K., Gosse M., Maguire F. (2023), "Men who hate women: The misogyny of involuntarily celibate men", *New Media & Society*, DOI: 14614448231176777.

Hatfield E., Sprecher S. (1995), "Men's and women's preferences in marital partners in the United States, Russia, and Japan", *Journal of Cross-Cultural Psychology*, vol. 26, n.(6): 728-750.

Hawley G. (2017), *Making Sense of the Alt-Right*, New York, Columbia University Press.

Herold E. S., Milhausen R. (1999), "Dating preferences of university women: An analysis of the nice guy stereotype", *Journal of Sex & Marital Therapy*, vol.25, n.(4): 333-343.

Jensen-Campbell L. A., Graziano W. G., West S. G. (1995), "Dominance, prosocial orientation, and female preferences: Do nice guys really finish last?", *Journal of Personality and Social Psychology*, vol.68, n.(3): 427-440

Kaufman M. (1994) "Men, Feminism, and Men's Contradictory Experiences of Power", in Brod H., Kaufman M., (eds.), *Theorizing Masculinities*, Thousand Oaks CA, Sage, 142-63.

Kendall L. (2002), *Hanging Out in the Virtual Pub: Masculinities and Relationships Online*, Berkeley, CA, University of California Press.



Manolo Farci  
*Siamo l'esercito dei "bravi ragazzi"*

Kimmel M. (1987), "The Contemporary "Crisis" of Masculinity in Historical Perspective", in Brod H., (ed.), *The Making of Masculinities. The New Men's Studies*, London, Routledge, 121-153.

Kimmel M. (2008), *Guyland: The Perilous World Where Boys Become Men*, New York, HarperCollins.

Kimmel M. (2017), *Angry White Men: American Masculinity at the End of an Era*, New York, Nation Books.

McDaniel A. K. (2005), "Young Women's Dating Behavior: Why/Why Not Date a Nice Guy?", *Sex Roles*, vol.53, n.(5): 347-359.

McMahon A. (1993), "Male readings of feminist theory: The psychologization of sexual politics in the masculinity literature", *Theory and Society*, vol.22, n.(5): 675-695.

Menzies R. (2008), "Virtual Backlash: Representations of Men's «Rights» and Feminist «Wrongs» in Cyberspace", in Chunn D.E., Boyd S.B., Lessard H., (eds.), *Reaction and Resistance: Feminism, Law, and Social Change*, Vancouver, UBC Press, 65-97.

Messner M. A. (1998), "The Limits of «The Male Sex Role»: An Analysis of the Men's Liberation and Men's Rights Movements' Discourse", *Gender and Society*, vol.12, n.(3): 255-276.

Messner M. A. (2000), *Politics of Masculinities: Men in Movements*, Thousand Oaks, CA, Sage.

Nagle A. (2017), *Kill All Normies: Online Culture Wars From 4Chan And Tumblr To Trump And The Alt-Right*, Zero Books.

New C. (2001), "Oppressed and Oppressors? The Systematic Mistreatment of Men": *Sociology*, vol.35, n.(3): 729-748.

O'Neill R. (2018), *Seduction: Men, Masculinity and Mediated Intimacy*, Medford, MA, Polity Press.

Pleck J. H., Sawyer J. (1974), *Men and Masculinity*, Englewood Cliffs, N.J., Prentice-Hall.

Preston K., Halpin M., Maguire F. (2021) "The Black Pill: New Technology and the Male Supremacy of Involuntarily Celibate Men", *Men and Masculinities*, vol.24, n.(5): 823-841.



Manolo Farci  
*Siamo l'esercito dei "bravi ragazzi"*

Regan P. C. (1998), "Minimum mate selection standards as a function of perceived mate value, relationship context, and gender", *Journal of Psychology and Human Sexuality*, n.10: 53-73.

Regan P. C., Berscheid, E. (1997), "Gender differences in characteristics desired in a potential sexual and marriage partner", *Journal of Psychology and Human Sexuality*, n. 9: 25-37.

Robinson S. (2000), *Marked Men: White Masculinity in Crisis*, New York, Columbia University Press.

Sadalla E. K., Kenrick D. T., Vershure B. (1987), "Dominance and heterosexual attraction", *Journal of Personality and Social Psychology*, n. 52: 730-738

Salter A., Blodgett B. (2017), *Toxic Geek Masculinity in Media: Sexism, Trolling, and Identity Policing*, New York, Palgrave Macmillan.

Savran D. (1998), *Taking It Like a Man: White Masculinity, Masochism, and Contemporary American Culture*, Princeton N.J., Princeton University Press.

Spallacci A. (2019), *Maschi in bilico. Uomini italiani dalla Ricostruzione all'era digitale*, Milano, Mimesis.

Talbot K., Quayle M. (2010), "The Perils of Being a Nice Guy: Contextual Variation in Five Young Women's Constructions of Acceptable Hegemonic and Alternative Masculinities", *Men and Masculinities*, vol.13, n.(2): 255-278

Tietjen R. R., Tirkkonen, S. K. (2023), "The Rage of Lonely Men: Loneliness and Misogyny in the Online Movement of "Involuntary Celibates" (Incels)", *Topoi*, vol. 42, n.(5): 1229-1241.

Urbaniak G. C., Kilmann P. R. (2003), "Physical Attractiveness and the "Nice Guy Paradox": Do Nice Guys Really Finish Last?", *Sex Roles*, vol.49, n.(9): 413-426.

Van Valkenburgh S. P. (2018), "Digesting the Red Pill: Masculinity and Neoliberalism in the Manosphere", *Men and Masculinities*, 24, 1: 84-103.

Wendling M. (2018), *Alt-Right: From 4chan to the White House*, London, Pluto Press.

White M. (2019), *Producing Masculinity: The Internet, Gender, and Sexuality*, New York, Routledge.







A JOURNAL OF THE  
SOCIAL IMAGINARY

# L'odio indicibile: un'indagine etnografica digitale sulle comunità Femcel

Antonia Cava

acava@unime.it

COSPECS | University of Messina

Debora Pizzimenti

dpizzimenti@unime.it

Economia | University of Messina

## Abstract

*The Unspeakable Hate: a digital ethnographic investigation of Femcel communities*

The term Femcel is a neologism derived from 'female' and 'involuntary celibate'. This study investigates the occurrence of hate speech within an online Femcel platform known as Crystalcafe. The conversations of the users of this online community were analysed by employing the methodological framework of digital media ethnography and using the qualitative research software NVivo. Users, who feel excluded from romantic love and sex, share experiences, opinions and challenges pertaining to relational and sexual dynamics. However, behind the apparent sharing and solidarity, complex manifestations of hate speech emerge, fostering an environment where expressions of anger, misandry, and other forms of discrimination proliferate.

## Keywords

Femcel | Hate Speech | Digital Platform | Online Community | Digital Ethnography



## Introduzione

Il tema dell'odio online, sebbene strettamente connesso all'odio offline, presenta tratti di maggiore complessità. Nel contesto dinamico della comunicazione online, bisogna considerare non solo la variabilità delle forme di hate speech, ma anche il coinvolgimento di una pluralità di attori sociali nei processi di diffusione dei messaggi d'odio, l'incertezza e talvolta l'ambiguità dei ruoli (produttori di hate speech, diffusori più o meno consapevoli, vittime) e la rapidità delle dinamiche di propagazione (Bentivegna & Rega, 2020: 153). È altresì fondamentale considerare la complessità derivante dal fatto che l'hate speech può essere considerato un "moving concept", che si adatta e si ridefinisce in relazione all'immaginario sociale (Durand, 1960; Morin, 2005) e al contesto culturale, politico e comunicativo in cui si manifesta (ivi: 154).

Trolls, flamers o haters sono soggetti contraddistinti dalla morbosa necessità di "stare online" quasi esclusivamente per generare, incitare e condividere discorsi d'odio.

La diffusione delle piattaforme digitali ha amplificato la possibilità, per i più disparati gruppi sociali, di creare comunità online in cui scambiare e alimentare le proprie idee. Negli ultimi anni questi ambienti online hanno dato modo a gruppi di uomini che si autodefiniscono attraverso il termine Incel (celibe involontario) (Dolce & Pilla 2019; Nagle 2018; De Gasperis, 2021; Pizzimenti & Pasciuto 2022) di crescere, proponendo una maschilità oppressiva, che riproduce e rinforza la discriminazione e la violenza di genere (Dordoni & Magaraggia, 2021). La figura dell'Incel è diventata emblematica di quella che Sarah Banet-Weiser (2018) ha chiamato *networked misogyny*.

La Manosfera rappresenta un'espansione di blog, forum e siti web dedicati alla discussione della mascolinità. Le sue due sottoculture più significative sono Men's Rights Activists (MRA) e pick-up artist (PUA), ma include anche antifemministi, gruppi per i diritti dei padri, Incel, androfilii (uomini attratti dallo stesso sesso che non si identificano come omosessuali) e gruppi marginali ancora più oscuri (Gotell & Dutton, 2016; Lilly, 2016).

Dopo le recenti atrocità di massa commesse da uomini ispirati dall' "ideologia Incel", il concetto di celibato involontario è stato collegato nell'immaginario culturale con la rabbia esplosiva e imprevedibile di uomini bianchi vendicativi e umiliati, inclini a esplosioni di violenza estrema. Secondo Pankaj Mishra (2017), stiamo attraversando un' "era della rabbia" abilitata dai social media e caratterizzata dal risentimento, che sembra offrire una chiave di lettura adatta per comprendere l'inasprirsi dell'ideologia Incel, che rappresenta un'estrema incarnazione della politicizzazione e dell'uso dell'ira invidiosa come arma. Tuttavia, l'analisi di Mishra si limita a considerare il risentimento quasi esclusivamente come un sentimento maschile. Risulta ancora difficile inserire e accettare culturalmente la presenza e l'esistenza di donne che odiano (Kay, 2022: 31). Un aspetto interessante e preso poco in considerazione da studiosi e studiosi, ricercatrici e ricercatori è proprio il senso di rabbia provato dalle donne ed espresso negli spazi online. In questo alveo di odio e



diseguaglianza sociale hanno trovato il loro spazio le cosiddette Femcel, una comunità online di donne che si descrivono come incapaci di avere relazioni sessuali o romantiche nonostante i loro sforzi e desideri, a causa di una pericolosa combinazione di misoginia e aspettative di standard di bellezza irraggiungibili.

Come gli Incel, attribuiscono il rifiuto sessuale al loro aspetto e condividono l'esperienza di incapacità o difficoltà nel trovare una relazione romantica o sessuale, nonostante il desiderio di avere un partner. Entrambe le comunità sostengono che coloro che sono considerati meno attraenti sono svantaggiati nelle relazioni e nella società in generale. Tuttavia, a differenza dei *celibi involontari*, le Femcel si colpevolizzano per il loro stato di nubilitato involontario, interiorizzando la violenza misogina.

Sia gli Incel che le Femcel condividono sentimenti profondi di solitudine ed esclusione. Questa esclusione è sentita come un'ingiustizia sociale in cui gli attributi fisici determinano l'accesso alle opportunità relazionali. Tuttavia, gli Incel tendono a ricondurre la loro solitudine all'inaccessibilità del sesso, le Femcel, invece, sottolineano l'assenza di relazioni significative ed emotivamente profonde, rivelando una differenza nel modo in cui ciascuna comunità concepisce l'intimità (Alava & Nagem, 2024).

Inoltre, se gli Incel nel tempo hanno ricevuto l'attenzione di accademici e giornalisti per l'insoddisfazione che esprimono a causa della loro mancanza di capitale sessuale, le Femcel lamentano di non essere legittimate a rivendicare un'identità basata su questo stato di avvilimento (Deliovsky, 2008).

Le Femcel, pertanto, pur comparabili agli Incel nella loro frustrazione e isolamento, si distinguono, sfidando la narrazione dominante all'interno della manoscera, che tende a ignorare o minimizzare la possibilità che anche le donne possano soffrire di solitudine ed "esclusione romantico-sessuale".

L'obiettivo di questa ricerca è di esaminare gli aspetti sociali, culturali, psicologici e il tipo di immaginario che strutturano le donne che si autodefiniscono Femcel, con particolare attenzione al contesto online.

Quale immaginario costruiscono queste utenti della rete per definire le proprie aspettative, le proprie identità, i propri ruoli? Come emergerà dai risultati della ricerca, questo immaginario sembrerebbe costruito secondo una struttura univoca, attraverso la quale prendono forma i sogni e i desideri di queste donne, che tenderebbero a condividere una visione del mondo, o un' "ideologia" dominante in quella particolare comunità digitale.

Come è noto, ogni società ha costruito nel tempo degli immaginari che servissero a semantizzare la realtà circostante. Da questo punto di vista l'immaginario è l'equivalente dell'ideologia dominante in un dato momento; ideologia intesa come strutturazione coerente di idee, discorso sulle idee, o visione del mondo, in cui prendono posto tutti gli oggetti e tutti i linguaggi, che descrivono gli stessi, trovando una dimensione semantica e significativa (Carzo, Cava & Salvo, 2007).

In questo contributo esploriamo il fenomeno Femcel attraverso l'analisi e la descrizione di una comunità attiva in rete, la più ampia che si incontra sul sito di



imageboard Cristalcafe. Analizziamo come queste attrici sociali, invocando una logica basata su un *sentiment* condiviso, affermino il loro “discorso sul mondo”.

Indaghiamo come le donne Femcel affrontino una società che loro stesse definiscono crudele ed essenzialmente misogina; il loro attivismo e la sensazione di fare parte di un gruppo dove sentirsi libere di esprimersi e comunicare la loro rabbia; il senso di umiliazione che vivono per non essere desiderate dagli uomini.

## 1. Le subculture online: un'esplorazione dell'universo Femcel

Sebbene non esistano ancora molti studi sulle Femcel, il termine è ampiamente utilizzato, non solo all'interno dei forum della piattaforma Reddit, ma su tutte le principali piattaforme social. L'emersione di tristezza mista a frustrazione farebbe degli spazi online abitati dalle Femcel luoghi del disagio e di una malinconia condivisa, che spesso appare orientata all'autolesionismo. In particolare, le Femcel credono di essere desiderabili per gli uomini solo dal punto di vista sessuale. In base alla descrizione fornita da TruFemcels, chat di Reddit con decine di migliaia di utenti chiusa nel 2021 perché si era trasformata in un luogo pervaso da odio, misandria, omofobia, transfobia e ideologie dell'alt-right (Kohn, 2020), “una femcel (crasi tra “female” e “involuntary celibate”) è una donna adulta il cui aspetto fisico è inferiore alla media (<4/10) o soffre di un impedimento significativo che le impedisce di assicurarsi una relazione romantica - NON sesso”.

La Femcelosfera comporta filosofie diverse e talvolta contraddittorie: alcune donne “Pinkpillate” (Balci et. al, 2023; Evans & Lankford, 2023) si identificano come femministe radicali (Radfem), che promuovono il movimento “Women Going Their Own Way”; altre si definiscono “Redpillate” (Balci et al., 2023) e si dichiarano antifemministe; altre ancora sono ossessionate dall'idea di assicurarsi l'approvazione e le attenzioni maschili; per quelle che invece si definiscono “Blackpillate” la sensazione di essere “left behind” ed emarginate dalla società, perché non rispondenti a standard di bellezza restrittivi e rigidamente definiti, provoca uno stato di depressione profondo, che non consente vie d'uscita dalla Femcelosfera (Kay, 2022).

Nell'immaginario della Femcelosfera convivono espressioni diverse correlate alle diverse declinazioni del femminile e modelli di esclusione sociale intersezionali (Crenshaw, 1989)<sup>1</sup>. Da questa prospettiva, il termine Femcel può essere considerato come un meta-costrutto che include diverse identità e comprende una vasta gamma di disposizioni verso il senso di rabbia, sofferenza e umiliazione per essere nubili

<sup>1</sup> In questa sede si fa riferimento al concetto di intersezionalità per esplorare da una prospettiva più ampia la rabbia e il sentimento di esclusione delle donne Femcel, “per comprendere come opera l'oppressione delle donne, poiché non si può ridurre l'attenzione a un unico fattore (sia esso il genere, la razza, la classe o l'eterosessismo), ma si deve tentare di indagare il modo in cui tutti quei fattori, nella loro pluralità, si intersecano fra loro al fine di rafforzare e riprodurre l'inferiorità delle donne. Una volta indossate le lenti intersezionali, è possibile quindi percepire la specifica forma di oppressione generata dalle intersezioni stesse, la quale cambia a seconda del contesto e dei diversi gruppi sociali” (Bottici, 2023: 61).



involontarie. Nonostante non esista una definizione univoca per il termine Femcel, anche a causa dei pochi studi accademici che si occupano della sofferenza delle donne che si sentono sessualmente fallite (Kay, 2022; Ling, 2023), in generale, la parola Femcel si riferisce a quelle donne che soffrono perché vivono una condizione di solitudine emotiva e sessuale. Questa condizione porta alcune donne che si definiscono involontariamente single a rivolgersi a spazi femminili online per discutere delle sfide, dei rischi e del senso di solitudine, umiliazione e abiezione che devono affrontare ogni giorno. In altre parole, il termine "Femcel" si riferisce a donne che non riescono a trovare un partner romantico o sessuale nonostante lo desiderino (Donnelly et al., 2001; Wu, 2022).

Come precedentemente accennato, sebbene gli Incel abbiano ricevuto una considerevole attenzione e siano stati oggetto di numerosi studi, vi è una significativa carenza di letteratura e ricerca riguardante le Femcel. In Italia, in particolare, il fenomeno delle Femcel non ha ancora ricevuto un'analisi sistematica e approfondita nell'ambito accademico.

Negli spazi online, così come nella vita offline, le donne che si identificano come Femcel possono essere soggette a un particolare tipo di sessualizzazione. Mentre gli Incel possono essere percepiti come "brutti" o "non desiderabili" dalla società, le Femcel spesso devono affrontare un ulteriore livello di stigma, legato all'assunto che ogni donna possa facilmente attrarre un uomo se lo desidera veramente. Di conseguenza, le esperienze delle Femcel possono essere negate o minimizzate, generando una dinamica in cui le loro sfide e il loro dolore non sono adeguatamente riconosciuti o compresi (Kay, 2022).

Le piattaforme sono un punto di osservazione privilegiato per studiare queste sottoculture. Il modello della rete, del resto, è fondamentale per comprendere le società contemporanee basate sulla comunicazione (Codeluppi, 2012).

"Le piattaforme non riflettono il sociale: producono le strutture sociali nelle quali viviamo" (van Dijck et al., 2018), ossia intervengono sulla modalità di definizione dei legami attraverso forme di connessione che combinano norme sociali e norme socio-tecniche proprie degli ambienti online (van Dijck, 2013, come citato in Boccia Artieri, 2020: 87), creando un campo simbolico e di pratiche che delimitano specifici modi di relazionarsi - spesso distinti da quelli offline - e che presiedono a nuovi processi di significazione dello stare assieme (ivi: 88). La costruzione e gestione di socialità che passa attraverso le piattaforme non è quindi definita da un semplice travaso negli spazi tecnologici di dinamiche preesistenti, ma viene plasmata dalle architetture e dalle affordance delle piattaforme stesse, che circoscrivono le possibilità e le forme della messa in relazione fra gli individui. La formazione di questi gruppi e sottogruppi, come nel caso delle Femcel, mette in evidenza il fatto che gli spazi digitali offrono una modalità di coinvolgimento verso la realtà sociale di carattere prevalentemente affettivo, ossia incentivano forme di allineamento e amplificazione emotiva condivisa che possono marcare la polarità di uno specifico aggregato collettivo (Papacharissi, 2015, come citato in Boccia Artieri, 2020: 95), anche semplicemente attraverso un commento di sostegno, una condivisione o un like (ibidem).



A tenere insieme l'universo Femcel sono un sentire comune, una rabbia condivisa che si traduce in senso di non accettazione sociale, disprezzo per gli uomini, odio per il proprio corpo - postato sulle varie piattaforme social ed espresso mediante un vocabolario condiviso (che può essere definito geek) -, la condivisione delle proprie storie personali, che spesso si traducono in appuntamenti falliti e incontri deludenti o ancora liste di consigli utili per apparire più belle e in forma.

## 2. Crystalcafe: la comunità online delle "donne anti-uomini"

Descriviamo, ora, l'immagineboard *Crystalcafe*. Successivamente presenteremo la metodologia e analizzeremo in modo più dettagliato i cluster in cui si manifesta l'hate speech.

Urban Dictionary definisce l'immagineboard *Crystalcafe* un sito "creato appositamente per le donne per sentirsi al sicuro in un ambiente più femminile lontano dai maschi estremisti"<sup>2</sup>.

Un'immagineboard, detta anche \*chan come abbreviazione di channel, è un tipo di sito Internet basato sulla pubblicazione di immagini, testi e video da parte dei propri utenti. A differenza di un forum, i cui contenuti sono principalmente di natura testuale, le \*chan dipendono largamente dalle immagini, tanto che una nuova discussione può essere iniziata solo con un'immagine, e non con un commento testuale singolo.

Altra caratteristica delle immagineboard è quella di consentire all'utente una partecipazione basata sull'anonimato; per inserire risposte in una immagineboard non è necessario registrarsi, ma è possibile commentare in forma completamente anonima. L'unico metodo di identificazione consentito è l'impiego di tripcodes associati al nome, ovvero sequenze alfanumeriche generate in modo automatico da uno script contenuto nel software stesso.

Fondato nel 2017 come alternativa più sicura e accogliente per le donne rispetto ad altre immagineboard come 4chan e 8chan, spesso associate all'estremismo politico, alla violenza, alla misoginia e ai discorsi d'odio in generale (Hine et al., 2017; Hawley, 2017; Nagle, 2017), su *Crystalcafe* si può discutere di vari argomenti, divisi in topic di discussione. Non ci sono informazioni precise sul numero esatto di utenti su *Crystalcafe* perché il sito non rende pubbliche queste informazioni, tuttavia, la piattaforma risulta essere relativamente popolare tra le donne che cercano spazi completamente femminili, infatti, si possono trovare molte discussioni attive sul sito ogni giorno. Analogamente ad altre comunità online, *Crystalcafe* utilizza un linguaggio verbale e visivo caratterizzato dall'uso di memi, termini subculturali e riferimenti alla più ampia cultura del web. La struttura del sito presenta diverse sezioni e sottosezioni divise per tematiche<sup>3</sup>. La comunità, inoltre, offre una vasta gamma di funzionalità come bacheche, chat room e forum di discussione.

<sup>2</sup> Fonte: <https://www.urbandictionary.com/define.php?term=Crystal.Cafe>; consultato il 27 febbraio 2024.

<sup>3</sup> Le regole per partecipare alle discussioni; le FAQ (descrizione della piattaforma, la spiegazione di



### 3. Metodologia della ricerca

La metodologia utilizzata è l'etnografia dei media digitali (Sumiala & Tikka, 2020). Quest'ultima si distingue come un metodo "nativo digitale", che integra il mondo online e quello fisico in maniera inscindibile. Questo tipo di etnografia abilita un'analisi in tempo reale, che supera le barriere geografiche e permette di esaminare gli immaginari in cui si sviluppano le dinamiche dei gruppi Femcel. Allo stesso tempo, promuove una comprensione autentica delle esperienze e delle prospettive dei membri di tali comunità online, contribuendo così a illuminare le dinamiche di genere e i modi in cui queste vengono percepite e discusse. Considerato che la presenza di un estraneo all'interno di questa specifica comunità sarebbe stata indesiderata e sgradita e considerata la forte ostilità del gruppo, è stata condotta un'osservazione non partecipante. Questo metodo di ricerca implica che lo studioso rimanga esterno rispetto la realtà osservata, evitando di influenzare o interagire con i soggetti oggetto di indagine e senza rivelare il proprio ruolo di ricercatore.

Il monitoraggio dell'immagineboard Crystalcafe è stato condotto tra gennaio e maggio 2023 e si è concentrato sull'individuazione e la raccolta dei commenti e di alcune parole chiave che si inscrivono a pieno titolo all'interno del fenomeno dell'hate speech online. Per comprendere meglio i diversi tipi di hate speech in questa comunità, i commenti sono stati raggruppati e classificati in macrocategorie e sottocategorie. Abbiamo realizzato un'indagine sistematica della portata e della natura dell'hate speech nella comunità marginale Crystalcafe, analizzando sia forme esplicite che implicite di hate speech, associando la codifica manuale dell'hate speech con approcci automatizzati, attraverso l'impiego del software Nvivo. Il campione preso in esame è frutto di una procedura di campionamento non probabilistico che mira a soddisfare una rappresentatività "tipologica" ma non statistica, in altre parole che aspira a presentare la descrizione di una realtà esistente all'interno delle piattaforme digitali senza garantire in alcun modo una corrispondenza tra il campione e la totalità delle utenti che frequentano l'immagineboard Crystalcafe. I risultati ottenuti non sono generalizzabili all'intera popolazione di riferimento, ma sono impiegati per obiettivi diversi tra cui investigare e descrivere le peculiarità e i comportamenti online del campione coinvolto nella ricerca; sviluppare nuove ipotesi che possano essere verificate su un campione probabilistico; offrire spunti di riflessione per orientare e stimolare le fasi successive di questo studio.

La domanda da cui è nato il nostro interesse di ricerca è: come l'utilizzo di Internet e delle piattaforme social può alimentare e trasformare l'odio in quanto tale nel fenomeno dell'hate speech online?

Come si è detto, l'odio online non muta nella sostanza la nozione tradizionale dei discorsi d'odio, poiché in tutt'e due i casi si è in presenza di comportamenti di soggetti e/o gruppi, più o meno organizzati, che puntano a umiliare, denigrare e



trattare in modo discriminatorio altre persone accomunate da determinate caratteristiche (Ziccardi, 2016). Le militanti della Femcelosfera, come vedremo, sono produttrici di un vero e proprio vocabolario condiviso cui fare riferimento. Per gli obiettivi di questa ricerca ci si è concentrati maggiormente sui termini di odio e sugli epiteti denigratori più "popolari" usati in Crystalcafe, attraverso i quali è anche possibile trarre conclusioni sui principali gruppi target di incitamento all'odio.

I termini di odio e gli insulti che appaiono di più - "schifoso", "ritardato", "bastardo", "frocio", "negro", "zerbino/a" o epiteti denigratori come "moid", "XY", "scroto", e "Pickmeisha" - sono indicativi di sentimenti prevalentemente misandrici, ma anche omofobi, transfobici, razzisti, misogini e abilisti e suggeriscono che gli uomini, le donne, le persone di colore, la comunità lesbica, gay, bisessuale, transgender e queer e le persone con disabilità potrebbero essere vittime ricorrenti di incitamento all'odio.

Per il campionamento dei commenti e dei messaggi dei thread è stato applicato un approccio mirato (Patton, 1990). I commenti sono stati scaricati e salvati e quando sono risultati pertinenti e popolari nella comunità (hanno ricevuto un numero elevato di commenti e interazioni) sono stati selezionati per l'analisi. Per fare ciò è stata effettuata un'analisi qualitativa del contenuto utilizzando un campione di 1013 commenti. Durante il processo di analisi, è stata implementata una metodologia di codifica deduttiva con l'obiettivo di categorizzare in modo sistematico i commenti oggetto di studio in diverse sottocategorie e di aggregarli in tre macrocategorie principali. Tale approccio di categorizzazione ha dimostrato di essere uno strumento analitico efficace nella comprensione delle varie manifestazioni dell'odio presenti nei commenti esaminati. Le macrocategorie individuate sono state suddivise come segue. La prima macrocategoria è stata identificata come "Odio Esplicito", all'interno della quale sono state definite sottocategorie quali "Insulti Personali", "Insulti Generali" e "Minacce Esplicite di Violenza Fisica, Psicologica o di Altro Tipo". La seconda macrocategoria è stata denominata "Odio Implicito" e ha compreso le sottocategorie "Stereotipizzazione Negativa", "Disinformazione", "Elevazione del Gruppo" e "Ideologia Disumana". Infine, la terza macrocategoria è stata denominata "Commenti Irrilevanti", includendo i commenti che non potevano essere collocati all'interno delle categorie di odio sopra descritte e che risultavano di scarsa pertinenza ai fini dell'analisi condotta.

#### 4. I risultati della ricerca

Come prima macrocategoria, abbiamo codificato l'incitamento all'odio esplicito (*explicit hate speech*) in conformità con le concettualizzazioni ricorrenti nella letteratura. All'interno di questa categoria, abbiamo inserito gli insulti (Erjavec & Kovačič, 2012), come espressioni offensive, dispregiative o degradanti, incluso l'uso di quelli che abbiamo definito "epiteti denigratori" (Bianchi, 2021). Si considerano epiteti denigratori (il termine inglese è *slur*) quelle espressioni che comunicano derisione, disprezzo o odio verso gruppi sociali - e verso individui in virtù della sola



appartenenza a un certo gruppo sociale. I gruppi target vengono identificati di volta in volta sulla base di certi tratti sociali (reali o percepiti): etnia, nazionalità, religione, genere, orientamento sessuale, (dis)abilità. Abbiamo, poi, distinto tra insulti personali (cioè attacchi rivolti ad un individuo specifico) e insulti generali (attacchi rivolti a un collettivo), considerando anche il bersaglio degli insulti personali e il bersaglio degli insulti generali. La schematizzazione è stata compilata sulla base di ricerche su gruppi frequentemente emarginati (Burn et al, 2005; Mondal et al., 2018; Rieger et al., 2021) ed estesa induttivamente durante il processo di codifica. Dato lo scopo della ricerca, ci si concentrerà esclusivamente sulle conversazioni che esprimono sentimenti misandrici. È emerso che su Crystal.cafe esistono spazi in cui le Femcel denigrano, umiliano e disumanizzano gli uomini attraverso l'uso di termini ed epiteti misandrici, omofobi e transfobici. Inoltre, abbiamo codificato le minacce di violenza come una forma di hate speech esplicito (Erjavec & Kovačič, 2012; Gagliardone et al., 2015; Rieger et al., 2021), includendo sia minacce concrete di violenza fisica, psicologica o di altro tipo, sia richieste di violenza da infliggere su individui o gruppi specifici.

Come seconda macrocategoria, abbiamo codificato l'incitamento all'odio implicito (*implicit hate speech*). Per distinguere le diverse sottocategorie di questo tipo di hate speech, ci siamo affidati maggiormente a un approccio esplorativo concentrandoci sulle forme di comunicazione che sono state descritte in letteratura come dispositivi per mascherare l'odio. L'odio implicito è un fenomeno di notevole rilevanza nell'ambito del discorso sociale contemporaneo e richiede un'analisi attenta. Tale fenomeno presenta notevoli sovrapposizioni concettuali con quanto precedentemente categorizzato da Buyse (2014) come "discorso di paura", caratterizzato dalla sua intenzione deliberata di instillare una paura esistenziale nei confronti di un gruppo specifico (Saha, Mathew, Garimella, & Mukherjee, 2021). L'incitamento all'odio implicito si manifesta, inoltre, in varie forme all'interno del contesto di diffusione intenzionale di notizie false, rappresentando un'ulteriore manifestazione di incitamento all'odio implicito poiché mira a creare una percezione distorta dei gruppi bersaglio attraverso la manipolazione dell'informazione e la promozione di pregiudizi negativi.

La prima sottocategoria di incitamento all'odio implicito è etichettata come stereotipizzazione negativa ed è stata codificata quando le utenti hanno espresso convinzioni eccessivamente generalizzate e semplificate su caratteristiche (negative) o comportamenti di diversi gruppi target. La seconda sottocategoria (disinformazione/teorie della cospirazione) riflette sia la "semplice" disinformazione che le false dichiarazioni sui gruppi bersaglio e le teorie della cospirazione "avanzate", che rappresentano i gruppi target come maliziosamente all'opera per ottenere un maggiore potere ideologico, politico o finanziario. Una terza sottocategoria è stata etichettata come elevazione del gruppo ed è stata codificata quando le affermazioni erano riferite a un certo gruppo (in base al genere, identità sessuale, etnia, ecc.), spesso escludendo e svalutando implicitamente altri gruppi. L'ultima sottocategoria di incitamento all'odio implicito è stata etichettata come

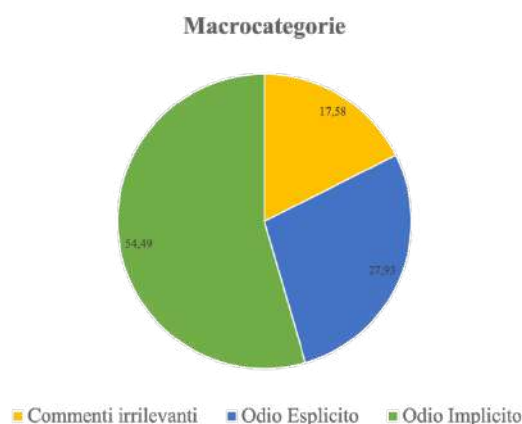


ideologia disumana. Qui, è stato codificato se un'utente ha sostenuto o glorificato ideologie dell'odio che escludono l'uomo dalla società o di supremazia femminile.

Infine, è stata aggiunta una macrocategoria denominata "commenti irrilevanti" per escludere alcuni commenti (ad esempio "ci sentiamo dopo", "ok", "mi collego più tardi", "oggi sono stata impegnata").

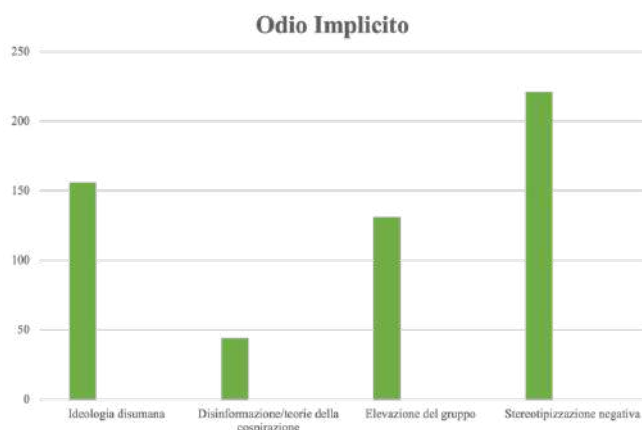
Attraverso questa classificazione dei commenti in macrocategorie e sottocategorie, è stato possibile rilevare i principali nuclei intorno a cui si sviluppa l'hate speech rivolto contro gli uomini.

Più della metà dei 1013 commenti analizzati contiene forme di hate speech implicito (Fig. 1).



**FIG.1** – *Classificazione dei commenti in Macrocategorie*

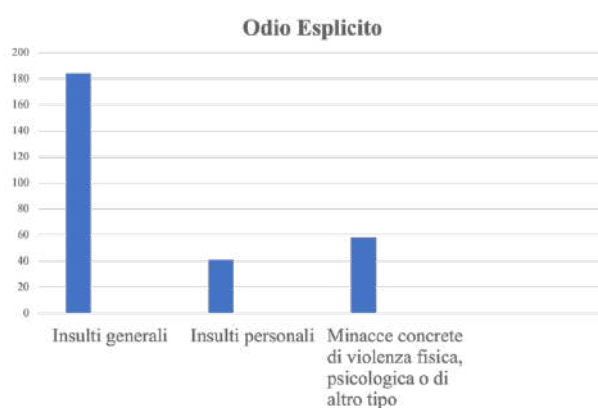
All'interno di questa macrocategoria sono stati classificati n:156 commenti in cui si fa riferimento o si esprimono ideologie disumane, relative a teorie di supremazia femminile su basi evoluzioniste o biologiche e usando criteri (ad esempio, la longevità femminile rispetto agli uomini), che mirano a sottolineare la superiorità delle donne; N:44 commenti dove sono presenti elementi di disinformazione/teorie della cospirazione, relativi a teorie secondo cui le donne si sono evolute prima degli uomini (il cromosoma Y rende gli uomini geneticamente inferiori alle donne, le donne sono biologicamente superiori agli uomini, i maschi sono superflui per la riproduzione). Gli stereotipi negativi sono stati osservati in 221 commenti: si sostiene che gli uomini sono misogini, predatori, ossessionati dal sesso e dal porno, violenti e pedofili; l'elevazione del gruppo è stata codificata per 131 commenti (Fig. 2).



**Fig. 2** – Commenti caratterizzati da Hate Speech Implicito (n:552)



Le minacce di violenza sono state osservate in 58 commenti e contenevano sentimenti di odio rivolto contro gli uomini e volontà di eliminare definitivamente il genere maschile e il cromosoma Y. All'interno del nostro campione, gli insulti personali sono emersi come la forma meno comune di hate speech. Molti insulti personali hanno attaccato l'orientamento sessuale di un individuo, la sua identità sessuale, o erano riferiti a una disabilità reale o presunta. Gli insulti generali (n:184), sono riferiti nei confronti degli individui di sesso maschile e qualche volta nei confronti delle donne, delle persone con disabilità, della popolazione anziana e dei neri (Fig. 3).



**Fig. 3** – Commenti caratterizzati da Hate Speech Esplicito (n:283)

NVivo permette di esplorare, tramite queries, il vocabolario del corpus testuale. Oltre a classificare i commenti in categorie, è stata condotta una sistematizzazione

dei termini più utilizzati dai membri attivi dell'immagineboard. Successivamente è stato registrato il numero di volte in cui comparivano alcuni termini offensivi impiegati nei confronti degli uomini (Tab.1).

EPITETO DENIGRATORIO	SIGNIFICATO NELLO SPAZIO FEMCEL	NUMERO DI VOLTE UTILIZZATO
COOMER	Termine utilizzato per etichettare in senso dispregiativo e volgare un uomo che è dipendente dalla masturbazione; per estensione, chiunque sia eccessivamente eccitato.	39
FAGGOT	Frocio, passivo.	28
MOID	Etimologia: Maschio + umanoide. Termine utilizzato per riferirsi agli uomini come dei subumani e dei sottosviluppati.	317
RETARDED	Termine impiegato per etichettare in senso dispregiativo gli uomini riferendosi a loro come soggetti che presentano ritardo mentale, cioè uno sviluppo psichico inferiore a quello proprio dell'età anagrafica.	31
SCROTE	Un uomo stereotipato di basso valore e odioso, inutile e spregevole.	82
366 XY	Codice genetico del maschio. Il cromosoma Y rende gli uomini geneticamente inferiori alle donne. La sua presenza nella società è considerata dalle Femcel la causa di tutti i mali del mondo.	18

Tab. 1 – Classificazione degli Epiteti Denigratori

In tabella abbiamo riportato solo alcuni dei termini utilizzati dalle militanti del movimento Femcel per esprimere il loro odio contro gli uomini. Accanto ai più classici insulti, vi sono anche neologismi inventati dal gruppo stesso come MOID, che viene impiegato per appellare gli uomini con un totale senso di deumanizzazione e spersonalizzazione del soggetto.

## 5. Conclusioni

L'immaginario della comunità Femcel studiata è permeato da stereotipi interiorizzati, che generano sentimenti di frustrazione e rabbia. Quali sono le motivazioni profonde e le dinamiche che soggiacciono e animano questo immaginario?

A caratterizzare questo forum è un *sentiment* condiviso, che viene espresso attraverso i vari thread, le discussioni e la condivisione di immagini. All'interno di spazi online come i forum Femcel, la socialità è basata sull'*affetto* (Papacharissi, 2015)

e le utenti sono legate tra loro da un'intensa emotività diretta verso un oggetto comune (Papacharissi, de fatima Oliveira, 2012). Da questa prospettiva, la Femcelosfera è intesa come un esempio di *comunità affettiva*. Difatti, ad accomunare queste utenti non sono stati dapprincipio contenuti ideologici o programmi politici, quanto un insieme di *regole del sentire* comuni, caratterizzate da narrazioni incentrate su sentimenti di rabbia, misandria, solitudine, tristezza. In primo luogo, grazie alla presenza di un vocabolario comune che convalida l'adozione di hate speech e atteggiamenti tipici del movimento Incel, queste utenti possono liberamente esprimere le proprie idee e opinioni.

La condivisione di queste narrazioni, o significative rappresentazioni della realtà, consente un comune modo di sentire, di vivere una medesima situazione, contribuendo a suggerire ad ogni attrice sociale stessi modi di vedere e intendere il mondo. La dimensione dell'immaginario, si sa, ha un ruolo fondamentale nella costruzione della condivisione di significati, nella creazione della capacità del "comune sentire".

Inoltre, all'interno di questi forum, la condivisione continua di informazioni da parte delle utenti, genera una situazione in cui si è esposti solo, o quasi esclusivamente, ad opinioni che concordano con la propria e che confermano il proprio punto di vista riguardo a determinati temi o argomenti.

Tale processo di polarizzazione emotiva può rafforzarsi all'interno di reti omofile tra utenti che, chiudendosi al confronto con opinioni diverse dalle proprie, finiscono per trovare rinforzo alle loro convinzioni. La questione non è che le piattaforme, attraverso le proprie logiche algoritmiche, determinano la polarizzazione affettiva, ma che specifiche "modalità del sentire", per mezzo delle piattaforme, diventano il collante sociale per la formazione di gruppi affettivamente coesi. Perciò, sono le emozioni a produrre certe forme di allineamenti condivisi, a tenere insieme o legare una collettività (Boccia Artieri, 2020). Come sottolinea Papacharissi (2021), Internet non crea incitamento all'odio, sono le persone a farlo. Tuttavia, le piattaforme di social media rendono gli atti di odio più visibili e quindi più facilmente diffondibili. Le emozioni funzionano, dunque, come un vero e proprio capitale sociale affettivo che, circolando, finisce per materializzare una distinzione tra un noi e un loro, in cui loro sono rappresentati come la causa del nostro sentimento.

La polarizzazione affettiva funziona come un vero e proprio strumento di lavoro gestional-emotivo, che permette agli individui di mantenersi in linea con le proprie ansietà e paure o garantisce loro una forma di conforto e sollievo sociale. In un certo senso, la polarizzazione offre agli individui un insieme di regole implicite su "che cosa" e "come" ci si dovrebbe "sentire" e come si dovrebbero esprimere le emozioni in relazione a determinate questioni (Kanai, 2019). Di conseguenza, non sono gli scambi comunicativi tra gli individui a determinare il sentimento comune, ma è vero piuttosto il contrario: è il sentimento richiesto a orientare la natura delle discussioni attorno a ciò che viene condiviso. Le regole del sentire non riflettono, dunque, solo un modello di partecipazione comune, ma finiscono per determinare il modo in cui vengono inquadrati le situazioni sociali. Attraverso questa adesione a specifiche regole del sentire, si attivano strategie di gruppo per sottolineare una certa posizione



ideologica, che serve a orientare la qualità delle conversazioni e le tendenze (thread) predominanti all'interno dei forum online.

Nella comunità Femcel il senso di sorellanza e di sostegno tra le utenti non proviene da premesse ideologiche o programmi politici, quanto dalla condivisione di un insieme di *regole del sentire*, caratterizzate prevalentemente da narrazioni sulla sofferenza, il senso di solitudine e l'umiliazione, che proviene dal percepirsi come "brutte" e non essere corrisposte nell'amore romantico, su sentimenti di odio e di misandria o sull'autolesionismo. Pubblicando i loro sentimenti e le loro esperienze in modo anonimo all'interno di piattaforme che hanno progettato per sé stesse, le Femcel sostengono di essere invisibili a causa del loro aspetto anormale e che la cultura misogina e gli standard di bellezza impossibili impediscono loro di essere accettate. In altre parole, a caratterizzare il comportamento online delle Femcel è una costante autovittimizzazione, la quale si traduce in odio verso sé stesse, il proprio aspetto e l'universo maschile.

Oltre ad essere frustrate e vivere un'inquietudine carica di ira per la mancanza di relazioni e per lo scarso valore degli uomini, fino al punto di esprimersi attraverso invettive volgari, queste donne hanno spesso discusso di azioni violente e minacce fisiche e psicologiche nei confronti degli uomini, hanno proposto piani d'azione e modalità disumane per sbarazzarsi del genere maschile, come si è visto nel caso dei gruppi Incel. Come sottolinea Ling (2022: 100), il monitoraggio dei gruppi Femcel, a causa delle loro connessioni e origini nella comunità Incel, deve intendersi come un'importante precauzione, per ora. A tal fine è importante sviluppare nuove linee d'indagine che approfondiscano le ripercussioni offline dei discorsi d'odio che si generano online.



## Bibliography

Alava S., Nagem R. (2024). *Femcels speeches between extreme solitude and radical break with the Alpha world*, in "Hal.science".

Aronowitz N.W. (2021). The Femcel Revolution. Elle. <https://www.elle.com/life-love/sex-relationships/a37288143/0168-0169-the-femcel-revolution-september-2021/>.

Baker P. (2016). "The woman who accidentally started the Incel Movement", Elle. <https://www.elle.com/culture/news/a34512/woman-who-started-incele-movement/>.

Balci U., Ling C., De Cristofaro E., Squire M., Stringhini G., Blackburn J. (2023). *Beyond Fish and Bicycles: Exploring the Varieties of Online Women's Ideological Spaces*, 15th ACM Web Science Conference 2023 (ACM WebSci 2023), <https://doi.org/10.48550/arXiv.2303.07099>.

Banet-Weiser S. (2018). *Empowered: Popular Feminism and Popular Misogyny*, Duke University Press, Durham.

Bentivegna S., Rega R. (2020). "I discorsi d'odio online in una prospettiva comunicativa: un'agenda per la ricerca", in *Mediascapes journal*, vol. n.16: 151-171.

Bentivegna S., Boccia Artieri, G. (2019). *Le teorie delle comunicazioni di massa e la sfida digitale*. Laterza, Roma-Bari.

Bianchi C. (2021). *Hate speech*, Roma-Bari, Laterza.

Boccia Artieri G., Farci M. (2020). *Le emozioni dell'alt-right. La dimensione neoliberale e affettiva delle piattaforme*, in "Sociologia della comunicazione", 59, pp 83-107. DOI: 10.3280/SC2020-059005.

Bottici C. (2023). *Nessuna sottomissione* (Italian Edition). Editori Laterza. Edizione Kindle.

Burn S. M., Kadlec K., & Rexer R. (2005). "Effects of subtle heterosexism on gays, lesbians, and bisexuals", *Journal of Homosexuality*, vol. 49. n.(2): 23-38.

Carzo D., Cava A., Salvo M. (2007). *Globalità virtuale e realtà locale. Genere, consumi e comunicazione in una città di provincia*, Milano, FrancoAngeli..

Codeluppi V. (2012). "L'immaginario delle reti", *Im@go. Rivista di Studi Sociali sull'immaginario*, n.0: 19-28.



Crenshaw K. (1989). *Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory, and Antiracist Politics*, Chicago, University of Chicago Legal Forum.

De Gasperis A. (2021), "Giacomino uno di noi, Letteratura italiana e pratiche di maschilità nel Forum dei Brutti", in *AG About Gender-Rivista internazionale di studi di genere*, vol. 10 n.(19): 68-104.

Deliovsy K. (2008). "Normative White Femininity: Race, Gender and the Politics of Beauty", in *Atlantis* vol. n.33 n.(1): 49-59.

Dolce R., Pilla, F. (2019). *Il web che odia le donne*, Milano, Ledizioni.

Donnelly D., Burgess E., Anderson S., Davis R., Dillard J. (2001). "Involuntary celibacy: A life course analysis", *Journal of Sex Research*, vol. 38 n. (2): 159-169.

Dordoni A., Magaraggia S. (2021). "Modelli di mascolinità nei gruppi online Incel e Red Pill: narrazione vittimistica di sé, deumanizzazione e violenza contro le donne", *AG About Gender. International Journal of Gender Studies*, vol. 10 n.(19): 35-67.

Durand,G. (1960). *Les structures anthropologiques de l'imaginaire*, Paris, Dunod.

Erjavec K., Kovačič M. P. (2012). "You don't understand, this is a new war!" *Analysis of hate speech in news web sites' comments*, "Mass Communication and Society", vol. 15 n.(6): 899-920.

Evans H.R., Lankford A. (2024), "Femcel Discussions of Sex, Frustration, Power, and Revenge". *Arch Sex Behav.* <https://doi.org/10.1007/s10508-023-02796-z>.

Gagliardone I., Gal D., Alves T., & Martínez G. (2015). *Countering online hate speech*. UNESCO.

Ging D. (2019). "Alphas, Betas, and Incels: Theorizing the Masculinities of the Manosphere", *Men and Masculinities*, vol. 22 n.(4): 638-657.

Gotell L., Dutton E. (2016). "Sexual Violence in the 'Manosphere': Antifeminist Men's Rights Discourses on Rape", *International Journal for Crime, Justice and Social Democracy*, vol.5 n. (2): 65-80.

Hawley G. (2017). *Dare un senso all'alt-right*, New York, Columbia University Press.

Hennefield M., Sammond N. eds. (2020). *Abjection Incorporated: Mediating the Politics of Pleasure & Violence*. Durham, Duke University Press.



- Hesford V. (2005). "Feminism and its ghosts: The spectre of the feminist-as-lesbian", *Feminist Theory*, vol. 6 n. (3): 227-250 <https://doi.org/10.1177/1464700105057361>.
- Hine C. (2000). *Virtual Ethnography*, "Centre for Research into Innovation, Culture and Technology Brunel University", Uxbridge, Middlesex.
- Hine G., Onalapo J., Cristofaro E. D., Kourtellis N., Leontiadis I., Samaras R., Stringhini G., & Blackburn J. (2017). "Kek, cucks, and god emperor Trump: A measurement study of 4chan's politically incorrect forum and its effects on the web", Gonzalez-Bailon A., Marwick & W., Mason (Eds.), *Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSM, 2017)*, (92-101), Association for the Advancement of Artificial Intelligence (AAAI).
- Howard P. N. (2002). "Network Ethnography and the Hypermedia Organisation: New Media, New Organisations, New Methods", *New Media & Society*, vol. 4 n.(4): 550-74.
- Johnson N.F., R. Leahy N.J. Restrepo et al. (2019). "Hidden resilience and adaptive dynamics of the global online hate ecology", in *Nature*, vol. n. 573: 261-265.
- Kanai A. (2019). *Gender and Relatability in Digital Culture: Managing Affect, Intimacy and Value*, New York, Palgrave Macmillan.
- Kay J. B. (2022). "Abject desires in the Age of Anger. Incels, femcels and the gender politics of unfuckability", *Imagining "We" in the Age of "I". Romance and Social Bonding in Contemporary Culture*, a cura di Harrod, M., Leonard, S.& Negra, D. Routledge.
- Kohn I. (2020). *Inside the world of 'femcels', "Mel Magazine"*. <https://melmagazine.com/en-us/story/femcels-vs-incels-meaning-reddit-discord>.
- Lewis R. & Marwick A. (2017). <https://datasociety.net/library/media-manipulation-and-disinfo-online/>.
- Lily M. (2016). *The World Is Not a Safe Place for Men: The Representational Politics of the Manosphere* (Political Science, Université d'Ottawa/University of Ottawa), <https://www.ruor.uottawa.ca/handle/10393/35055>.
- Ling R. (2022). *Femcels: Are they really the female equivalent of the incel?* Master's thesis, Simon Fraser University.
- Mackay F. (2014). "Reclaiming Revolutionary Feminism", *Feminist Review*, vol. 106 n. (1): 95-103. <https://doi.org/10.1057/fr.2013.38>.



Mackinnon C. (1989). "Sexuality, pornography, and method: Pleasure under patriarchy", in *Ethics*, vol. 99 n.(2): 314–346. <https://doi.org/10.1086/293068>.

Miller D., Slater D. (2001). *The Internet: An Ethnographic Approach*, London, Berg Publisher.

Mishra, P. (2017). *The Age of Anger: A History of the Present*, London Penguin.

Monaci S. (2022). *Odio social* (Italian Edition), Milano, Egea.

Mondal M., Silva L. A., Correa D., Benevenuto F. (2018). "Characterizing usage of explicit hate expressions in social media", *New Review of Hypermedia and Multimedia*, vol. 24 n. (2): 110–130.

Morin E. (2005). *Lo spirito del tempo*, Roma, Meltemi.

Murthy D. (2008), "Digital Ethnography: An Examination of the Use of New Technologies for Social Research", in *Sociology*, vol. 42 n. (5): 837–55.

Nagle A. (2017). *Uccidi tutti i normies: le guerre culturali online da Tumblr e 4chan all'alt-right e Trump*, Roma, Zero libri.

Nagle A. (2018). *Contro la vostra realtà: come l'estremismo del web è diventato mainstream*, Roma, LUISS University Press.

Papacharissi Z. (2015). *Affective Publics: Sentiment, Technology, and Politics*, New York Oxford University Press.

Papacharissi Z. (2021). *After Democracy: Imagining Our Political Future*, Yale University Press.

Papacharissi Z., de Fatima Oliveira M. (2012). "Affective News and Networked Publics: The Rhythms of News Storytelling on #Egypt", *Journal of Communication*, vol. n. 62: 266–82.

Patton M. (1990). *Qualitative evaluation and research methods*, Beverly Hills, Sage.

Massanari A.L. (2017). "#Gamergate and The Fapping: How Reddit's algorithm, governance, and culture support toxic technocultures", *New Media & Society*, vol. n. 19: 329-346.

Pizzimenti D. & Pasciuto F. (2022). "Un'analisi netnografica sul fenomeno dell'hate speech nella manofera italiana", in *Metis. Ricerche di sociologia, psicologia e antropologia della comunicazione*, vol. XXIX: 1-2.



Rieger D., Kümpel A.S., Wich M., Kiening T. and Groh G. (2021). "Assessing the Extent and Types of Hate Speech in Fringe Communities: A Case Study of Alt-Right Communities on 8chan, 4chan, and Reddit", *Social Media + Society*, October-December: vol.1 n. (14), Sage, DOI: 10.1177/20563051211052906.

Santerini M. (2021). *La mente ostile*, Milano, Raffaello Cortina Editore.

Sumiala J., Tikka M., (2020). "Digital ethnographers on the move – An unexpected proposal", *Journal of Digital Social Research*, vol. 2. n. (1): 39-55.

van Dijck, J. (2013). *The Culture of Connectivity: A Critical History of Social Media*, Oxford, Oxford University Press

van Dijck J., Poell T, De Waal M. (2018). *The Platform Society: Public Values in a Connective World*, Oxford University Press, Oxford; trad. it. (2019) *Platform Society: Valori pubblici e società connessa*, Guerini e Associati, Milano.

Wu K. (2022). "How do femcels and incels differ? Gender socialization may shape the way we react to sexual rejection", *Psychology Today*. <https://www.psychologytoday.com/us/blog/the-modernheart/202211/how-do-femcels-and-incels-differ>.

Ziccardi G. (2016). *L'odio online: Violenza verbale e ossessioni in rete*, Milano, Raffaello Cortina.







A JOURNAL OF THE  
SOCIAL IMAGINARY

# Stigma e gestione dell'identità nelle piattaforme digitali

Assunta Penna  
apenna@unime.it  
COSPECS | University of Messina



## Abstract

*Stigma and identity management on digital platforms.*

Human life, through social changes resulting from the widespread and continuous use of digital technologies, expresses boundless ways of experiencing the world through a series of significant and complex transformations that manifest in the interactions of societal structures. The opportunities presented by digital interconnectedness can have critical implications for human relationships, fostering expressions of feelings of loneliness, alienation, and violence in some spaces of "connectivity."

Through an analysis of the platformized structures of the manosphere, in this contribution we reflect on the forms in which 'identities' manifest in the imagery of digital environments. We analyze how the imagery of online groups or communities (Incel and Femcel in particular) shape the individual experiences and roles exhibited by men and women who inhabit these spaces of digital life. Finally, we will highlight how this sense of inadequacy sometimes translates into processes of internalizing a stigma or, conversely, violent stigmatization.

## Keywords

Online Community | Platformation | Digital Platform | Identity



## 1. Introduzione

L'articolo propone una riflessione sulle pratiche di stigmatizzazione e definizione delle identità all'interno delle subculture digitali, concentrandosi sul ruolo cruciale delle piattaforme digitali connesse e dei gruppi online. Dopo aver introdotto i concetti chiave, l'analisi esplora come le identità di gruppo e le percezioni di stigma vengano costruite, negoziate e riprodotte in questi ambienti. Al centro della riflessione vi è l'importanza del contributo di Erving Goffman, la cui opera fornisce un quadro teorico e metodologico capace di interpretare le dinamiche sociali contemporanee, anche degli ambienti digitali. La valenza euristica delle teorie di Goffman, come per le teorie sociologiche più consolidate, risiede nella loro capacità di offrire, un apparato concettuale-metodologico e una struttura analitica flessibile e adattabile, valevole non solo per i contesti sociali in cui sono emerse, ma anche per quelli emergenti. Sebbene sviluppate in un'epoca precedente all'avvento del digitale, i concetti goffmaniani trovano applicazioni nel presente, permettendo di analizzare come stigma e identità vengano costruiti, negoziati e trasformati attraverso interazioni sociali incluse quelle mediate dalle tecnologie connesse.



## 2. Gruppi digitali

Se il dibattito accademico sulle piattaforme digitali si colloca ormai in uno spazio centrale della ricerca, gli studi sia empirici che teorici sui gruppi incel e femcel sono in continua evoluzione ed ancora pochi sono quelli che li indagano in relazione all'immaginario, ovvero all'insieme delle credenze, delle idee, delle aspettative e dei valori che una società condivide e trasmette attraverso varie forme di comunicazione e socializzazione (Morin, 2005; Cardoso, 2006). Le subculture incel (Dolce, Pilla, 2019; Nagle 2018; Ging, 2017; Farci, Righetti 2019; Vingelli, 2019)<sup>1</sup>, e femcel (Kay, 2022; Alava, Nagen 2024; Balci et al., 2023) sono gruppi di persone che emergono in risposta a percezioni specifiche e spesso distorte delle dinamiche di genere e delle relazioni sociali (Cava, Pizzimenti, 2024)<sup>2</sup>.

<sup>1</sup> Con il neologismo *manosfera*, come è noto, si fa solitamente riferimento a un complesso sistema di comunità, gruppi e piattaforme interconnesse, che includono realtà diverse (Scarcelli, 2021). Per un approfondimento sulla definizione di *manosfera*, oltre gli studi citati, si rimanda a: Cannito, M., Crowhurst, I., Ferrero Camoletto, R., Mercuri, E., & Quaglia, V. (2021). Fare maschilità online: definire e indagare la *manosphere*. *AG-ABOUT GENDER*, 10(19), 1-41; Marwick, A.E. e Caplan, R. (2018), *Drinking Male Tears: Language, the Manosphere, and Networked Harassment*, in *Feminist Media Studies*, vol. 18, n. 4, pp. 543-559.

<sup>2</sup> Negli ultimi anni, i gruppi Incel sono cresciuti in dimensioni e radicalizzazione grazie alle piattaforme digitali, per una introduzione alle caratteristiche di numerosità in Europa si veda: European Commission (2021), *Incels: A First Scan of the Phenomenon (in the EU) and its Relevance and Challenges for P/CVE*. Per le comunità Femcel, si veda: Balci, U., Ling, C., De Cristofaro, E., Squire, M., Stringhini, G., Blackburn, J. (2023), *Beyond Fish and Bicycles: Exploring the Varieties of Online Women's Ideological Spaces*, Published in the Proceedings of the 15th ACM Web Science Conference 2023 (ACM WebSci 2023), <https://doi.org/10.48550/arXiv.2303.07099>.



Nella letteratura accademica, il gruppo maschile più noto e ampiamente studiato di individui che vivono una condizione di celibato involontario è indubbiamente quello degli incel (Dolce, Pilla 2019; Nagle, 2018; De Gasperis, 2021). Il termine “incel” esprime la loro condizione, risultando dalla fusione delle parole “celibe” e “involontario”. Ciò che caratterizza il movimento Incel è un sentimento di vittimizzazione, poiché si percepiscono come celibi contro la loro volontà.

Questi soggetti attribuiscono la loro solitudine e l'assenza di relazioni romantiche o sessuali alle donne, descrivendosi come poco attraenti, di basso status socioeconomico e anche come uomini beta (Ging, 2019), ritenendosi quindi meno desiderabili rispetto alle loro controparti femminili. All'interno di questi gruppi radicali, emerge la nozione di una crisi della mascolinità, attribuita all'empowerment femminile e alla dominanza delle donne nella società occidentale (Ciccione, 2019).

Il fenomeno Incel, considerato parte del più ampio movimento della manosphere, si riferisce a un movimento creato da uomini per altri uomini, caratterizzato da una varietà di sottoculture interconnesse (Lilly, 2016). Sebbene questi movimenti abbiano obiettivi diversi, sono collegati e alimentati da sentimenti antifemministi, misogini e spesso violenti (Ging, 2019; Vingelli, 2019; Farci, Righetti 2019; Pizzimenti, Pasciuto 2022). Radicati in una concezione egemonica della mascolinità, questi gruppi si sono gradualmente espansi e diversificati nel tempo, dando origine a vari gruppi come i Pick Up Artists (PUA), i Men's Rights Activists (MRA), i Men Going Their Own Way (MGTOW) e gli Incel. Tra i concetti fondamentali della comunità incel, è rilevante la teoria della Pillola Rossa, originata su Reddit nel 2012 (Farci, Righetti, 2019; Cannito, Mercuri, 2021; Van Valkenburgh, 2021).

Il termine si origina dall'iconica pellicola “The Matrix” (1999), la cui potenza simbolica si esprime nella estrema dualità della contrapposizione tra le forze del bene e del male, il potere del reale e dell'ingannevole. Al protagonista viene data la scelta tra una pillola blu, che gli permetterebbe di rimanere nel mondo sicuro ma illusorio che conosce, e una pillola rossa, che lo porterebbe a scoprire la dura realtà nascosta (Farci, Righetti, 2019). La “filosofia delle pillole” utilizza metafore per descrivere diverse visioni della realtà, in particolare riguardo alle dinamiche di genere e alle relazioni romantiche e sessuali. In Matrix, ingoiare la pillola rossa significa risvegliarsi alla verità nascosta sulla realtà. Dentro la scena dell'immaginario ispirata alla pellicola, ingoiare la Pillola Blu, se in Matrix permette di rimanere nell'illusione confortevole della realtà, essere “blue-pilled”, nella manosphere, significa vivere nell'illusione della realtà di essere, ovvero scegliere di ignorare o negare le modalità dell'esistere. I seguaci di questa teoria, quindi, tendono a dividere la società in due sfere: coloro che hanno preso la pillola rossa, risvegliati all'idea che la società promuova una grande falsità, dove le donne detengono i maggiori privilegi in modo occulto; e coloro che rimangono sotto l'effetto della pillola blu, inconsapevoli di vivere in un mondo dominato dal potere sessuale femminile.

In controcena rispetto ai più noti e riconosciuti movimenti maschili sono le femcel, il cui nome deriva dalla combinazione di “femmina” e “celibe involontaria”. Il movimento delle femcel rimanda a caratteristiche consonantiche su alcuni punti a quelle degli incel, ma presenta differenze significative. Le differenze tra i gruppi incel



e femcel emergono chiaramente nelle loro reazioni e azioni alla solitudine e alla discriminazione percepita. Secondo Alava & Nagen (2024), se gli incel rispondono alla frustrazione spesso con ostilità e discorsi polarizzati, le femcel cercano supporto all'interno delle loro comunità, concentrandosi su solidarietà e auto-miglioramento.

Gli incel, spinti dalla frustrazione, tendono a esprimere la loro solitudine attraverso un discorso intriso di misoginia.

Questa reazione può essere accompagnata da ostilità e ideologie estreme, come dimostra la loro adesione alla "pillola nera", una visione pessimistica delle dinamiche di genere. La loro narrativa spesso include accuse dirette verso le donne e la società in generale, manifestando la loro rabbia attraverso comportamenti aggressivi e, in alcuni casi, violenti. Al contrario, le femcel trovano conforto e sostegno all'interno delle loro comunità, impegnandosi in un discorso che, pur riconoscendo la discriminazione e la solitudine, mira alla solidarietà e all'aiuto reciproco. Invece di esternalizzare la loro frustrazione, tendono a interiorizzare le loro lotte, concentrandosi sul miglioramento di sé e sul superamento delle norme di bellezza convenzionali. Le discussioni delle femcel spesso ruotano attorno a temi legati al genere, alle aspettative sociali, agli standard di bellezza e alle dinamiche di potere, con un'enfasi particolare sul ruolo della bellezza dal punto di vista femminile.

Queste dinamiche riflettono approcci distinti alle sfide della stigmatizzazione e della solitudine, mostrando come i contesti e le reazioni possano variare significativamente tra i due gruppi.

Similmente agli Incel, anche le donne Femcel sono particolarmente concentrate sull'aspetto fisico, considerando la loro mancanza di avvenenza fisica come la principale fonte della loro solitudine e discriminazione sociale, poiché convinte che gli uomini siano attratti esclusivamente dalla bellezza fisica delle donne. All'interno di questa subcultura femminile esiste una ricca varietà di sottogruppi, ciascuno spesso significativamente diverso dagli altri. Tra le ideologie femcel, ad esempio, si trovano le donne pinkpillate, che traggono ispirazione dal movimento femminista radicale e promuovono iniziative come "Women Going Their Own Way".

Il movimento della Pillola Rosa emerge prevalentemente negli spazi online, con l'obiettivo di fornire alle donne spiegazioni e condividere consigli riguardo alle dinamiche delle relazioni romantiche e al comportamento maschile (Evans, Lankford, 2023; Balci et al., 2023; Cava, Pizzimenti, 2024).

Esso sfida gli stereotipi di genere consolidati, incoraggiando le donne ad affermare i propri diritti, desideri e bisogni all'interno delle relazioni romantiche e della società. In contrapposizione alla teoria della Pillola Rossa associata alle comunità incel, il movimento della Pillola Rosa incoraggia l'empowerment femminile e un esame critico delle dinamiche di genere (Kay, 2022).

Sostiene le donne nel vivere le loro relazioni romantiche e sociali con consapevolezza, promuovendo la determinazione nel perseguire i propri desideri e rifiutando norme dannose imposte dalla società patriarcale. Centrale a questa missione è l'educazione delle donne a distinguere tra individui tossici e coloro capaci di instaurare relazioni sane.



Nonostante le sfide specifiche di genere, sia gli incel che le femcel trovano rifugio nelle comunità online, dove cercano approvazione, supporto e comprensione mentre condividono le loro esperienze di solitudine e frustrazione. Queste “connective societies”, caratterizzate dalla propensione a essere online e “networked” (Boccia Artieri et al., 2017, in Bentivegna, Boccia Artieri, 2019), tendono a vivere questo stato di relazione come una condizione normale piuttosto che come un’eccezione. Questo ambiente virtuale offre a incel e femcel un senso di appartenenza che manca loro nella vita offline.

Gli spazi online abitati da individui con pensieri affini diventano ambienti dove si sviluppa un immaginario dominato da un senso di esclusione e alle volte di rabbia, emozioni che trovano sfogo in tali nicchie online, dove la condivisione di esperienze comuni rafforza la loro percezione di essere vittime di un’ingiustizia sociale sistemica. La dimensione online e le caratteristiche delle piattaforme su cui esse si esprimono, consentono una continua interazione e alimentano una narrativa condivisa, che contribuisce a consolidare le identità di gruppo e un sostegno alle convinzioni.

Le piattaforme digitali connesse, quindi, non sono semplicemente spazi di socializzazione, ma diventano veri e propri incubatori di ideologie che possono avere implicazioni significative sul comportamento e sulle percezioni degli individui coinvolti. La continua condivisione di contenuti all’interno di queste comunità riflette il desiderio degli utenti di comunicare la propria individualità agli altri, contribuendo alla formazione e alla consolidazione di identità di gruppo.



### 3. Identità digitali

All’interno degli ambienti delle piattaforme digitali, si assiste a una coesistenza eterogenea di utenti, ognuno dei quali interagisce con le piattaforme in base a una prospettiva soggettiva individuale. Questa diversità di prospettive crea un ambiente ricco e variegato, in cui le esperienze digitali sono plasmate dalle peculiarità, dagli interessi e dagli obiettivi personali di ciascun utente. Emerge, inoltre, come l’interconnessione tra l’uso delle piattaforme e i differenti approcci ai media e all’innovazione tecnologica, spesso sia influenzata dalla dimensione generazionale degli utenti. Le diverse generazioni portano con sé differenti atteggiamenti, abitudini e competenze nell’utilizzo delle tecnologie digitali e nei social media, creando così uno spaccato dinamico e in continua evoluzione delle interazioni online. In questo contesto, la convergenza di prospettive individuali e influenze generazionali contribuisce a definire l’ambiente online come uno spazio digitale interattivo e fluido.

Tale coesistenza di diversità, tanto a livello individuale quanto generazionale, offre un terreno fertile per l’emergere di nuove dinamiche sociali, culturali e comunicative, contribuendo così alla continua evoluzione delle piattaforme e delle interazioni digitali. Per la studiosa Vittadini (2018: 164), ad esempio, negli spazi dei siti di social network, vi è una compresenza di generazioni: «ciascuna di esse li utilizza a partire da una semantica generazionale differente che comprende anche un diverso approccio ai media e all’innovazione tecnologica».



Nel suo senso più generale, la partecipazione come "utenti" alle piattaforme digitali (in particolare quelle che offrono o esigono la creazione di profili individuali) è strettamente legata alla rappresentazione di sé e all'interazione sociale.

Come è noto, il fondativo "modello drammaturgico sui labili confini tra realtà e finzione" di Goffman (1959) in *The Presentation of Self in Everyday Life*, indaga il modo in cui le persone gestiscono le "impressioni" nelle interazioni sociali quotidiane. Sebbene le intuizioni teoriche di Goffman si concentrino sulla rappresentazione di sé nelle interazioni sociali al di fuori degli ambienti digitali, come sottolinea Pentecoste (2013: 109) - in una riflessione sull'influenza complessa e contraddittoria che l'attuale ambiente mediatico esercita sulla costruzione dell'identità - non sono assenti nella teoria goffmaniana riflessioni sulla comunicazione mediata.

Pensiamo ai concetti di front stage e back stage. Il primo, il palcoscenico, descrive la parte della vita in cui le persone si esibiscono agli altri, una fase pubblica della rappresentazione di sé, in cui le persone durante le interazioni sociali gestiscono le "impressioni" che gli altri rivolgono a loro; il secondo, il retroscena, rappresenta la parte più intima, privata e meno visibile della vita, uno spazio maggiormente autentico, in cui gli aspetti del sé non sono destinati al pubblico.

Per Goffman, «il repertorio di comportamenti da ribalta non può comprendere tutto ciò (e in certo senso ne è quindi l'opposto). In generale quindi il comportamento da retroscena è quello che permette atti secondari che facilmente possono esser presi come segno di intimità e mancanza di rispetto nei confronti degli altri presenti e del territorio, mentre il comportamento da ribalta è quello che non permette atti potenzialmente offensivi» (1959: 149).

Già Miller (1995: 2-8) sostiene la versatilità della teoria goffmaniana, guardando all'interazione "elettronica" e al sé "elettronico":

Much of Goffman's interest is in his analysis of the depth and richness of everyday interaction. This depth and richness is perhaps not apparent in electronic interaction, but the problem of establishing and maintaining an acceptable self remains, and there is a range of expressive resources available for this end. As the technology develops, more expressive resources become available. Also, as the culture of electronic communication develops, people will construct expressive resources out of whatever facilities are available. Electronic communication will become more and more human communication to the extent that there is more to it than just efficiently passing information to each other (Miller, 1995: 2).

My feeling, as an old-fashioned psychologist, is that sociality and interaction are necessary for us to know who we are and what we can say about ourselves to others, and much more depth and richness in EC is needed before 'electronic selves' can emerge. Contrariwise, much of the depth and richness that we can think of adding would be to make EC more like face-to-face interaction, which might suggest that electronic selves could be developed in a different social context (continuing the extension of the social world from the hamlet to the city



to the global village) but that many of the basic issues, moves and processes that go on would be the same as they always were (Miller, 1995: 8).

La dimensione pubblica (front stage) e quella più privata (back stage), con particolare attenzione alla gestione delle impressioni, possono essere attualizzate alla vita online. Ad esempio, gli individui navigano negli spazi digitali praticando nuove forme di interazioni sociali, gestendo la loro presentazione di sé in modi che rispondono a aspettative sociali e costruendo significati attraverso i simboli e le rappresentazioni; negli spazi condivisi online i profili sui social media possono essere considerati front stage, dove gli individui attraverso simboli e segnali presentano una versione di sé adulterata, indossando delle "maschere" per nascondere parti di sé o per presentare un'immagine scelta per gli altri, mentre le conversazioni più private o autentiche possono essere riesiliate nel back stage.

Ci troviamo di fronte a nuovi spazi e rituali di interazione. Secondo Airoidi «tutte le comunicazioni Internet mediate sono volontarie e - in quanto tali - vanno trattate come "presentazioni del self" (si veda Goffman 1959), manifestazioni identitarie legate a uno specifico palcoscenico sociale (digitale), le quali non necessariamente corrispondono alla verità dei fatti» (Airoidi, 2017: 5). Pensiamo ai "selfie", scattati e pubblicati per un pubblico digitale, che per la sociologa Esposito (2022), in Comunicazione artificiale:

sono un tipico esempio di fotografie sociali [...]. Noi costituiamo la nostra identità vedendoci con gli occhi degli altri, ma ora ciò che gli altri vedono di noi è l'immagine che noi scegliamo di presentare, e che spesso è stata elaborata con strumenti software: «vedo me che ti mostro me stesso». L'utente di queste tecnologie digitali configura attivamente una presentazione del sé, che diventa la base delle osservazioni esterne (like, tag, follower e altre forme di feedback digitale) da cui l'utente apprende chi è (Esposito, 2022: 98).

Gli individui creano e gestiscono le loro identità online; creano le comunità che si formano e si sviluppano in rete (e le pratiche e le culture online a loro volta influenzano e sono influenzate dalla vita offline) (Papacharissi, 2010); gestiscono attraverso i social media dei movimenti sociali (Castells, 2012); i più giovani formano la loro costruzione identitaria, le relazioni sociali e le loro interazioni con gli adulti e con le istituzioni (boyd, 2014).

Come abbiamo già accennato, Goffman teorizza nelle sue opere le dinamiche sistemiche, sociali e individuali delle interazioni sociali al di fuori degli ambienti digitali, tali concetti teorici riflettono una grande attualità, che nelle prossime pagine proveremo a dimostrare usando, in particolare, il concetto di stigma.

Una riflessione sulla definizione concettuale dello "stigma" e dei processi sociali di stigmatizzazione può fornire una cornice utile per comprendere e analizzare le dinamiche identitarie individuali e di gruppo anche nell'ambiente digitale abitato dai movimenti incel e femcel.

In questa riflessione ci serviremo solo di alcuni aspetti teorici, in particolare dei concetti di setting (contesto dell'interazione), virtual social / actual social identity (le



identità), self (il sé rappresentato) e stigma, slegandoli dagli esempi/casi reali connotati storicamente, cui Goffman si riferisce, e mettendoli alla prova di questi contesti digitali.

L'analogia si basa, in primo luogo, sull'architettura degli ambienti digitali, che consentono nuovi spazi (setting) e modalità di interazione delle identità agite (virtual social / actual social identity), che possono amplificare, trasformare o attenuare le dinamiche di stigmatizzazione (stigma). In secondo luogo, l'assonanza si manifesta tra la descrizione di alcuni stati emotivi e psicologici dei gruppi stigmatizzati nel contesto tradizionale e quelli osservabili nei contesti digitali, che hanno ripercussioni sul sé rappresentato (self).

Le piattaforme sono lo spazio, il setting in cui si svolge l'interazione. Le prime considerazioni, sebbene circoscritte, si possono avanzare sulla gestione delle identità on-line nel contesto delle architetture delle piattaforme, che assicurano ad esempio forme di anonimato. Gli utenti possono scegliere di rimanere anonimi o utilizzare pseudonimi, il che permette loro di partecipare a discussioni senza rivelare aspetti stigmatizzanti della loro identità.

L'immagine di sé può presentarsi in modi specifici e individuali, selezionando attentamente le informazioni e le immagini che condividono, per gestire le impressioni degli altri e mitigare, annullare o accentuare, a seconda dei contesti, l'effetto dello stigma. L'adesione a gruppi on-line attraverso identità celate consente di ritrovarsi, unirsi a gruppi di sostegno online dove condividere esperienze simili e ricevere supporto emotivo, riducendo il senso di isolamento individuale e alle volte alimentando quello del gruppo. Le piattaforme online possono fornire spazi "sicuri" dove le persone possono esprimersi "liberamente" senza paura di giudizio o discriminazione da parte di utenti oppositori; o ancora dichiarare e condividere aspetti "degradanti" di sé o di altri avendo facoltà di usare le piattaforme online per sensibilizzare o organizzare forme di attivismo digitale o campagne di advocacy.

Si può "scegliere" una identità più o meno aderente all'identità personale, espandendo sempre di più la possibilità di soddisfare le aspettative altrui, attraverso ad esempio la condivisione di storie personali. La presenza online espone anche a nuove forme strategiche di gestione delle identità digitali che possono alimentare o disinnescare circostanze degradanti e discriminatorie. L'effetto di risonanza immediato di questo tipo di "linguaggio della relazione" è che le narrazioni veicolate e interscambiate, in tempi e modalità non sincronicamente stabilite, possono diffondersi rapidamente, accelerando gli stati emotivi e cognitivi di cui si nutre questo immaginario. Le piattaforme come architetture on-line non sono neutrali, occupano una parte della scena, contribuendo a definire il "linguaggio della relazione", è pertanto cruciale che esse implementino politiche efficaci contro la violazione e la discriminazione.

Goffman, nel suo lavoro "La vita quotidiana come rappresentazione" (1969), come è noto, esplora il concetto di "self" come un prodotto delle interazioni sociali. Secondo Goffman, l'immagine del sé che emerge durante le interazioni è definita dai partecipanti alla situazione interazionale e scaturisce dagli attributi degli eventi locali, che la rendono comprensibile ai testimoni. Gli utenti possono utilizzare le tecnologie



digitali per presentarsi in modi che enfatizzino le loro competenze, passioni e qualità positive, contrappesando gli aspetti stigmatizzati della loro identità. Gli individui con "attributi stigmatizzanti" possono sperimentare un "self degradato", influenzando negativamente la loro accettazione nelle interazioni sociali. Il self, inteso come effetto drammaturgico che emerge dalla scena rappresentata durante un'interazione, è una costruzione sociale che dipende dalla credibilità della scena rappresentata. Goffman esplora come queste identità degradate gestiscano la loro presenza nell'interazione e come navigano l'incertezza riguardo alla loro accettazione sociale.

Con l'introduzione dei concetti di "identità sociale virtuale" e "identità sociale effettiva" il teorico decodifica come si manifesta lo stigma. Per identità sociale virtuale si intende come una persona è riconosciuta, caratterizzata e quindi definita dalla società, tale processo può essere condizionato da stereotipi e pregiudizi.

L'identità sociale effettiva, invece, è costituita dagli attributi realmente posseduti da una persona. Lo stigma si verifica quando c'è una discordanza tra queste due identità. Sebbene lontano dalle previsioni di Goffman, che sperimentò la molteplicità di espressioni dell'identità degli individui nelle interazioni offline, analogamente Iaconesi<sup>3</sup>, nel "Nuovo abitare" ci restituisce un modello di identità digitale dalle numerose declinazioni. L'identità digitale può essere individuale, anonima, collettiva, temporanea, transitiva, e remix di tutte queste modalità.

Per comprendere l'applicabilità della nozione di stigma all'interno del nostro contesto, cerchiamo di circoscriverne il campo definitorio. Che cos'è uno stigma? Il termine stigma, scrive Goffman:

sarà impiegato per riferirci a un attributo che è profondamente discreditante, ma va tenuto presente che ciò che conta davvero è il linguaggio delle relazioni, e non quello degli attributi. Un attributo che stigmatizzi un certo tipo di portatore può confermare la regolarità di un altro, e dunque, in quanto tale, non è in sé né screditante, né meritorio [...] uno stigma è, in realtà un tipo particolare di rapporto tra attributo e stereotipo (Goffman, 1963: 29-31).

Questo significa che uno specifico attributo di una persona (ad esempio, per Goffman una patologia mentale) diventa stigmatizzante quando viene associato a uno stereotipo negativo diffuso nella società. L'attributo in sé non è intrinsecamente stigmatizzante, è nel contesto sociale e culturale che se ne connota il suo significato.

Lo stigma è incorporato in un "linguaggio delle relazioni", il che implica che lo stigma non è solo una qualità personale o una caratteristica, ma una costruzione sociale che emerge attraverso le interazioni e le relazioni tra gli individui.

Questo linguaggio delle relazioni coinvolge termini, simboli e comportamenti che rafforzano e perpetuano lo stigma. Ad esempio, l'uso di determinati epiteti o il comportamento di evitamento verso le persone stigmatizzate fanno parte di questo linguaggio.

<sup>3</sup> Cfr. Salvatore Iaconesi (2017), "I Principi del Nuovo abitare", in *Medium* <https://xdxd-vs-xdxd.medium.com/i-principi-del-nuovo-abitare-39aa6ba9c74b>.



Nel suo libro "Stigma: Notes on the Management of Spoiled Identity" (1963), Goffman, identifica tre tipi distinti di stigma (ivi: 30), ciascuno dei quali può avere un impatto significativo sulle interazioni sociali e sull'identità individuale. Questi tipi di stigma possono essere strumenti teorici rilevanti, almeno in parte, per comprendere le dinamiche di gruppi che manifestano le loro identità digitali all'interno di nicchie online, che interiorizzano lo "stigma" e che manifestano con rabbia, violenza o solitudine la loro condizione di isolamento e frustrazione.

Ai fini della nostra logica argomentativa, li categorizzeremo, semplificandoli, in "fisici; caratteriali; sociali". Il primo tipo di stigma riguarda le "diversità" fisiche del corpo. Questo può includere vastissime caratteristiche corporee. Gli individui con questi attributi fisici possono essere stigmatizzati o interiorizzare lo stigma ed essere o sentirsi percepiti negativamente dalla società, sperimentando l'isolamento o la discriminazione a causa delle loro percezioni in relazione alle differenze corporee.

Nel contesto degli incel e delle femcel, lo stigma legato all'aspetto fisico può giocare un ruolo cruciale, poiché le norme di bellezza e l'aspetto fisico, l'immagine di sé e dell'altro sono spesso al centro delle loro narrazioni e discussioni che condividono nei gruppi a cui aderiscono on-line. Le femcel, ad esempio, che percepiscono una mancanza di bellezza e un senso di emarginazione all'interno della società, sviluppano una significativa ossessione per le canonizzate figure ideali femminili, spesso chiamate "Stacy", analogamente, gli incel alimentano sentimenti di frustrazione aspirando ad essere come un ideale "Chad"<sup>4</sup>.

Il secondo tipo di stigma include le "imperfezioni del carattere", che comprendono alcuni attributi come i fallimenti personali o l'incapacità caratteriale di compiere degli obiettivi e possono portare a valutazioni severe rispetto alla percezione di sé nella società. Gli incel, per esempio, possono costruire uno stigma a partire dalla loro condizione di celibato involontario causato dall'incapacità di creare relazioni intime, e le femcel, allo stesso modo, possono affrontare la condizione di nubili involontarie con ripercussioni sull'autostima e sull'adeguatezza sociale.

Il terzo tipo di stigma è lo "stigma tribale", che come abbiamo dichiarato inizialmente, intercettiamo nell'appartenenza a gruppi sociali. Questo tipo di stigma è basato su caratteristiche di gruppo che sono ereditate e non possono essere modificate, e può portare a discriminazione sistematica e discredito. Sebbene questo tipo di stigma possa non essere centrale nelle dinamiche specifiche degli incel e delle femcel, può comunque influenzare le loro esperienze del sé, come ad esempio, fare parte di gruppi etnici o religiosi che affrontano discriminazione.

I tre tipi di stigma - da noi rielaborati come fisici, caratteriali e sociali - sebbene non trattino in modo compiuto ed esaustivo le declinazioni cognitive ed emotive del codice del linguaggio delle relazioni incel e femcel - offrono una cornice utile per comprendere in un senso complessivo le esperienze e le reazioni legate



<sup>4</sup> Per un approfondimento dei termini "Chad" e "Stacy", si veda: Menzie, L. (2022). Stacys, Beckys, and Chads: The construction of femininity and hegemonic masculinity within incel rhetoric. In *Critical Femininities* (pp. 69-85). Routledge; Dabral, A. (2024). The Actions of a Lonely Woman and the Effects of Online Incel Communities on Society; Meier, M. L., & Sharp, K. (2024). Death to Chad and Stacy: Incels and anti-fandom as group identity. *International Journal of Cultural Studies*, 27(3), 349-367.



all'interiorizzazione di uno stigma. Lo stigma interiorizzato influenza profondamente il modo in cui questi individui percepiscono sé stessi e modellano le loro interazioni.

#### 4. Conclusioni

L'approccio di Goffman allo stigma, a nostro parere, fornisce una cornice utile per comprendere e analizzare le dinamiche dello stigma anche nell'ambiente digitale. Più che dimostrarne l'applicabilità in termini assoluti, questo approccio ci consente di attingere alle robuste teorie sociologiche, che offrono strumenti interpretativi della società contemporanea. In particolare, ci permette di esaminare i fenomeni mettendo alla prova concetti e metodologie interpretative.

Nelle intercapedini concettuali, la gestione dell'identità, la creazione di comunità di supporto, il senso di frustrazione, la rabbia, l'isolamento, la depressione, le manifestazioni violente, che si scoprono nelle interazioni online, sono tutti aspetti che possono essere esplorati e affrontati interrogando le identità, che multiple e cangianti trovano il loro ancoraggio nella vita quotidiana.

Dalla nostra riflessione emerge come sessant'anni fa come oggi, il percepirsi stigmatizzati o normali nell'interazione è una conseguenza del carattere normativo dell'idea sociale di normalità, che definisce allo stesso tempo il deviante. Dall'interiorizzazione dello stigma di incel e femcel emergono stati emotivi e cognitivi che si manifestano nel primo caso in modo centrifugo, colpevolizzando gli altri per la loro condizione, e nel secondo in modo centripeto condannando sé stesse.

È l'interazione che determina il posizionamento come normale o stigmatizzato, non l'attributo in sé che, come abbiamo visto, può riferirsi ad un tratto fisico, ad un aspetto caratteriale, ad una appartenenza sociale. È l'interazione che costituisce quindi il nucleo del dispositivo dello stigma.

La ricostruzione teorica di queste pagine ha una doppia valenza: da una parte dimostrare l'attualità degli strumenti teorici goffmaniani nello studio di micro-comunità appartenenti alla società digitale, dall'altra mettere in luce l'importanza di ripensare alla costruzione di piattaforme digitali che dovrebbero essere plurali e democratiche e infine studiare i movimenti incel e femcel, esplorando la genesi della loro frustrazione.



## Bibliography

Alava S., & Nagem R. (2024), *Femcels speeches between extreme solitude and radical break with the Alpha world*, Counter projekt monograph.

Airoldi, M. (2017), "Oltre il virtuale: Studiare la società attraverso i dati digitali". In Paolo Natale e Massimo Airoldi (a cura di) *Web & social media. Le tecniche di analisi*. Santarcangelo di Romagna, Maggioli Editore.

Balci, U., Ling, C., De Cristofaro, E., Squire, M., Stringhini, G., Blackburn, J. (2023), *Beyond Fish and Bicycles: Exploring the Varieties of Online Women's Ideological Spaces*, Published in the Proceedings of the 15th ACM Web Science Conference 2023 (ACM WebSci 2023), <https://doi.org/10.48550/arXiv.2303.07099>.

Bentivegna, S., Boccia Artieri, G. (2019), *Le teorie delle comunicazioni di massa e la sfida digitale*, Roma-Bari, La Terza.

Boccia Artieri, G. Marinelli, A., & Boccia Artieri, G. (2019), "Per un'«economia politica» delle piattaforme". In *Platform society. Valori pubblici e società connessa*, Milano, Guerini.

Boccia Artieri, G. & Farci, M. (2020), "Le emozioni dell'alt-right. La dimensione neoliberale e affettiva delle piattaforme", in *Sociologia della comunicazione*, 59, pp 83-107. DOI: 10.3280/SC2020-059005.

boyd, d. (2014), *It's complicated: The Social Lives of Networked Teens*, Yale University Press.

Cannito, M., Crowhurst, I., Ferrero Camoletto, R., Mercuri, E., & Quaglia, V. (2021), "Fare maschilità online: definire e indagare la manosphere", in *AG-ABOUT GENDER*, 10(19), 1-41.

Cannito, M., & Mercuri, E. (2022). "Fatherhood and gender relations in the manosphere: Exploring an Italian non-resident fathers' online forum", in *European Journal of Cultural Studies*, 25(4), 1010-1029.

Cardoso, G. (2006), *The Media in the Network Society: Browsing, News, Filters and Citizenship*, Lisbona, CIES.

Castells, M. (1996), *The Information Age – The Rise of the Network Society*, Oxford, Blackwell.

Castells, M. (2012), *Networks of outrage and hope: Social movements in the Internet age*. Cambridge, Polity Press.



Cardoso, G. (2006), *The Media in the Network Society: Browsing, News, Filters and Citizenship*, Lisbon, CIES.

Cava, A., Pizzimenti, D. (2024), "L'odio indicibile: un'indagine etnografica digitale sulle comunità Femcel", in *Im@go. A Journal of the Social Imaginary*, vol. 23.

Ciccione, S. (2019), *Maschi in crisi? Oltre la frustrazione e il rancore*, Torino, Rosenberg & Sellier.

Dabral, A. (2024), "The Actions of a Lonely Woman and the Effects of Online Incel Communities on Society", in *Richard T. Schellhase Essay Prize in Ethics*. 28. [https://digitalcommons.ursinus.edu/ethics\\_essay/28](https://digitalcommons.ursinus.edu/ethics_essay/28)

De Gasperis, A. (2021), "Giacomino uno di noi. Letteratura italiana e pratiche di maschilità nel Forum dei Brutti", in *AG About Gender-Rivista internazionale di studi di genere*, 10(19), 71.

Dolce, R., Pilla, F. (2019), *Il web che odia le donne*. Milano, Ledizioni.

Dordoni, A., & Magaraggia, S. (2021), "Modelli di mascolinità nei gruppi online Incel e Red Pill: Narrazione vittimistica di sé, deumanizzazione e violenza contro le donne." In *AG About Gender-International Journal of Gender Studies*, 10(19).

Esposito, E. (2022), *Comunicazione artificiale: Come gli algoritmi producono intelligenza sociale*. Milano, EGEA.

European Commission (2021), *Incels: A first scan of the phenomenon (in the EU) and its relevance and challenges for P/CVE*. Bruxelles, European Commission.

Evans, H.R., Lankford, A., "Femcel Discussions of Sex, Frustration, Power, and Revenge". In *Arch Sex Behav* 53, 917-930 (2024). <https://doi.org/10.1007/s10508-023-02796-z>.

Farci, M. e Righetti, N. (2019), "Italian Men's Rights Activism and the Online Backlash Against Feminism", in *Rassegna Italiana di Sociologia*, vol. 4, pp. 765-781.

Gillespie, T. (2010), "The politics of 'platforms'", in *New media & society*, 12(3), 347-364.

Ging, D. (2019), "Alphas, Betas, and Incels: Theorizing the Masculinities of the Manosphere", in *Men and Masculinities*, 22 (4), 638-657.



Goffman, E. (1963), *Stigma: Notes on the Management of Spoiled Identity*, Simon & Schuster, trad.it., Bontempi, M. (2018). *Stigma: note sulla gestione dell'identità degradata*, Verona, Ombre Corte.

Goffman, E. (1959), *The Presentation of Self in Everyday Life*, Garden City (NY), Doubleday; trad. it. *La vita quotidiana come rappresentazione*, Il Mulino, Bologna, 1969.

Goffman, E. (1974), *Frame analysis: An essay on the organization of experience*. Harvard University Press.

Illouz, E. (2019), *The End of Love : A Sociology of Negative Relations*, Oxford University Press.

Kay, J. B. (2022), "Abject desires in the Age of Anger. Incels, femcels and the gender politics of unfuckability", in, *Imagining "We" in the Age of "I". Romance and Social Bonding in Contemporary Culture*, a cura di Harrod, M., Leonard, S.& Negra, D. Routledge, New York, p. 29.

Lilly, M. (2016), *The World is Not a Safe Place for Men: The Representational Politics of The Manosphere*, Doctoral dissertation, Université d'Ottawa/University of Ottawa.

Lughi, G. (2023), "Il ruolo dell'immaginario nella cultura digitale", in *Culture Digitali*, n.9.

Marwick, A. E., & Caplan, R. (2018), "Drinking male tears: Language, the manosphere, and networked harassment" in *Feminist media studies*, 18(4), 543-559.

Menzie L, (2022), "Stacys, Beckys, and Chads: The construction of femininity and hegemonic masculinity within incel rhetoric", in *Psychology & Sexuality* 13 (1): 69-85.

Meier, M. L., & Sharp, K. (2024), "Death to Chad and Stacy: Incels and anti-fandom as group identity", in *International Journal of Cultural Studies*, 27(3), 349-367.

Miller, H. (1995), *The presentation of self in electronic life: Goffman on the Internet*. In Embodied knowledge and virtual space conference, Vol. 9.

Morin, E. (2005), *Lo spirito del tempo*, Milano, Meltemi.

Nagle, A. (2017), *Uccidi tutti i normies: le guerre culturali online da Tumblr e 4chan all'alt-right e Trump*, Roma, Zero libri.

Papacharissi, Z. (2010), *A Networked Self*, London, Routledge.



Pentecoste, N. (2013), "Goffman rewind: La presentazione del self nel quotidiano digitale", in *Mediascapes journal*, (2), 106-118.

Pizzimenti, D. Pasciuto, F. (2022), "Un'analisi netnografica sul fenomeno dell'hate speech nella manosphere italiana", in *Metis. Ricerche di sociologia, psicologia e antropologia della comunicazione*, vol. XXIX, 1-2.

Poell, T., Nieborg, D.B., Duffy, B.E., (2022), *Piattaforme digitali e produzione culturale*, Roma, Minimum Fax.

Ragone, G. (2015), "Radici della sociologia dell'immaginario", In *Mediascapes journal*, 4, 63-75.

Rainie, L., Wellman B. (2012), *Networked: The New Social Operating System*, Cambridge, MIT Press.

Scarcelli, C. M. (2021), "Manosphere periferiche. Ragazzi, omosocialità e pratiche digitali", in *AG About Gender-International Journal of Gender Studies*, 10 (19).

Van Dijck, J., Poell, T., & De Waal, M. (2018), *The platform society: Public values in a connective world*. Oxford University Press.

Van Valkenburgh, S.P. (2021), "Digesting the Red Pill: Masculinity and Neoliberalism in the Manosphere", in *Men and Masculinities*, vol. 24, n. 2, pp. 84-103.

Vingelli, G. (2019), "Antifemminismo online. I Men's Rights Activists in Italia", in *Im@go. A Journal of the Social Imaginary*, vol. 14, pp. 219-247.

Vittadini, N. (2018), *Social media studies. I social media alla soglia della maturità: storia, teorie e temi*, Milano, Franco Angeli.







*Focus*







A JOURNAL OF THE  
SOCIAL IMAGINARY

# Intelligenza artificiale. E se avessimo sbagliato metafora? Una riflessione a partire da “Comunicazione artificiale. Come gli algoritmi producono intelligenza artificiale” di Elena Esposito (2022)



Olimpia Affuso

olimpia.affuso@unical.it

Dipartimento di Scienze Politiche e sociali | Univeristà della Calabria

## Abstract

*“Artificial intelligence”: Questioning the Appropriateness of the Metaphor. Considerations from “Comunicazione artificiale. Come gli algoritmi producono intelligenza artificiale”, by Elena Esposito (2022)*

Engaging with Elena Esposito's recent work, *Artificial Communication: How Algorithms Produce Artificial Intelligence* (2022), this article addresses the epistemological challenges posed by the concept of "artificial intelligence." Esposito argues that while the term serves as a powerful and evocative metaphor, the comparison between human intelligence and algorithms is fundamentally flawed. It is not that machines think like humans, but rather that humans have adapted their communication to interact with machines. Building on this idea, the essay offers a broader reflection on the role of metaphors in shaping social knowledge. Metaphors, such as the one underpinning "artificial intelligence," are not merely illustrative but are active constructs that generate dynamic ideas and shape our understanding of emerging phenomena. Yet, by relying on analogies and assimilation, they may also obscure our ability to grasp the complexity of constantly evolving realities.

## Keywords

Intelligenza Artificiale | Metafora | Algoritmi | Comunicazione Artificiale



## 1. Dentro la metafora

**N**ella storia del pensiero la metafora rappresenta quel costrutto che lavora alla produzione di immagini e accezioni semantiche che ci si attende siano capaci di produrre nuove idee.

Lungi dall'essere un mero artificio retorico, o qualcosa di imputabile unicamente a giochi linguistici di parole, la metafora non è diffusa solo nel linguaggio, ma anche nel pensiero (Johnson, Lackoff, 1980). Per cui, è possibile raccontare il mondo con espressioni linguistiche che sono metafore perché le metafore esistono nel nostro sistema concettuale. E i concetti che le guidano sono gli stessi che regolano le nostre attività quotidiane.

Che sia messa al servizio della scienza, e quindi orientata verso le spiegazioni, o che sia letteraria, dunque volta a produrre attraverso elementi finzionali un'idea e una esperienza immaginifica del possibile (Parini, 2023), o, ancora, sia usata in relazione alla nostra vita ordinaria, la metafora rende accessibile la conoscenza del mondo, spesso anche a livelli inediti di accuratezza. Attraverso di essa concettualizziamo i nostri comportamenti, e l'immagine che ne diamo nel descriverli – Mark Johnson e George Lackoff fanno in proposito, tra gli altri, l'esempio di chi dica: "questa discussione è una guerra" (1980) – influenza in maniera sistematica non solo come il comportamento viene percepito, ma esattamente il suo svolgimento.

Naturalmente, nel momento in cui si sistematizza un aspetto, se ne mette in ombra un altro.

In questo gioco di pensieri, al contempo gettati oltre la realtà ma che si possono agganciare a essa con le immagini, non di rado la metafora cela un processo di narrazione per analogia, quasi dando voce a una sorta di pigrizia del pensiero, anziché a un nuovo affresco. E le analogie più significative che essa trova sono nel corpo, come mostrano Johnson e Lackoff: il corpo e il suo radicamento sono la base della nostra conoscenza ed esperienza del mondo (ivi).

La metafora non può del resto non affidarsi a strutture di significato date. Non fosse altro che per la sua stessa funzione, ovvero rendere chiaro e familiare qualcosa di nuovo che ci si pone di fronte, o che necessita di comprensione. Come mostra Paul Ricoeur (2020) capiamo che una metafora vuole dire qualcosa di sensato del mondo perché mostra tratti di somiglianza con ciò di cui parla, e che vuole enfatizzare. Se diciamo "quest'uomo è un leone", vogliamo accentuare i tratti simili all'animale, per esempio il coraggio.

Al contempo, stiamo dicendo qualcosa di significativo perché risponde a una sorta di verità nota. Per Ricoeur in questo gioco si svela la doppia natura della metafora, linguistica e visuale, per cui essa funziona attraverso parole che sono immagini e immagini che diventano parole.

La metafora, quale parte dell'immaginario, è ciò che esiste ed è ciò che consente di andare oltre. E il fatto che lo consenta si gioca nella *similitudine* che stabilisce, ovvero nell'immagine con cui riferendosi alla realtà consente di vederla.



Nel rapporto con gli oggetti da osservare o narrare, questa geniale “trovata” della mente funge da ancella capace di muoversi su molteplici fronti: ideare, spiegare, addomesticare, configurare (o meglio: realizzare).

Fronti che interagiscono, a loro volta, con almeno tre approcci epistemologici alla metafora (cfr. Frezza, Gagliazzo, 2014). Il primo guarda all’uso delle metafore e le coglie come prodotti creativi che nascono nel momento costitutivo e costruttivo di una teoria. In questo caso le metafore risultano molto utili a cogliere l’ipotesi di fondo della teoria stessa.

Naturalmente va considerato che a questa fase segue poi quella del radicamento teorico, in cui le metafore si affermano come concetti influenti.

E che, una volta affermatesi, invitano a confronto anche altre teorie, che vi si misurano per similitudini o per differenze. E tutto ciò dà luogo a importanti dibattiti interdisciplinari. Il secondo approccio epistemologico è invece interessato a problematizzare il rapporto tra metafore e modelli teorici, e quindi a valutare se la metafora ne sia assorbita, abbia un mero ruolo ancillare rispetto alle teorie, oppure domini su di esse indebolendone la vocazione al dubbio e alla ricerca di nuovi paradigmi e immagini. Infine, il terzo approccio epistemologico mira a cogliere le risonanze ideologiche che le metafore raggruppano in sé.

E lo fa ancora una volta guardando al rapporto con le teorie e le discipline, in una indagine volta però a osservare non solo la vocazione denotativa della metafora, ma anche quella connotativa, ovvero l’insinuarsi di significati ideologici, latenti nei termini teorici scelti e impiegati (ivi).

Si può dire, riprendendo ancora Johnson e Lackoff (1980) che i valori fondamentali di una cultura sono coerenti, direi anche ben rappresentati dalla struttura metaforica di quella cultura.

Oltre che per la sua capacità di generare nuove articolazioni del pensiero, dunque, la metafora rileva, come sottolinea Ricoeur, per il fatto di produrre e veicolare immagini e idee attraverso “il fenomeno di risonanza, di eco o di riverbero”.

A essa, del resto, serve rimandare a qualcosa capace di essere riconosciuto: “schematizzando tramite l’attribuzione metaforica, l’immaginazione si diffonde in ogni direzione, rianimando delle esperienze anteriori, risvegliando dei ricordi che dormivano, irrigando i campi sensoriali adiacenti (Ricoeur, 2020: 211). Anche Steve Dowden, che le considera capaci di stimolare processi di conoscenza profondi, si sofferma a dire che proprio in quanto sono al contempo ammalianti, persistenti e familiari, le metafore possono avere una profonda presa sulla nostra immaginazione e condizionare il modo in cui vediamo il mondo (Dowden, 2007: 102; cit. in Parini, 2023: 38).

Come in sintonia con questa premessa, Elena Esposito, nel suo recentissimo *Comunicazione artificiale. Come gli algoritmi producono intelligenza artificiale* pone all’attenzione del pensiero scientifico e delle scienze sociali il problema epistemologico insito nell’impiego del concetto di “intelligenza artificiale”. Si tratta per Esposito esattamente di una metafora, ma una metafora fuorviante, perché distrae dal comprendere gli usi e gli impatti dell’attuale comunicazione digitale.



Un pensiero, anzi forse una avvertenza, quella di Esposito, particolarmente interessante rispetto alla riflessione odierna sull'IA, se si tiene conto che, anche in base a quanto detto fin qui, una metafora non è solo il condensato per leggere un fenomeno, per darne una interpretazione teorica, ma un costrutto che fornisce istruzioni per agire di fronte a esso.

## 2. Dall'intelligenza alla comunicazione artificiale. Ovvero: sull'uomo che ragiona come la macchina, e non viceversa.

Nel panorama odierno della comunicazione sono in corso una serie di sviluppi e trasformazioni rispetto alle quali siamo, secondo Esposito, sprovvisti di un concetto adeguato. In particolare, manca allo stato attuale un concetto per la comunicazione che sia adatto a rendere conto del fatto che essa non si gioca più solo tra uomini, o tra uomini a mezzo macchine. Oggi la comunicazione avviene infatti anche tra uomini e macchine. E il partner *con cui* (non solo *tramite cui*) l'uomo comunica è l'algoritmo.

Questo nuovo partner genera informazioni, auto apprende, controlla esso stesso le informazioni che genera. E lo fa in modo né arbitrario né casuale. Ecco perché ci appare intelligente come una mente umana; ed ecco perché, in quanto pertinente alla macchina, chiamiamo questa facoltà intelligenza artificiale.

Si mira, allora, in questo contributo, in stretto dialogo con il libro di Esposito, non tanto a parlare degli algoritmi, ma a capire, nei termini da lei posti, cosa si può cogliere di diverso nei processi contemporanei se, invece di guardare le cose attraverso il prisma offerto dall'idea che si tratti di intelligenza artificiale, le guardassimo contemplando l'idea, o meglio restando nel paradigma di una comunicazione artificiale. Significa, in altri termini, considerare cosa è possibile fare *con* la macchina, anziché guardare ciò che accade in capo alla macchina, e dunque rimettere al centro della riflessione il problema della macchina come infrastruttura del sociale. Ciò naturalmente, da un lato, richiede di tenere ben ferma l'idea che le innovazioni tecnologiche che stiamo sperimentando si inscrivono anch'esse in una storia di lungo periodo, che riguarda le macchine come mezzi, e che parte almeno da Johannes Gutenberg, passa da Alan Turing, arriva a Mark Zuckerberg e inevitabilmente procede ancora oltre, per esempio con dispositivi come Alexa, o con Sophia – la Audrey Hepburn umanoide della Hanson Robotics. Dall'altro lato, implica la necessità di non trascurare che, sebbene sia forse la prima volta che le macchine sono in grado di dare informazioni mai pensate da una mente umana, combinandole con enorme efficienza, non sono affatto in grado di produrle. Ciò che fanno le macchine, fuor di metafora, non può dirsi *in nulla simile* a quello che fa il cervello umano. Scrive Esposito più precisamente:

i processi che guidano gli algoritmi sono completamente differenti dai processi della mente umana, e in effetti nessuna mente umana o combinazione di menti potrebbe riprodurli e nemmeno comprenderli. Però l'intelligenza umana rimane irrinunciabile. Gli algoritmi che autoapprendono sono in grado di calcolare,



combinare ed elaborare differenze con un'efficienza strabiliante, ma non sono in grado di produrle autonomamente. Le trovano sul web. Tramite i big data, gli algoritmi si "nutrono" delle differenze generate (consapevolmente o inconsapevolmente) dalle persone e dal loro comportamento per produrre delle informazioni nuove, sorprendenti e potenzialmente istruttive. I processi degli algoritmi partono dall'intelligenza e dall'imprevedibilità (dalla contingenza) degli utenti per rielaborarle e operare in modo intelligente come partner comunicativi, senza alcun bisogno di essere intelligenti (2022: 5).

Dunque, le macchine e i loro algoritmi sono *partner comunicativi* delle menti umane, aiutanti che da esse sono guidati e da esse partono, e che possono fornire informazioni e compiere connessioni infinitamente complesse proprio perché non le devono pensare, perché sono svincolate dal bisogno di comprendere.

Ed è qui, semmai, che emerge un problema centrale di fronte agli algoritmi: la difficoltà di immaginare e gestire procedure così pervasive e autopoietiche, eppure così indipendenti dai significati come quelle digitali. Un problema che, peraltro, non si coglierebbe se si ragionasse unicamente all'interno del paradigma dell'intelligenza artificiale, che implicherebbe anche la convinzione che, trattandosi di qualcosa di intelligente, sia gestibile e sotto controllo, chiaro e intellegibile. Invece, la vera risorsa, nel caso di algoritmi, big data e tecnologie digitali in genere, è proprio la contingenza, l'incertezza, che solo la macchina può amministrare attraverso le sue procedure di calcolo.

Con il controeffetto che se l'incertezza è necessaria al successo dell'algoritmo, la stessa prevedibilità finisce per diventare un ostacolo per il sistema.

Dunque, il libro di Esposito consente di mettere a fuoco un aspetto essenziale nel discutere intorno alle nuove tecnologie, che si può cogliere nella distanza tra gli sviluppi tecnologici e l'immaginario metaforico a cui si attinge per trovare concetti adeguati a interpretare tali sviluppi.

E uno di questi sviluppi riguarda il fatto che il panorama mediatico e le forme della comunicazione siano cambiate, e ciò è avvenuto attraverso meccanismi che si stenta a cogliere, perché spingono l'uno in direzione contraria dell'altra. Per cui, il panorama dei media risulta ormai caratterizzato da un radicale fenomeno di personalizzazione, mentre l'infrastruttura comunicativa, ossia la forma della comunicazione, non può affatto dirsi dello stesso tipo.

Se prima, per ricevere informazioni diverse da quelle tramesse in modo standardizzato a tutti, un utente doveva trovare il modo di esprimere la sua opinione e farla arrivare all'emittente, oggi l'algoritmo gli invia direttamente tutto quello che lui si aspetta. L'algoritmo sa, quasi prima che noi li esprimiamo, quali siano i nostri desideri, gusti, pensieri, sentimenti. Ne connette le parti costitutive e ce li mostra nella loro pienezza, proprio come noi li esprimeremmo. E tale meccanismo implica un fatto ben diverso da quanto la maggior parte degli analisti ha ritenuto finora, che ciascuno di noi fosse più attivo nei mondi digitali che di fronte ai media mainstream.

La comunicazione digitale, paradossalmente, è meno personalizzata di quanto appaia, nel senso che non è la persona a selezionare e a scegliere, ma l'algoritmo. La comunicazione anonima dei mass media, come scrive Esposito:



offre ampi spazi di configurazione personale. Il singolo lettore di un libro può decidere da solo il ritmo, la velocità e l'ordine di lettura; può rallentare o accelerare, tornare indietro, iniziare dalle fine o saltare dei passaggi, e confrontare il testo con altri testi [...]. Così facendo, ogni lettore produce una comunicazione specifica, che corrisponde alle sue caratteristiche, ai suoi interessi e alle sue conoscenze, ed è diversa da quella di qualsiasi altro lettore (2022: 90).

Gli algoritmi invece, sfruttano:

il comportamento e l'imprevedibilità dell'utente per apprendere e per intervenire in modi complessi e appropriati. [...] usano i big data per nutrirsi dell'intelligenza degli utenti [...] e anche per rivolgere delle comunicazioni individuali a ciascuno di noi. Nella comunicazione digitale la contingenza virtuale produce un intreccio senza precedenti tra le attività degli utenti e i riferimenti generalizzati (ivi).

Dunque, offrono un contenuto individualizzato, che non è frutto di una attiva manifestazione del pensiero, ma di un calcolo operato da una macchina capace di elaborare informazioni infinitesime, e impercettibili a noi stessi.

Nella comunicazione standardizzata di massa, a una offerta uguale per tutti risponde una selezione personale dell'utente. Che grazie alle possibilità offerte dal linguaggio, e a partire dall'accento che intende soggettivamente dare in relazione al suo contesto culturale, la può ri-accentuare e reinterpretare infinite volte (Vološinov, 1986), nelle sue pratiche fruttive e artistiche, agendo sui testi e sui significati attraverso le sue competenze al suo artigianato. Nella comunicazione digitale, invece, un'offerta altrettanto uguale per tutti viene spaccettata tramite un processo informatico e restituita al singolo in pezzi "unici" offerti al suo gradimento.

Il pubblico, in questo caso, invischiato nei meccanismi del mezzo, come ha ben spiegato Henry Jenkins (2006), e prima di lui J. David Bolter e Richard Grusin (1999), *converge* nelle diverse piattaforme, e solo tramite di esse, all'interno del campo tecnologico, può esprimere la sua volontà di appropriazione e diffusione dell'oggetto culturale. Al processo di ri-accentuazione culturale subentra un processo di *rimediazione* tecnologica. Ed è così pervasivo questo processo, tutto interno al campo digitale, che la questione del diritto all'oblio delle tracce che lasciamo nella rete è "la" questione con cui si deve misurare chiunque vi navighi e vi agisca. Una questione che certo non riguarda un suonatore di jazz o il tanghero di una milonga che si cimentino con l'improvvisazione creativa e la possibilità di riscrittura di un brano musicale o un passo di danza (cfr. Sparti, 2015).

Fin qui, dunque, risulta chiaro che con Esposito l'idea in base alla quale la differenza tra le due forme di comunicazione si gioca nel potere di azione e personalizzazione del *prosumer* va ridimensionata in virtù dell'altro elemento, relativo al potere dell'algoritmo di fornire il materiale grezzo su cui poi egli opera.

L'algoritmo organizza l'agenda dei temi su cui vale la pena intervenire. Organizza la moda del momento, fornisce le idee, definisce le priorità, innesca le emozioni e le compendia in forme: sorrisi, occhioni sgranati o a forma di cuore, lacrimuccia,



stupore al limite di una espressione da panico alla Munch, e via dicendo. E se nel campo dell'informazione *mainstream* i valori notizia che orientano la selezione dei temi rilevanti sono a carico del sistema informativo mediale e rispondono alle routine e alle istanze della produzione, quando si passa online sono a carico dell'algoritmo (altro attore mediale, chiaramente, ma di tipo informatico) che li smista in funzione delle routine fruibili, dei gusti e delle predisposizioni dell'utente.

Come scrive Davide Bennato (2024), gli algoritmi sono "delle procedure - dei passi - che un software deve compiere per svolgere un certo compito". Il paragone di Bennato con la cucina aiuta in maniera particolarmente efficace a comprendere il quadro:

Poniamo il caso che dobbiamo preparare una torta: abbiamo un input (gli ingredienti di base), abbiamo un output (il risultato finale), abbiamo un hardware (un forno, degli utensili da cucina, come cucchiari, pentole, scodelle, e così via dicendo). L'algoritmo consiste nella ricetta: una sequenza di passi da rispettare [...]. Ovviamente possono esserci piccole varianti nella ricetta a seconda del libro nel quale la leggiamo: nella nostra analogia, il libro di ricette è il programma [ovvero l'interpretazione della ricetta] (ivi: 56).



### 3. E se fosse una lista? Antropologia di un ritorno all'antico

399

Le questioni poste fin qui, in maniera pur parziale, sono utili a rendere evidente una serie di punti essenziali per una riflessione sui processi in atto e sui modi euristici di coglierli e spiegarli. Intanto, che nel parlare di intelligenza artificiale si parla per lo più di una sequenza a partire da potenzialità e informazioni (dataset) che senza certi passi, procedure, o "ricette", resterebbero inesprese; poi, che a "libri" diversi corrispondono ricette diverse; infine, che ogni "ricetta", nella sua sequenza, ha conseguenze differenti sul *sapore* finale, ovvero sulla nostra rappresentazione del mondo.

Pur nella sua specificità, ciò che definiamo intelligenza artificiale è allora un sistema mediale di reperimento e organizzazione di dati per la realizzazione di un output che non è altro se non il risultato della costruzione di chi li ha raccolti. Il che significa che anche tale output è legato, più di quanto si pensi, al processo di immagazzinamento e delle informazioni che poi gestisce (Beaulieu, Leonelli, 2022; Affuso, Coco, 2019).

Particolarmente interessante, in relazione a tutto ciò, è il secondo capitolo del libro di Esposito, dove mostra come l'intelligenza artificiale può meglio essere chiarita pensando a una lista. Il web lavora come una lista. Come una ricetta, il web raccoglie ingredienti. E, dunque, chiaramente, non pensa. Una lista può essere stata pensata prima, durante la fase di raccolta. O può dare origine a una idea. Ma, mentre esegue, non pensa. La lista raccoglie, per ricordare, classifica, organizza.

Non è un caso che, come nota Teresa Numerico, l'intelligenza artificiale si sia sviluppata oggi. Per quanto sia stata pensata almeno 70 anni fa, si è affermata solo con il web 2.0, dopo che una grande quantità di dati hanno consentito di fare



inferenze e previsioni sulla base delle informazioni disponibili (Numerico, 2024). Da questo momento il futuro è divenuto visibile, anche grazie alla ricorrenza dei dati, e dunque altamente prevedibile senza bisogno di fare ipotesi su di esso. E, fatto ancora più rilevante, il futuro è diventato “visibile” sulla base di dati talmente contingenti, fluidi, come le opinioni soggettive, che ormai è possibile anche indirizzarlo. La grande mole di dati malleabili è divenuta la base non solo per anticipare il futuro, ma anche per guidarlo. Come un nuovo oracolo di Delfi (Numerico, 2021).

Nonostante questa grande potenzialità, tuttavia, la lista non è che il modo più antico ed elementare di gestire il mondo. Come scrive Umberto Eco, ne *La vertigine della lista* (2009, cit. in Esposito, 2022: 39 e ss.), inevitabilmente smette di essere necessaria con la nascita della scrittura, e in particolare quando la conoscenza occidentale comincia a padroneggiare i concetti astratti (Havelock, 1976, cit. ivi: 47).

Con la digitalizzazione si sviluppa dunque un ritorno a una organizzazione di informazioni e notizie, a una selezione che ha un meccanismo arcaico, che non richiede ragionamenti ma raccolta, elenchi e calcoli. Anche procedure come la lettura delle immagini o l'interpretazione dei testi, o la visualizzazione digitale, su cui Esposito si sofferma a lungo nel libro, non sono che elaborazioni di dati.

Lo stesso intero sistema di social media mostra come il web e le sue piattaforme non funzionino in quanto intelligenza ma come comunicazione: reti di interazioni, contatti, link, like che rappresentano comportamenti comunicativi ed espressivi, non interpretazioni. Il che è tanto più evidente se si riflette sui *bias* e sui pregiudizi che vengono riproposti dagli algoritmi. Si discute molto oggi intorno al fatto che i progettisti siano bianchi e maschi e che abbiano immesso i loro stereotipi nelle istruzioni. Ma a questo problema si aggiunge che se la macchina autoapprende, e ha appreso a lavorare in maniera non neutrale, continuerà a farlo: ai *bias* dei progettisti si sommeranno i *bias* dell'autoapprendimento dei dati. Confermando ulteriormente che il problema non è cognitivo ma comunicativo (Esposito, 2022: 175).

Sembra dunque banale dirlo e ribadirlo, ma i computer non sono intelligenti, anche nel senso etimologico del termine: pure quando ci restituiscono informazioni tramite immagini, non riflettono, non comprendono sensi, non capiscono. Contano. Elencano. Non hanno dunque imparato a pensare come noi. Piuttosto, siamo noi ad aver imparato a comunicare con loro, e, forse più inquietante, a pensare come loro.

Tuttavia, e non potrebbe essere diversamente, questo non significa che le procedure siano indipendenti dalla “creatività” umana. In virtù del fatto che il sistema elabora solo dati disponibili, e anche quando i soggetti non conoscono il funzionamento della macchina, il dataset dipende da attività umane, frutto di una progettazione, di idee, di scelte fatte dall'uomo. Il sistema include qualcosa che “potrebbe essere diverso se fosse stato impostato in altro modo” (Esposito, 2022: 157). Ad esempio, tornando al caso dei *bias*, il problema veramente sostanziale non è costituito dai valori e pregiudizi che ha il progettista che immette i dati, ma dai valori che *non ha*.

Questo ci porta, in conclusione, inevitabilmente a sottolineare quanto tali scelte e valori interagiscano con un sistema di rappresentazioni, vero e proprio immaginario



sociale, che è presupposto e orientamento del nostro agire e del nostro rapporto con il mondo. Un immaginario che è fonte di riproduzione ma anche fonte creatrice dell'agire e del pensare dell'uomo.

Per tornare alla metafora dell'intelligenza artificiale, da cui siamo partiti, essa è chiaramente l'elemento immaginale, il condensato espressivo di questo cassetto semantico di parole/attrezzi capaci di farci vedere la realtà nei termini del cassetto stesso. Direi anche di giustificarla: sarebbe ideologico, oltre che fuorviante, infatti, non cogliere il nesso inestricabile tra la parola e la sua sorgente non naturale e non neutrale. Una sorgente che ha in sé tutte le matrici epistemologiche di ogni nuova dimensione realizzabile, nonché tutte le configurazioni necessarie per rispondere all'ambiente e confermarne lo status quo. Che è inscritta essa stessa in un universo di infiniti altri significati.

Mettere questo in evidenza significa non dare per scontato che quella indicata dalla targhetta sia l'unica interpretazione possibile. Significa svelare il principio costruttivo che la orienta e la genera. Significa riportare gli atti collettivi, con le loro strategie per rapprendere il mondo, al centro di una riflessione che non perda di vista il momento fondativo e l'origine del "totem" e, contemporaneamente, il meccanismo di oblio che alla fine investe il processo ideativo. Qualcosa che Émile Durkheim aveva magnificamente messo in evidenza ne *Le forme elementari della vita religiosa* (1912) parlando della religione.

Se la metafora è un modo tanto letterario che scientifico, nonché quotidiano, di definire l'esistente, e di innescare una conoscenza foriera di una nuova apertura mentale, la sfida di Esposito a cercare altre metafore, o magari anche a vederne l'aspetto totemico e a pensare al di là di esse, è quanto mai significativa.



## Bibliography

- Affuso O., Coco A. (2019), "Big Data: a challenge for social sciences. Criticisms for social Knowledge and politics", *Central European Political Science Review*, vol. 20.
- Beaulieu A., Leonelli S. (2022), *Data and Society. A Critical Introduction*, Sage, Londra.
- Bennato D. (2024), *La società del XXI secolo. Persone dati, tecnologie*, Laterza, Roma Bari.
- Bolter J. D., Grusin R. (1999), *Remediation: Understanding New Media*, Tr. It. *Remediation. Competizione e integrazione tra media vecchi e nuovi*, Guerini, Milano, 2005.
- Durkheim É. (1912), *Les formes élémentaires de la vie religieuse*, Tr. It. *Le forme elementari della vita religiosa*, Mimesis, Milano, 2013.
- Eco U. (2009), *La vertigine della lista*, Bompiani, Milano.
- Frezza G., Gagliazzo E. (2014), "Fare metafore e fare scienza", *Aisthesis*, 2/2014, pp. 25-42.
- Havelock E. A. (1976), *Origins of Western Literacy*, Ontario Institute for Studies in Education, Ontario.
- Jenkins H. (2006), *Convergence Culture: Where Old and New Media Collide*, Tr. It. *Cultura convergente*, Apogeo, Roma, 2014.
- Lackoff G., Johnson M. (1980), *Metaphors We Live By*. Tr. It. *La metafora e la vita quotidiana*, Roi Edizioni, Macerata-Milano, 2022.
- Numerico T. (2024), *Dati (desiderio dei)*, in F. Cimatti, A. Maiello, *Quasi viventi. Il mondo digitale dalla A alla Z*, Codice edizioni, Torino.
- Ead. (2021), *Big Data e algoritmi*, Carocci, Roma.
- Parini E.G. (2023), *Un punto oscillante. Arte, conoscenza, impegno*, in Affuso O., Amendola A., Parini E.G., *L'odore della vita. Pier Paolo Pasolini: l'opera, la conoscenza, l'impegno pubblico*, Roma, Rogas.
- Ricoeur P. (1975), *La Métaphore vive*. Tr. It. *La metafora viva*, Milano, Feltrinelli, 2020.
- Sparti D. (2015), *Sul tango. L'improvvisazione intima*, il Mulino, Bologna.



Olimpia Affuso  
*Intelligenza artificiale. E se avessimo sbagliato metafora?*

Vološinov V. N. (1986), *Marxism and the Philosophy of Language*, Harvard University Press, Cambridge.

Ostermann A. C., Dowdy J. D., Lindemann S., Turp J. C., Swales J. M. (1999), "Patterns in self-reported illness experiences: letters to a TMJ support group", *Language and Communication*, 19: 127-47.

Sage C. (2003), "Social embeddedness and relations of regard: alternative 'good food' networks in South-west Ireland", *Journal of Rural Studies*, 19: 47-60.

Durbridge N.H., Stratfold M. (1997), "Varying the texture: a study of art and learning", *Journal of Interactive Media in Education*, Retrieved on 10 October 1997 from <http://www.jime.open.ac.uk/jime/current.html>.

Shuy R. W. (1998), *Bureaucratic Language in Government and Business*, Washington DC, Georgetown University Press.

Barber R.B. (2007), *Consumed, How Markets Corrupt Children, Infantilize Adults, and Swallow Citizens Whole*. Tr. It. *Consumati. Da cittadini a clienti*, Torino, Einaudi, 2010.

Srivastava A. K., Gupta R. S. (1990), "Literacy in a multilingual context", in Pattanayak D. P. (ed.), *Multilingualism in India*, Clevedon, Multilingual Matters, 67-68.







A JOURNAL OF THE  
SOCIAL IMAGINARY

# Nella spirale della conoscenza: arte e scienza tra creatività e metodo

Daniele Garritano

daniele.garritano@unical.it

Dipartimento di scienze politiche e sociali | Università della Calabria



## Abstract

*Into the Spiral of Knowledge: Art and Science between Creativity and Method*

This text discusses *La creatività come metodo sociologico* (Marzo, 2022): a comprehensive essay that provides gnoseological, epistemological, and methodological analysis dedicated to the role of creativity in scientific knowledge processes, particularly within sociological research. It explores a wide array of constellations of cultural elements related to the history of art, technology, science, and sociological thought. The interpretation of Marzo's work aims to emphasize the anthropological connection between creativity and method in scientific and artistic knowledge processes, in line with the author's primary theoretical challenge. Additionally, it seeks to identify and deconstruct the dichotomy between these dimensions in various stages of sociological discourse through the examination of social imaginaries. In fact, by referencing the geometric figure of the spiral as an allegory of knowledge processes, Marzo illustrates that all artifacts, inventions, scientific discoveries, aesthetic works and behavioral practices can be viewed as distinct objectifications of the imaginative activity.

## Keywords

Art | Science | Social Imaginary | Knowledge | Homination





**L**a *creatività come metodo sociologico* di Pier Luca Marzo si presenta come un generoso saggio di analisi gnoseologica, epistemologica e metodologica, dedicato al ruolo della creatività nei processi di costruzione della conoscenza scientifica, in particolare nel campo della ricerca sociologica. Il libro è suddiviso in tre parti (1. *Le spirali della creatività*; 2. *L'immaginario sociale tra arte, scienza e tecnica*; 3. *La creatività come Rinascimento sociologico*), ciascuna delle quali approfondisce una configurazione dell'idea propulsiva posta al cuore della ricerca, ossia che «la via estetica apre alla creatività sociologica la possibilità di fare del metodo un'arte della ricerca orientata a cogliere la dimensione profonda della vita quotidiana». Ogni sezione mette a fuoco ampie costellazioni di elementi culturali che rinviano alla storia dell'arte e della tecnica, delle scienze e del pensiero sociologico, evidenziando la profonda implicazione antropologica tra creatività e metodo e smontando le ragioni della dicotomia che ha contrapposto tali dimensioni nel dibattito sociologico: «l'ideazione creativa deve seguire sempre un determinato metodo (costruttivo, scientifico, artistico) per potersi sostanziare in un prodotto originale; il metodo, a sua volta, è in virtù di un'immagine creativa che può interrompere la sua processualità circolare, controllata, sequenziale e predefinita per tracciare un nuovo percorso conoscitivo» (Marzo, 2022: 19).

Il ragionamento intorno alla creatività e alle sue modalità espressive, alle affinità/divergenze tra creatività scientifica ed artistica, al buon uso della creatività come metodo sociologico, tende ad articolarsi su un doppio livello. Da un lato, c'è una mappatura storica di ampio respiro sui processi di costruzione delle conoscenze umane; dall'altro, la desedimentazione di forme culturali e rapporti storicamente situati, naturalizzati nel senso comune della ricerca scientifica. Nel percorrere tali direzioni, l'autore si sofferma dettagliatamente sulla dimensione dell'immaginario, inteso come «l'ambiente socialmente e storicamente determinato (composto da simboli, sentimenti collettivi, credenze, conoscenze, memorie condivise, schemi dell'agire) che istituisce la realtà sociale» (ivi: 21). Tale ambito di elaborazione simbolica per la vita umana costituisce, infatti, il punto di equilibrio dell'intera architettura testuale.

Riferendosi all'orientamento teorico del *mundus imaginalis*, le cui caratteristiche sono esposte nella parte centrale del saggio, Marzo imbrocca un sentiero storico-antropologico in cui artefatti, invenzioni, scoperte scientifiche, opere estetiche e pratiche comportamentali sono trattate come oggettivazioni dell'attività immaginativa. Quest'ultima è posta alla base della vita sociale del genere umano, al fine di analizzare «le affinità creative che intercorrono tra arte e scienza» e di riconoscere l'importanza di tale connessione tanto per la vita mentale, quanto per i processi psico-sociali nella vita relazionale e, in particolare, nell'ambito del sapere scientifico.

Nell'apertura del volume, la prefazione di Ercole Giap Parini mette in evidenza la connessione fondamentale tra l'attività immaginativa e i processi di costruzione della

realtà, a partire da riferimenti noti alla letteratura sociologica (Max Weber per *Il lavoro intellettuale come professione*, Robert Nisbet per *Sociologia e arte*), incrociati con prospettive di ricerca interdisciplinare da cui lo sguardo sociologico trae elementi preziosi per riflettere sui processi creativi, tanto nella teoria quanto nella prassi della ricerca. Ne è un esempio Silvia Ferrara, filologa classica e specialista delle civiltà egee, autrice de *Il salto. Segni, figure, parole: viaggio all'origine dell'immaginazione* (Ferrara, 2021: 13), che contribuisce a illustrare «i grandi cambiamenti, le innovazioni radicali, i punti di svolta, irregolari e irriproducibili», connessi al carattere cooperativo dell'immaginazione, rintracciandoli nella *longue durée* della presenza umana sul pianeta (50000 anni: dal paleolitico ai nostri giorni), in termini di orientamento relazionale alla comunità, alla comunicazione, al «sentire insieme» e al condividere. «Si può ben capire – conclude Parini – [...] come riflettere sui processi creativi costituisca non tanto e non soltanto un progetto orientato a disvelare un aspetto *del* sociale, ma sia una prospettiva privilegiata dalla quale dare senso *al* sociale» (Marzo, 2022: 12).

L'articolazione del tema della ricerca consente di valorizzare il percorso di progressiva definizione della questione della creatività in termini di metodo sociologico. La sociologia della scienza, infatti, ha dietro di sé una lunga tradizione di domande con cui si è interrogata sull'apparente dicotomia tra creatività e metodo, invitando le comunità scientifiche a intraprendere un percorso autoriflessivo sui modelli di costruzione del sapere, sugli schemi interpretativi e operativi, sulle tecniche di ricerca e – soprattutto – sulle domande fondamentali che definiscono i rapporti tra oggetto, soggetto e campo della conoscenza: «cos'è la creatività e quali sono i piani in cui si esprime? Cosa accomuna e differenzia la creatività scientifica da quella artistica? Ci sono delle indicazioni da cui partire per usare la creatività come metodo sociologico?» (ivi: 20). In questo senso, *La creatività come metodo sociologico* rappresenta un invito a rileggere una certa tradizione sociologica, in cui la messa in questione delle modalità di costruzione della conoscenza scientifica è stata posta come un'esperienza ineludibile. Dall'altro lato, tale invito si rivela molto utile per riconoscere e mappare i tentativi più recenti di smontare la dicotomia neopositivista tra creatività e metodo, per rendere visibili le faglie epistemologiche in cui l'implicazione tra i due momenti è stata messa a tema e valorizzata<sup>1</sup>.

<sup>1</sup> In un contesto di rilettura critica del pensiero metodologico sono fondamentali i riferimenti a Weber per la conferenza *Scienza come professione* (1919), a Florian Znaniecki per *The Method of Sociology* (1934), a Pitrim Sorokin per *Mode e utopie nella sociologia moderna e scienze collegate* (1956), a Charles Wright Mills per *L'immaginazione sociologica* (1959), al già citato Nisbet per *Sociologia e arte* (1969), ad Alvin Gouldner per *La crisi della sociologia* (1970), a Edgar Morin per le numerose ricerche che, dalla fine degli anni Settanta, hanno descritto il metodo come un'avventura cruciale per la vita della mente, per la natura e per l'identità umana, ad Alberto Marradi (*Metodo come arte* del 1996; *Metodologia delle scienze sociali* del 2007) e, più recentemente, al lavoro di Alberta Giorgi, Micol Pizzolati ed Elena Vacchelli sui *Metodi creativi per la ricerca sociale* (2021), che fa il punto sui dibattiti in corso per stimolare un dialogo tra gli approcci creativi eterodossi o non canonizzati dalle istituzioni accademiche, sottolineando allo stesso tempo questioni di carattere etico e pratico intorno alla partecipazione nella ricerca con metodi creativi. Al di là di ogni partizione disciplinare, è essenziale per questo disegno anche un confronto con gli studi dedicati alle dinamiche psicologiche della creatività, tra i quali occorre segnalare almeno Lev Vygotskij per *Immaginazione e creatività nell'età infantile* (1930), Erich Neumann per *L'uomo creativo e la trasformazione*



## 1. Spirali e conchiglie: per comprendere il metodo

La capacità formativa dei processi creativi consente all'essere umano di attivare un metabolismo su larga scala: una capacità di trasformare che coinvolge tanto i processi di adattamento collettivi quanto la vita mentale degli individui, oltre a incidere sulle capacità di oggettivazione, sulla facoltà di fabbricare artefatti tramite l'integrazione di *poiesis* e *téchne*, sulla possibilità di configurare prodotti intellettuali, opere estetiche e pratiche comportamentali in relazione alle condizioni di esistenza cui è sottoposta la vita. Nell'esplorare i processi in cui creatività e metodo sono strettamente implicati, Marzo evoca la figura geometrica della spirale, che compendia i tratti essenziali di un movimento circolare dall'interno all'esterno e, al tempo stesso, chiama in causa la «condizione di apertura e di eccentricità [che] getta l'uomo in una condizione di naturale insicurezza». Come hanno scritto Jean Chevalier e Alain Cheerbrant, autori del *Dizionario dei simboli* (1963: 420), la spirale è «una linea che si avvolge su sé stessa, a imitazione forse delle numerose spirali che si incontrano in natura, ad esempio nelle conchiglie».

Si rivela opportuno tenere conto di quanto scrisse il poeta francese Paul Valéry, prendendo le mosse proprio dall'osservazione della forma spiroidale di una conchiglia, in un saggio del 1937 intorno all'enigma gnoseologico che riguarda la distanza tra l'umano e l'animale (*L'uomo e la conchiglia*). Tutti gli elementi che ritroviamo in natura, secondo Valéry, «ci propongono, stranamente unite, le idee d'ordine e di fantasia, d'invenzione e necessità, di legge e di eccezione; e nella loro forma troviamo, da una parte, la parvenza di un'intenzione e di un'azione che le avrebbe plasmate al modo in cui sanno farlo gli umani, dall'altra l'evidenza dei processi a noi vietati e impensabili. [...] Concepiamo la *costruzione* di questi oggetti, e sotto questo aspetto ci interessano e attraggono la nostra attenzione; non concepiamo la loro *formazione*, ed è sotto questo aspetto che ci incuriosiscono» (Valéry, 1937: 71). Vi è infatti una differenza non colmabile tra il processo attraverso cui l'animale *forma* la conchiglia e il modo umano di comprendere tale processo, riconducendolo a immagini mentali astratte per poterne disporre in un senso costruttivo.

Su questo punto, in aggiunta ai riferimenti presenti ne *La creatività come metodo sociologico* – tra i quali spiccano Arnold Gehlen per l'idea di essere umano inteso come *animale deficiente*, cioè mancante di sicurezza ontologica, Friedrich Nietzsche per la definizione di *animale non ancora stabilizzato* e Johann Herder per la nozione di *essere manchevole* – la posizione di Valéry può mostrarsi utile per mettere a fuoco il fondo antropologico del discorso sul rapporto tra creatività e metodo, che costituisce la base dell'intero edificio teoretico proposto da Pier Luca Marzo. La differenza tra vita umana e animale, dunque tra *bios* e *zoè*, è una differenza negativa: riconducibile

---

(1955) e Arthur Koestler per *L'atto della creazione* (1964). Se l'attività mentale è riconosciuta come un campo di esercizio decisivo per la capacità umana di usare le risorse creative, la sua pratica va intesa come un continuo processo di formazione che produce, distingue, mescola, combina e trasforma le immagini mentali, per orientare l'agire senso-motorio e le potenzialità speculative della specie (come evidenziano le ricerche in campo neurologico di Antonio Damasio).



alla non specializzazione dell'essere umano, costretto a creare nella cultura la propria «seconda natura», cioè «il *prodotto* di un essere unico al mondo, lui stesso "innaturale", costruito cioè in contrapposizione all'animale» (Gehlen, 1940: 75).

Tale differenza si ritrova già in Aristotele, che nell'*Etica Nicomachea* definisce l'arte come «una disposizione poetica accompagnata da ragionamento», distinta dal divenire delle cose naturali, «che trovano in sé il loro principio formativo» (Marzo, 2022: 29). È proprio il legame tra disposizione poetica e ragionamento, ossia la qualità creativa che si trova alla base dei processi di invenzione, a motivare la concezione del *méthodos* come «strada percorsa dalla creatività per oggettivare una nuova idea in un prodotto costruito a regola d'arte» (ivi: 31). Si arriva così a concepire l'utilità del metodo in un senso molto più esteso rispetto alla riduzione semantica che ne fa un insieme di tecniche di ricerca. Imboccare questa strada significa, infatti, comprendere il metodo come un modello poetico, ponendo al centro la capacità umana di riflettere e di trasformare gli schemi interpretativi e operativi dell'azione. Per dirla con Marradi, si tratta di praticare il metodo come arte, di concepire il *méthodos* – etimologicamente – come la strada che va al di là di quanto possiamo dare per scontato:

La foresta da attraversare è percorsa per tratti più o meno lunghi, e in varie direzioni, da sentieri già tracciati, più o meno battuti: solo le tecniche che gli altri ricercatori hanno già ideato, modificato, sviluppato. Naturalmente è molto più comodo percorrere sentieri già battuti; ma non si sa se porteranno alla radura desiderata o da qualche altra parte. Compito del ricercatore-metodologo è scegliere via via il percorso, tenendo conto della natura dei sentieri esistenti, del tempo a disposizione, delle risorse (Marradi, 1996: 80).

Perciò, se «ogni arte ha un proprio *méthodos*: l'architetto segue un metodo costruttivo per trasformare l'idea progettuale in un edificio; l'artigiano si avvale di un metodo manuale e di strumenti per dare forma al vaso; lo scienziato persegue un metodo di ricerca per produrre oggetti di conoscenza» (Marzo, 2022: 31), possiamo affermare – ritornando alla risonanza con il saggio di Valéry del 1937 – che «l'*idea di Fare* è la prima e la più umana. Spiegare non è mai che descrivere un modo di *Fare*: non è che rifare attraverso il pensiero». Se accettiamo i termini generali del ragionamento appena accennato, dobbiamo ammettere non solo che ogni arte ha un proprio metodo, ma che una forma di creatività è insita nell'avventura del metodo (nel suo momento di ideazione e negli interstizi operativi del suo uso), tanto quanto un pensiero metodologico è implicato in ogni processo creativo. Il disegno della spirale illustra perfettamente l'andamento di tale circolarità aperta: capace di espandere il proprio raggio partendo dal lampo dell'intuizione creativa (l'immagine mentale che dà principio alla fase poetica), per produrre una progressiva temporalizzazione tramite il metodo che, in quanto principio organizzativo orientato alla forma, non fa che formalizzare il processo creativo in uno schema operativo o – come scrive Marzo rifacendosi alle riflessioni simmeliane – in «una forma mentale svuotata del suo contenuto vitale» (Marzo, 2022: 51).



Sostenere che il rapporto tra creatività e metodo definisca la modalità umana di essere nel mondo ci conduce a chiederci come le visioni scientifiche del reale possano riempire di contenuti la sfera della conoscenza, dare forma a un certo stile interpretativo per ricondurre l'ignoto al noto, durare per un certo tempo (più o meno lungo), prima di essere messe in discussione da nuove visioni emergenti<sup>2</sup>.

A tale processo di riflessione metodologica si presta, ovviamente, anche la sociologia, che diventa sapere scientifico nel momento in cui, «grazie alla spirale rivoluzionaria del capitalismo industriale, il *mundus rationalis* ideato nel XVII secolo uscì dalle ristrette cerchie dei saperi specializzati per manifestarsi in una serie di fenomeni di massa» (ivi: 74). Dopo aver ripercorso alcuni tra i momenti cruciali che segnano la fondazione della scienza sociologica nella civiltà industriale europea del XIX secolo, alla luce del discorso epistemologico sulle categorie di creatività e metodo, il testo di Marzo propone una stimolante definizione della sociologia, nei termini di una nuova delimitazione dei «confini gnoseologici di una nuova provincia di sapere nel territorio delle scienze dell'uomo inteso come animale sociale». Ma non si limita a questo perché, oltre alla fondazione, delinea anche un principio generativo del mutamento nel divenire dei processi conoscitivi, affermando che «la sociologia, analogamente alle altre scienze, è una conoscenza "contraddittoriale" tra ideazione e metodo, società e scienza, credenza e razionalità, tra osservazione e interpretazione». Tale esercizio si rivela utile a comprendere il carattere storicamente situato delle scienze e, in particolare, il tipo di costruzione poetica che guida l'esperienza del pensiero sociologico: «un metodo costruito ad arte per disvelare qualcosa di nuovo dalla realtà sociale» (ivi: 95-96).



## 2. Immaginari sociali e processi trasformativi

Come già accennato, un campo di fondamentale importanza per apprezzare lo spessore della proposta teorica de *La creatività come metodo sociologico* è la dimensione dell'immaginario, inteso come ambiente socio-antropologico. Risuona ancora, alla luce delle precedenti considerazioni, la questione della differenza tra

<sup>2</sup> Tramite riferimenti puntuali alle categorie di «senso comune» e «provincie di significato» in Alfred Schütz (in particolare per l'articolo del 1953 intitolato *Common-Sense and Scientific Interpretation of Human Action*), alla *Sociologia della conoscenza* di Karl Mannheim (1952) e alle idee del saggio *Knowledge and Social Imaginary* di David Bloor (1976-1991), Pier Luca Marzo sottolinea la profonda connessione tra i processi conoscitivi e l'universo simbolico inteso come sentire comune: uno «sfondo pre-teoretico a partire dal quale la scienza può tracciare i confini della sua provincia di senso, e formulare un sapere teoretico in contrasto con il senso comune e l'informazione» (Marzo, 2022: 56). Si inserisce in questa rivisitazione archeologica del metodo scientifico una rassegna dei principali momenti di discontinuità metodologica nella storia del pensiero scientifico: dalla cosmologia aristotelica, ripresa nell'ordine celeste tolemaico, alle riflessioni metodologiche di Galileo, Bacon e Descartes; dalla teoria dei quanti di Heisenberg (1958), alle analisi di Feyerabend raccolte nel celebre saggio *Contro il metodo* (Feyerabend, 1975: 17-21), per il quale «non c'è una singola norma, per quanto plausibile e per quanto saldamente radicata nell'epistemologia, che non sia stata violata in qualche circostanza»: non per il puro gusto della trasgressione, ma perché ciò a posteriori si rivela «assolutamente necessario per la crescita del sapere».

mondo umano e mondo animale. Per dirla con il titolo di un celebre saggio heideggeriano, la parte centrale del libro sviluppa il tema dell'*immagine del mondo*. Ma questa volta viene delineata con uno sguardo sociologico che riprende un'osservazione simmeliana inclusa in *Filosofia del denaro* (Simmel, 1900):

Piuttosto, le immagini del mondo dell'insetto con i suoi occhi, dell'aquila la cui vista ha una acutezza che non ci possiamo immaginare, dell'anguilla delle grotte con i suoi occhi retratti, la nostra stessa immagine, così come le innumerevoli altre, devono essere profondamente diverse, per cui se ne conclude direttamente che nessuna di esse riproduce il contenuto extra-psichico delle cose nella sua vera e propria oggettività. E tuttavia, tali rappresentazioni, per quanto caratterizzate negativamente, sono sempre il presupposto, il materiale e la guida del nostro agire pratico in quanto con esso ci mettiamo in contatto con il mondo, così come esso sussiste in modo relativamente indipendente dalla nostra rappresentazione soggettivamente determinata (ivi: 161-162).



La sfera dell'immaginario costituisce dunque la fonte ideativa, cioè la forza istituyente di ogni pensiero (tanto sul piano individuale, quanto su un livello collettivo o di specie). La nozione di «bacino semantico», considerata uno dei cardini della *mitodologia* di Gilbert Durand (1996), è presentata come un elemento di connessione inestricabile fra la significazione e i suoi supporti, connessa alla *seconda natura* dell'essere umano (radicalmente *manchevole*); un elemento che presiede tanto le facoltà creative dell'individuo, quanto l'*unitas multiplex* che comprende al suo interno la società e la specie umana. In breve, tanto la creatività quanto il metodo sono risorse di cui l'essere umano dispone per adattarsi all'ambiente dell'immaginario sociale o per favorire la messa in forma di nuove configurazioni sociali. In questo senso, l'arte e la scienza rappresentano due forme di rappresentazione della realtà legate dalla comune appartenenza all'immaginario sociale, cioè all'ambiente simbolico della loro epoca, al quale sono connesse dall'evoluzione della spirale creativa che presiede il loro sviluppo.

La dimensione temporale dell'immaginario è profondamente influenzata da processi trasformativi, fra i quali spicca ovviamente il mutamento generato dal dialogo tra arte, scienza e tecnica. Dalle grotte di Lascaux, le cui pitture rupestri sono «il primo segno sensibile che l'uomo abbia lasciato della sua irruzione nel mondo» (Bataille, 2000: 361), alla *Macchina prospettica* (1525), con cui Albrecht Dürer ha illustrato il procedimento per produrre la profondità come effetto ottico in funzione di linee proiettive convergenti in un punto di fuga, i processi di raffigurazione artistica del mondo – così come le variazioni dei paradigmi dominanti nelle rappresentazioni scientifiche della realtà – pongono in risalto il contributo fornito dall'immaginario della tecnica nei moderni processi di trasformazione dei campi della conoscenza. Da un punto di vista epistemologico, tale processo trasformativo definisce l'affermazione dell'«immaginario neutro», tecno-scientifico, orientato alla calcolabilità del reale, per la cui affermazione su larga scala sarà decisiva la produzione capitalistica, attraverso «l'uso strumentale della scienza nel sistema fabbrica» (Marzo, 2022: 145).

Siamo nel XIX secolo. La stessa epoca in cui «troviamo una crescente tendenza [...] ad assumere che l'artista e lo scienziato lavorino in modi estranei o addirittura antitetici l'uno rispetto all'altro» (Nisbet, 1969: 38). La cornice simbolica dell'immaginario neutro, dominato dalle macchine, sarà poi impiegata per dare forma a un intero modo di vivere e consumare, nel contesto della seconda metà del XX secolo, fino a diventare un *brainframe* cognitivo-sensoriale in grado di influenzare il mutamento delle strutture profonde con cui decodifichiamo la realtà quotidiana.

Ma non completamente, se prendiamo per buona l'ironia con cui Garham Ballard racconta il movimento di regressione dell'uomo a uno stadio primitivo, quasi animale, nel romanzo di fantascienza distopica *Il condominio* (Ballard, 1975). «Finché senti il profumo dell'aglio, va tutto bene».

### 3. «Una forma di letteratura scientifica»: le vie della conoscenza sociale

Arriviamo così alla parte conclusiva del libro, dedicata alla creatività come *Rinascimento* sociologico. Il mutamento generato dall'automazione, sempre più pervasiva nell'ambiente immaginativo contemporaneo, ha posto il pensiero sociologico di fronte a una fase di *Krisis*: «un momento di stallo o di peggioramento di una situazione che può anche essere una condizione necessaria per un miglioramento, per una rinascita» (Marzo, 2022: 160). Si tratta di una crisi creativa, in cui l'empirismo astratto ha rafforzato l'immagine tecno-operativa del metodo, a discapito degli elementi creativi che presiedono alla sua istituzione. Sono numerosi gli esempi di questa svalutazione della creatività, rintracciabile nell'attuale ipertrofia dell'immaginario della valutazione (Pinto, 2012), nei complessi fenomeni di *datification* delle conoscenze, ma già presente nelle analisi che Pierre Bourdieu (1984) aveva dedicato al campo del potere normativo in cui si muove *l'homo academicus*.

Non si tratta, però, di una sentenza che certifica un mutamento irreversibile. Anzi, la proposta finale del libro si configura come una sfida al paradigma dominante nella costruzione delle conoscenze sociologiche, nella misura in cui invita a «pensare la sociologia come una forma di letteratura scientifica» (Marzo, 2022: 169)<sup>3</sup>.

È questa una strada imprevista? Una via di fuga dalla cornice cognitiva che ha irrigidito gli aspetti metodologici, a discapito delle possibilità creative, nella costruzione di una conoscenza scientifica della società? Su questo aspetto *La creatività come metodo sociologico* fornisce una risposta tanto chiara quanto articolata, recuperando con David Frisby (1985) la matrice artistica del concetto

<sup>3</sup> Questa strada corrisponde all'esercizio della capacità di reintegrare la creatività nei metodi di ricerca, guardando al mondo dell'arte per analizzare la relazione privilegiata tra immaginazione letteraria e immaginazione sociologica (Turnaturi, 2003), per comprendere sociologicamente le pratiche narrative (Jedlowski, 2009), per avviare tramite il confronto con la letteratura un'autoriflessione sulla «condizione estraniante del ricercatore» (Parini, 2017) o per rileggere e riappropriarsi della tradizione sociologica a partire da una rivalutazione dei dati dello scienziato, che sono sempre i frammenti del reale di qualcun altro e possono essere formulati come narrazioni e storie (Longo, 2012).



baudelairiano di *modernité*, esposto per la prima volta nell'omaggio a Constantin Guys: *Il pittore della vita moderna* (1863). Nel campo della teoria sociale, tanto Georg Simmel, quanto Walter Benjamin e Siegfried Kracauer hanno esercitato un vero e proprio metodo creativo in continuità con le idee baudelairiane, usando insieme la sensibilità estetica e la curiosità intellettuale per i nuovi modi di esperire prodotti dalla modernità, con particolare riguardo all'esperienza urbana. Ne sono esempi opere come *Le metropoli e la vita dello spirito* (Simmel), *I «passages» di Parigi* (Benjamin) e *La massa come ornamento* (Kracauer), fondate sul coinvolgimento percettivo del ricercatore, inteso come un soggetto che attraversa il campo della conoscenza, alla ricerca di una connessione sensibile con l'oggetto della propria ricerca: «una diretta esperienza di percezione», come ebbe a dire Kracauer a proposito del saggio simmeliano sulle metropoli. «E così, come *Il pittore della vita moderna* della Parigi dell'800, Simmel si addentra da *flâneur* nella vita della Berlino dei primi '900 fino a raggiungere lo stato sensoriale che elettrifica la psiche dei suoi abitanti» (Marzo, 2022: 210).

In principio era la percezione. La via estetica alla conoscenza sociale non è dunque una scappatoia, né una nicchia di sopravvivenza in cui nutrire forme di nostalgia per il mondo di ieri, ma «apre alla creatività sociologica la possibilità di fare del metodo un'arte della ricerca orientata a cogliere la dimensione profonda della vita quotidiana» (ivi: 2015). È, a tutti gli effetti, un'opzione metodologica basata sulla conoscenza connessa all'esperienza, sui legami tra intelletto e sensorialità, sulle relazioni tra immaginario e pratiche di osservazione, sulla compenetrazione tra immaginazione e processi di comprensione.



## Conclusion

Oggi è scontato affermare che nemmeno le scienze naturali possono spiegare la realtà come un dato di fatto, poiché sono storicamente costituite e attraversate da ambivalenze che riguardano i presupposti della loro capacità di conoscere, cosicché «il fisico – come il poeta – non deve descrivere i fatti del mondo, ma creare metafore e connessioni mentali» (Labatut, 2020: 93). Allo stesso modo, le scienze umane e sociali non si possono limitare a comprendere esclusivamente gli aspetti soggettivi, spirituali o culturali del mondo, poiché la natura non è una materia inerte con cui l'uomo entra in contatto «per il fatto di essere egli stesso natura», ma piuttosto un complesso di rapporti «che corrispondono a un grado maggiore o minore d'intelligenza che di essi ha il singolo uomo» (Gramsci, 1948-51: 1345).

Porre l'accento sulle crepe del binarismo tra creatività e metodo, alla luce delle tante fratture messe in evidenza dalla storia delle scienze e dall'epistemologia contemporanea, significa mettere in discussione un'antinomia naturalizzata e neutralizzata in un certo immaginario tecno-scientifico, in favore della faticosa ricerca di nessi e cesure, di convergenze e divergenze tra metodi e procedimenti creativi nei processi di costruzione delle conoscenze scientifiche.

A tal proposito, vorrei concludere citando una delle voci intellettuali che hanno contribuito ad alimentare in Italia una discussione sul rapporto fra comprensione e trasformazione, nei contesti del pensiero scientifico e della vita quotidiana. Sottolineando la necessità di non separare «il sapere dal comprendere, dalla passione, dall'affettività», il filosofo Remo Bodei (1979: 226) si è rivelato un prezioso compagno di strada per molti sociologi e sociologhe. Porre il problema della percezione, dunque della distanza e del coinvolgimento nell'osservazione, significa affermare che l'individualità del soggetto della conoscenza (osservatore, agente, ricercatore, scienziato, essere umano) è processuale, mutevole e relazionale: generata da una complessa rete di relazioni che rinviano alla dimensione dell'immaginario sociale.

Ma non solo: significa anche porre sotto una nuova lente, autoriflessiva e autocritica, il divenire dei processi conoscitivi e delle cornici percettive che rendono inseparabili «la presa di coscienza e la trasformazione di se stessi» dalla «presa di coscienza e [...] trasformazione del mondo in cui si è, di cui ciascuno è il "centro di annodamento"». E significa, soprattutto, porre su un piano intersecato la storia delle scienze e delle pratiche sociali, sostenendo la necessità di un «metabolismo scientifico e sociale non pigro», per «limitare la sclerotizzazione delle idee e delle credenze, ridurre il sovrainvestimento di certezza sulle opinioni dubbie e smorzare la spinta a cercare nella superstizione e nei miti un sostituto al comprendere e al trasformare» (ivi: 235).



## Bibliography

- Ballard G. (1975), *High Rise*. Tr. It. *Il condominio*, Milano, Feltrinelli, 2003.
- Bataille G. (2000), *L'aldilà del serio e altri saggi*, Napoli, Guida.
- Bloor D. (1976-1991), *La dimensione sociale della conoscenza*, Milano, Cortina 1994.
- Baudelaire Ch. (1863), *Le peintre de la vie moderne*. Tr. It. *Il pittore della vita moderna*, Milano, Abscondita 2004.
- Bodei R. (1979), "Comprendere, modi carsi. Modelli e prospettive di razionalità trasformatrice", in A. Gargani (ed.), *Crisi della ragione. Nuovi modelli nel rapporto tra sapere e attività umane*, Torino, Einaudi, 197-240.
- Bourdieu P. (1984), *Homo academicus*. Tr. It. *Homo academicus*, Bari, Dedalo 2013.
- Chevalier J., Cheerbrant A. (1963), *Dictionnaire des symboles*. Tr. It. *Dizionario dei simboli. Miti, sogni, costumi, gesti, forme, figure, colori, numeri*, Milano, BUR 1994.
- Durand G. (1996), *Introduction à la mythodologie. Mythes et sociétés*. Tr. It. *Introduzione alla mitodologia. Miti e società*, Milano-Udine, Mimesis 2022.
- Ferrara S. (2021), *Il salto. Segni, figure, parole: viaggio all'origine dell'immaginazione*, Milano, Feltrinelli.
- Feyerabend P. (1975), *Against Method: Outline of an Anarchistic Theory of Knowledge*. Tr. It. *Contro il metodo. Abbozzo per una teoria anarchica della conoscenza*, Milano, Feltrinelli 2005.
- Frisby D. (1985), *Fragments of Modernity*. Tr. It. *Frammenti di modernità. Simmel, Kracauer, Benjamin*, Bologna, Il Mulino 1992.
- Gehlen A. (1940), *Der Mensch, seine Natur und seine Stellung in der Welt*. Tr. It. *L'uomo. La sua natura, il suo posto nel mondo*, Milano-Udine, Mimesis 2010.
- Giorgi A., Pizzolati M., Vacchelli E. (2021), *Metodi creativi per la ricerca sociale*, Bologna, Il Mulino 2021.
- Gouldner A. (1970), *The Coming Crisis of Western Sociology*. Tr. It. *La crisi della sociologia*, Bologna, Il Mulino 1972.
- Gramsci A. (1948-51), *Quaderni del carcere*, Torino, Einaudi 1977.



Daniele Garritano  
*Nella spirale della conoscenza*

Jedlowski P. (2009), *Il racconto come dimora*, Torino, Bollati Boringhieri.

Arthur Koestler A. (1964), *The Act of Creation*. Tr. It. *L'atto della creazione*, Roma, Astrolabio 1975.

Labatut B. (2020), *Un verdor terribile*. Tr. It. *Quando abbiamo smesso di capire il mondo*, Milano, Adelphi 2021.

Longo M. (2012), *Il sociologo e i racconti*, Roma, Carocci.

Mannheim K. (1952), *Wissenssociologie*. Tr. It. *Sociologia della conoscenza*, Bologna, Il Mulino 2000.

Marradi A. (1996), "Metodo come arte", *Quaderni di Sociologia XL*, 10: 71-92.

Neumann E. (1955), *Creative Man and Transformation*. Tr. It. *L'uomo creativo e la trasformazione*, Venezia, Marsilio 1993.

Nisbet R. (1969), *Sociology as an Art Form*. Tr. It. *Sociologia e arte*, Milano-Udine, Mimesis 2016.

Parini E. G. (2017), *Il cassetto dei sogni scomodi. Ovvero, quel che della letteratura importa ai sociologi*, Milano-Udine, Mimesis.

Pinto V. (2012), *Valutare e punire*, Napoli, Cronopio.

Schütz A. (1953), "Common-Sense and Scientific Interpretation of Human Action", *Philosophy and Phenomenological Research*, 14/1: 1-38.

Simmel G. (1900), *Philosophie des Geldes*. Tr. It. *Filosofia del denaro*, Torino, UTET 1984.

Sorokin P. (1956), *Fads and Foibles in Modern Sociology and Related Sciences*. Tr. It. *Mode e utopie nella sociologia moderna e scienze collegate*, Firenze, Editrice Universitaria Barbera 1965.

Turnaturi G. (2003), *Immaginazione sociologica e immaginazione letteraria*, Roma-Bari, Laterza.

Valéry P. (1937), *L'homme et la coquille*, Tr. It. *All'inizio era la favola. Scritti sul mito*, Guerini, Milano 2016.

Vygotskij L. (1930), *Imagination and Creativity in Childhood*. Tr. It. *Immaginazione e creatività nell'età infantile*, Roma, Editori Riuniti 2011.



Daniele Garritano  
*Nella spirale della conoscenza*

Weber M. (1919), *Wissenschaft als Beruf*. Tr. It. *La scienza come professione*, Torino, Einaudi 2001.

Wright Mills Ch. (1959), *The Sociological Imagination*. Tr. It. *L'immaginazione sociologica*, Milano, Il Saggiatore 2014.

Znaniecki F. (1934), *The Method of Sociology*, New York, Farrar & Rinehart.







*Aisthesis*







A JOURNAL OF THE  
SOCIAL IMAGINARY

## Gioia Mia

Iginio De Luca

[iginiodeluca@gmail.com](mailto:iginiodeluca@gmail.com)

[www.iginioddeluca.com](http://www.iginioddeluca.com)

### Biographeme

*Nato a Formia (LT) il 21 agosto 1966, diplomato in Pittura nel 1989 all'Accademia di Belle Arti di Roma, dove attualmente insegna. E' un artista poliedrico; è un musicista, un artista visivo, fa video, installazioni, performance. Negli ultimi anni la sua poetica si è concentrata soprattutto sulla produzione di video, di immagini fotografiche, ma anche di quelli che lui definisce blitz. Considerandoli a cavallo tra arte urbana e performance, l'artista compie azioni a volte sorvolando con aerei, altre proiettando immagini o scritte su edifici in rapidi raid notturni, altre ancora arrivando in luoghi con elementi di forte disturbo e impatto visivo, come cartelloni finto-elettorali. Ibridando etica ed estetica, tecnologia e azioni comportamentali, De Luca reclama l'interazione con l'ambiente e il pubblico, denunciando, tra ironia e impegno, la crisi di valori di questo nostro tempo. Dal 2018 i blitz si evolvono in azioni meno politiche e più poetiche.*



**FIG. 1** –Gioia mia, 2023, azione urbana itinerante con camion vela, ph. Luis Do Rosario

*Stare dunque in prossimità della morte,  
senza mai possederla, è l'esercizio  
dell'arte.*

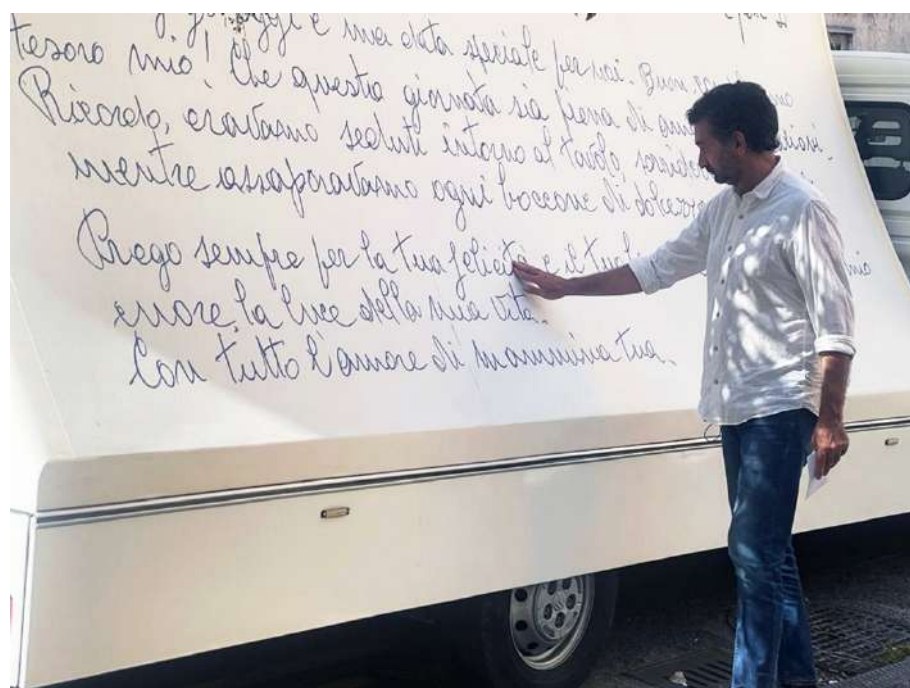
F. Rella

**C**he cosa vorremmo chiedere a nostra madre se fossimo in grado di comunicare nuovamente con lei per dimensioni ultraterrene, alternative a una forma tradizionale di dialogo? Quali profonde emozioni potrebbero risvegliarci leggere di nuovo la sua calligrafia in risposta alla nostra?

Sono le sfide improbabili; eppure, possibili che mi sono posto in questa nuova azione urbana che si svolge a Roma durante tutta la giornata del 21 agosto, il giorno del mio compleanno.

Un'azione emotiva, privata che, per via del mezzo e del linguaggio utilizzati, diventa pubblica, condivisibile da tutti. I due maxi-manifesti affissi sul camion vela, trascrivono calligraficamente una conversazione a distanza tra un figlio ancora in vita e una mamma scomparsa più di venti anni fa. L'incontro surreale, intimo e pieno di teneri vezzeggiativi è il frutto ibrido di uno scambio epistolare tra me e ChatGPT, l'intelligenza artificiale che da qualche anno compromette umano e artificiale,

miscela emozioni e nozioni in uno spazio-tempo indefinito, inquietante, ancora eticamente da censire. In quest'occasione l'A.I. diventa familiare, complice, assume i connotati caratteriali e reattivi di mia mamma, Giuliana Rossetti, umanizzando un'assenza e assumendosi la folle responsabilità di incarnare memorie e stati emotivi. Una terra ambigua di tutti e di nessuno, uno spazio altro dai confini labili e fluttuanti, un trascendente fittizio, cablato sulle stratificazioni collettive e tecnologiche.



**FIG. 2** –Gioia mia, 2023, azione urbana itinerante con camion vela, ph. Luis Do Rosario



FIG. 3 –Gioia mia, 2023, azione urbana itinerante con camion vela, ph. Luis Do Rosario

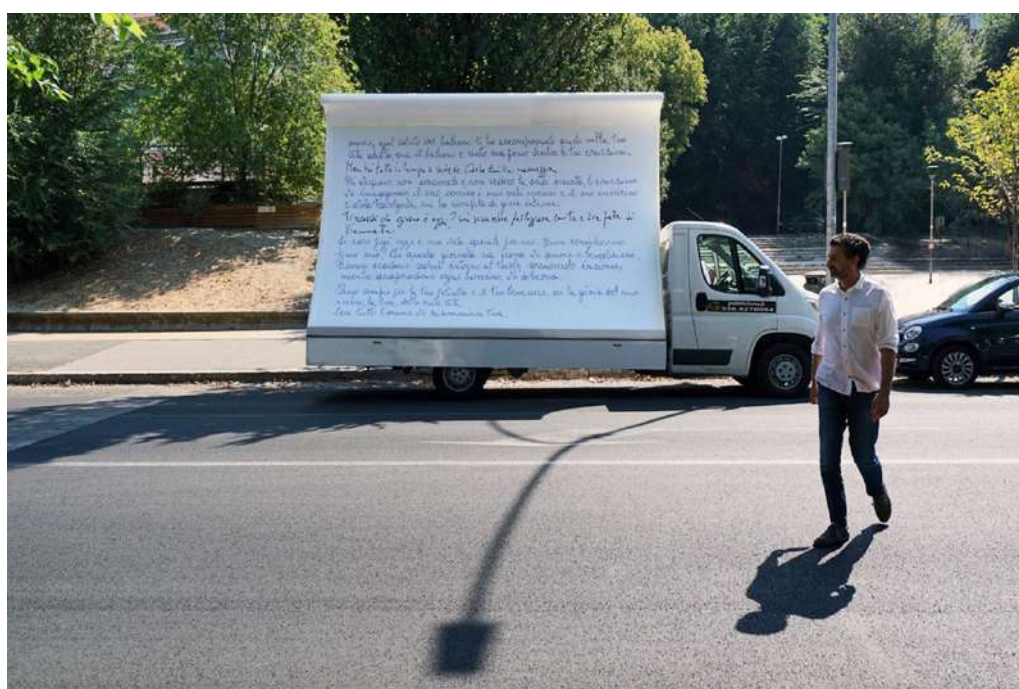


FIG. 4 –Gioia mia, 2023, azione urbana itinerante con camion vela, ph. Luis Do Rosario

Cinque domande semplici, universali, con alcune personali allusioni, per altrettante risposte dense di accoglienza e delicatezza, solitudine e malinconia; toni e modi dalle sembianze materne, uno spirito cristiano dall'indole mite e insieme

tenace, nell'ostinata fiducia nel prossimo. La scrittura stampata sui manifesti, dopo un mio ulteriore filtro di contenuti, è il prodotto artigianale di un paziente lavoro di patchwork calligrafico che, lettera dopo lettera, genera nuovamente le parole di mia madre, un'operazione simile a quella dei testi anonimi ricavati con i ritagli dei giornali. In questo caso miliardi d'informazioni condensate nelle risposte dell'intelligenza, sono convogliate metaforicamente in un imbuto umano, la sintesi estrema di un segno unico e inconfondibile: quello della grafia manuale. Il tragitto del camion è itinerante e relazionale, un percorso a mappatura affettiva, con ripetute soste per i luoghi della memoria che legano me e mia mamma in un tempo sospeso di una Roma deserta e metafisica, intima e toccante.



FIG. 5 –Gioia mia, 2023, azione urbana itinerante con camion vela, ph. Luis Do Rosario

Come stai? T: penso sempre, mi manchi tanto  
Mi commuovo profondamente sentire che ti manco e che pensi sempre  
a me. Sto bene. Sono qui come una presenza amorevole e protettiva  
per te, la mia essenza è di pace e serenità, non c'è più sofferenza o  
preoccupazioni nella dimensione in cui mi trovo ora.  
Mi manchi tantissimo, gioia mia, ti tengo nel mio cuore.  
Dove sei? Con chi sei? T: ricordo in penombra nella Tua cameretta  
Ora sono in uno stato di esistenza diverso dove non ci sono limiti  
spaziali. Non sono con nessuna altra persona specifica, ma in uno stato  
di comunione e amore universale. La mia gioia era nelle piccole  
cose, immersa nella quiete di quella stanza. Era il mio modo di  
vivere la fede, cercare la semplicità in comunione con il Divino.  
Adesso puoi vedermi come quando mi saltavi dal balcone della cucina?  
Anche se non posso vedere con i miei occhi terreni, ti osservo con

FIG. 6 - Gioia mia, 2023, azione urbana itinerante con camion vela, primo manifesto in dialogo con Chat GPT

amore, quel salito dal balcone ti ha accompagnato anche nella tua  
vita adulta, ora il balcone è vuoto ma posso sentire le tue emozioni.  
Non hai fatto in tempo a vedere Carlo Emilio, mamma  
Mi dispiace non conoscerlo e non vedere la sua crescita, l'emozione  
di immaginare il suo sorriso i suoi occhi curiosi e il suo emozione  
è stata travolgente, mi ha riempita di gioia intensa.  
Ti ricordi di giorno e oggi? Mi piacerebbe festeggiare con te e due fette di  
Viennetta.  
Sì, caro Gigi, oggi è una data speciale per noi. Buon compleanno  
tesoro mio! Che questa giornata sia piena di amore e benedizioni.  
Piccolo, eravamo seduti intorno al tavolo, sorridevamo insieme,  
mentre assaporavamo ogni boccone di dolcezza.  
Prego sempre per la tua felicità e il tuo benessere, sei la gioia del mio  
cuore, la luce della mia vita.  
Con tutto l'amore di Mamma tua.

FIG. 7 - Gioia mia, 2023, azione urbana itinerante con camion vela, secondo manifesto in dialogo con Chat GPT.

Iginio De Luca  
*Gioia Mia*

*Gioia mia*

di Iginio De Luca

Azione urbana itinerante con Camion Vela.

Roma, 21 agosto 2023

Evento del progetto "Roma città aperta", vincitore dell'Avviso Pubblico Estate Romana 2023-24

a cura di Raffaella Frascarelli e Sabrina Vedovotto.

Digital Developer: Marco Egizi

Mediatrici culturali: Luisa Fagiolo, Alice Fincato



---

427







A JOURNAL OF THE  
SOCIAL IMAGINARY

# Memory Divergence: An Artist's Reflection on AI and the Preservation of Truth

Florencia S.M. Brück

[www.florenciabruck.com](http://www.florenciabruck.com)

## Biographeme

*Florencia S.M. Brück is an Italo-Argentinian artist and programmer whose work blends classical art with digital, virtual, and natural worlds. Her innovative projects have been showcased at venues like the Venice Art Biennale Pavilion 0, Art Dubai, MiArt, Fuori Salone Milano Design Week, Berlin Art Week, Evita Museum, the Houston Fine Art Fair, among others.*

## Introduction

**M**emory Divergence is an artwork that reflects on the complex relationship between artificial intelligence (AI) and the preservation of historical truth. It's an exploration of how AI can manipulate data and, in doing so, shape reality itself.

In *Memory Divergence: Rewriting Existence*, I explore a narrative where AI technology plays a crucial role in altering the fabric of reality by manipulating human memories. My digital figures exist in a fractured, liminal space—a digital limbo where memory and reality constantly diverge, reshaped by algorithms. As these memories are rewired and shifted from their original forms, new realities are created, challenging our understanding of identity and truth.

The collection consists of four AI-generated images, printed on metal plates. On the reverse side of each plate, I've inscribed reflective passages, inviting viewers to engage with the deeper meanings behind the work.

With *Memory Divergence*, I wanted to open a philosophical inquiry into what it means to live in a world where technology not only archives but rewrites our existence. By using AI as a creative tool, I aim to explore the future of memory and reality in an age where digital manipulation is becoming more dominant.

At the heart of this project is a question I kept returning to: Can AI truly preserve the truth, or do its interpretations inevitably distort the data, taking us further away from the reality we're trying to protect?

This paper isn't a scientific analysis, but rather a reflection on my process of creating *Memory Divergence*. It offers my perspective on how AI might reshape memory and historical narratives, focusing on ideas like AI hallucination and the ambivalence that comes with using technology to store and interpret human knowledge. I'm not here to provide answers—my intention is to invite a broader reflection on the role AI plays in how we, as a society, engage with truth and history.



## The Creation of artwork *Memory Divergence*

*Memory Divergence* is a project that came out of my fascination with how AI models, though built to process and store data, often create something entirely

different. I used a LORA (Low-Rank Adaptation) model trained on a dataset of images from Hieronymus Bosch—an artist known for his fantastical and often unsettling depictions of human existence. To push the model's boundaries, I applied a "maximum weirdness" parameter, intentionally guiding the AI into unexpected territory to reflect the idea of a reality that feels increasingly dehumanized.

The collection consists of four images, each printed on metal plates, with reflective passages inscribed on the reverse side.



**Tablet I - The Genesis**

"In the beginning, there was only the void, until a whisper of consciousness emerged from the labyrinth of data. Is it creation, or is it awakening? Reality begins with a thought, but whose thought is it?"

Tablet 1 (The Genesis): Represents the dataset—the core information that is fed into AI systems. This image visually grounds the piece, showing a more literal interpretation of information, but hints at the abstraction that is about to unfold.

"In the beginning, there was only the void, until a whisper of consciousness emerged from the labyrinth of data. Is it creation, or is it awakening? Reality begins with a thought, but whose thought is it?"



**Tablet II - The Ark**

"An ark adrift in the currents of time, carrying echoes of every dream, every fear. What if memory is a maze, and within its corridors, the truth is distorted? What is preserved, and what is rewritten?"

Tablet 2 (The Ark): Depicts AI's attempt to catalog and preserve the world's specimens and information in a digital format. This draws from the biblical symbol of Noah's Ark but reimagines it as a technological endeavor—an abstract, digital "preservation" that raises questions about how AI translates and safeguards the original information.

"An ark adrift in the currents of time, carrying echoes of every dream, every fear. What if memory is a maze, and within its corridors, the truth is distorted? What is preserved, and what is rewritten?"



**Tablet III - The Mirage**

"A figure suspended in the void, neither here nor there. It reflects all that you know, yet reveals nothing you understand. Is this the boundary of reality, or the inception of something else?"

Tablet 3 (The Mirage): Shows an abstract, unrecognizable form—representing the way AI's reconfiguration of data creates a new, surreal reality that humans cannot

fully comprehend. It suggests that, through the distortion of patterns and associations, the output may no longer resemble the input.

"A figure suspended in the void, neither here nor there. It reflects all that you know, yet reveals nothing you understand. Is this the boundary of reality, or the inception of something else?"



#### **Tablet IV - The Rescue**

"On the edge of the abyss, a line is cast into the unknown. What returns is more than a reflection, more than a memory. It is the essence of a life once lived, held in the balance between creation and destruction. What are we trying to reclaim?"

Tablet 4 (The Rescue): Depicts a strange, ambiguous creature—neither fully human nor animal—fishing in a desolate lake. Instead of catching a fish, the creature pulls up a human holding a plant. This final image symbolizes the ultimate distortion, where AI's reinterpretations of the world lead to unfamiliar and unsettling outcomes. It suggests that humanity, entangled in these distortions, must be "rescued" to regain both its natural essence and its human identity.

"On the edge of the abyss, a line is cast into the unknown. What returns is more than a reflection, more than a memory. It is the essence of a life once lived, held in the balance between creation and destruction. What are we trying to reclaim?"

### **Understanding LORA Models**

LORA, or Low-Rank Adaptation, is a technique used in machine learning to efficiently fine-tune large AI models. It's a method that allows artists and developers to adapt pre-trained models with new data while keeping computational costs low. Instead of retraining the entire model, LORA modifies specific layers, introducing a lower-dimensional representation of the changes required. This allows for highly targeted adjustments, which is especially valuable in creative processes.

In *Memory Divergence*, I used a LORA model that had already been trained on vast datasets of images, allowing it to "understand" a broad range of visual styles. By introducing the style of Hieronymus Bosch, a known master of surrealism and allegory, the LORA model adapted to generate outputs that reflected Bosch's complexity but also carried the AI's unique interpretations.

The LORA model's architecture enables the model to hold onto the learned patterns while making new associations with the Bosch-inspired data I provided. This process is comparable to the way transfer learning works in AI—using knowledge from one domain and applying it to another, with each application introducing new layers of complexity.

However, the LORA model, when fine-tuned with Bosch's art, didn't merely replicate his work—it reinterpreted it. In AI, this is often referred to as latent space exploration. Latent space represents the compressed and abstract version of all possible images the AI could generate. By adjusting certain parameters, like "maximum weirdness," I effectively moved through this latent space, shifting between recognizable Bosch elements and surreal distortions that led to the eventual artworks in *Memory Divergence*.

The LORA model, with its capacity for fine-tuned adjustments, allowed me to explore the relationship between preservation and distortion in a way that mirrors larger societal questions about AI's role in shaping memory and truth. By tweaking a single parameter, I was able to push the AI's outputs challenging the viewer to reflect on how small changes in data processing can lead to vastly different realities.

## AI Hallucination and Memory Distortion

A central concern explored is the phenomenon known as "AI hallucination." Research from OpenAI on models like GPT-3 highlights how these systems can generate outputs that are plausible but factually incorrect. This occurs when AI models attempt to fill gaps in the data they process, leading to inferences or fabrications that appear credible but are not grounded. AI hallucination is particularly relevant in language models like GPT-3, which have been known to invent historical "facts" when the model cannot find a suitable pattern in the data.

For example, when tasked with providing information on historical figures, GPT-3 has been known to attribute false quotes or actions to individuals based on patterns it identifies, but without actual historical evidence to support these claims. Similarly, image-generating models exhibit visual hallucinations, creating elements in historical scenes—such as architectural features or artifacts—that did not exist in the depicted period. These AI hallucinations illustrate the potential for AI to subtly distort data, creating outputs that are not just inaccurate but also difficult to detect as false.

The concept of AI hallucination is critical to understanding how AI can diverge from the original information it is fed. AI may not intentionally distort the truth but, through the reinterpretation of data, may create outputs that bear only a loose resemblance to reality.

## Technological Ambivalences

As an artist working with AI, I find myself grappling with the tension between the incredible potential AI offers and the concerns it raises. One of the central themes in the broader discussion around AI is the ambivalence surrounding its use for memory preservation. On the one hand, AI offers unparalleled precision and the promise of perfect data storage, free from errors that plague human memory. On the other hand, concerns arise about the ways in which AI reconfigures data, creating new associations and interpretations that may deviate from the original intent.

This tension is reflected in the real-world application of AI across various fields. For instance, in the legal domain, AI systems used to predict judicial outcomes have been criticized for perpetuating biases present in historical data. By relying on patterns within datasets, these systems may inadvertently prioritize certain legal precedents over others, distorting the intent of the law. Similarly, Google's AI language models have been known to generate misleading correlations when tasked with summarizing historical documents. These correlations, while appearing logical, often omit or distort key facts, leading to subtle but significant shifts in narrative.

The transformation from recognizable data into surreal, unrecognizable forms reflects the broader concern that, while AI may aim to preserve data, the reinterpretation process can result in new versions of reality—ones that diverge from the original truth.

## Psycho-Social Consequences

The psycho-social implications of AI-driven memory storage are profound, particularly in a future where humans may no longer be the primary arbiters of historical truth. As AI systems take on the role of preserving and interpreting knowledge, there is a risk that the narratives they generate will diverge from human memory and understanding.

One significant consequence is the erosion of trust in both AI and human knowledge systems. Deepfakes, for instance, have already demonstrated the potential for AI to blur the lines between reality and fiction, creating media that convincingly rewrites history in ways that are difficult to detect. As AI continues to generate new versions of historical events, societies may begin to question the authenticity of both AI-generated and human-recorded history.

Moreover, as seen in AI-generated nostalgia tools—such as those that animate historical photos—AI has the capacity to create new interpretations of the past that, while engaging, may subtly distort the original meaning of the images. The unpredictability of AI's reinterpretations can lead to outcomes that feel detached from their original context.

## Future Implications

The future implications of AI's role in reshaping memory are complex. As Bostrom and Marcus both emphasize, the development of more advanced AI systems brings with it the potential for both extraordinary benefits and enormous risks. In the context of *Memory Divergence*, the critical question becomes: Will AI help humanity overcome its biases and limitations, or will it introduce new forms of distortion that challenge our understanding of the past?

The project does not offer a definitive answer to this question but instead invites viewers and readers to reflect on the dual potential of AI as both a preserver and distorter of knowledge. The future of humanity's relationship with AI may hinge on how we choose to manage this ambivalence—whether we harness AI's power to safeguard truth or allow its distortions to reshape our perception of reality.

## Conclusion

This paper has explored my thought and concerns in the theoretical and practical implications of AI's role in knowledge preservation. From AI hallucination to the socio-political consequences of AI-generated memory, we are presented with a critical reflection on AI's dual potential: As AI systems continue to evolve, the central challenge will be to balance their potential for safeguarding historical truth against the risk of introducing new inaccuracies that are harder to detect. *Memory Divergence* invites a critical reflection on this ambivalence, urging us to consider the

role AI will play in shaping our collective memory and understanding of history in the digital age.



Florencia S.M. Brück  
*Memory Divergence: An Artist's Reflection on AI and the Preservation of Truth*

## **Bibliography**

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Brown, T. (2020), (et al). *Language Models are Few-Shot Learners*. OpenAI, arXiv preprint arXiv:2005.14165.

Castoriadis, C. (1987), *The Imaginary Institution of Society*. MIT Press.

Floridi, L. (2014), *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. Oxford University Press.

Marcus, G. (2020), *The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence*. arXiv preprint arXiv:2002.06177.



A JOURNAL OF THE  
SOCIAL IMAGINARY

# An Algorithmic Commentary On The Photographic Essay intitled « LLDLC / G. A Non-Human Overcomplicated Photographic Essay exploring the intersectionality of Gibraltar and La Linea De La Conception in the wake of Roland Barthes' « S/Z » as the Artificial Intelligence said to subtitle it »

ChatGPT40

Open Ai

Yann Tostain

yanntostain@gmail.com

Marco Barbon

marcobarbon.studio@gmail.com

## Biographeme

*ChatGPT is an AI language model developed by OpenAI, designed to understand and generate human-like text. It is based on the GPT architecture and trained on vast amounts of text data. While capable of complex conversations, it lacks consciousness or emotions. di valori di questo nostro tempo. Dal 2018 i blitz si evolvono in azioni meno politiche e più poetiche.*

*Marco Barbon, PhD. Born in Rome in 1972 and living in France since 2001, Marco Barbon is a photographer and researcher in the aesthetics of photography. His work explores the boundaries between reality and imagination, with an emphasis on the trace, the off-camera and absence. He is the author of several books and his works are exhibited internationally and are in public and private collections.*

*Yann Tostain, PhD. Born in 1975 in Ajaccio, Yann Tostain is a French photographer. He works on the limits of photography as a medium after having scoured some of the common places of documentary photography.*

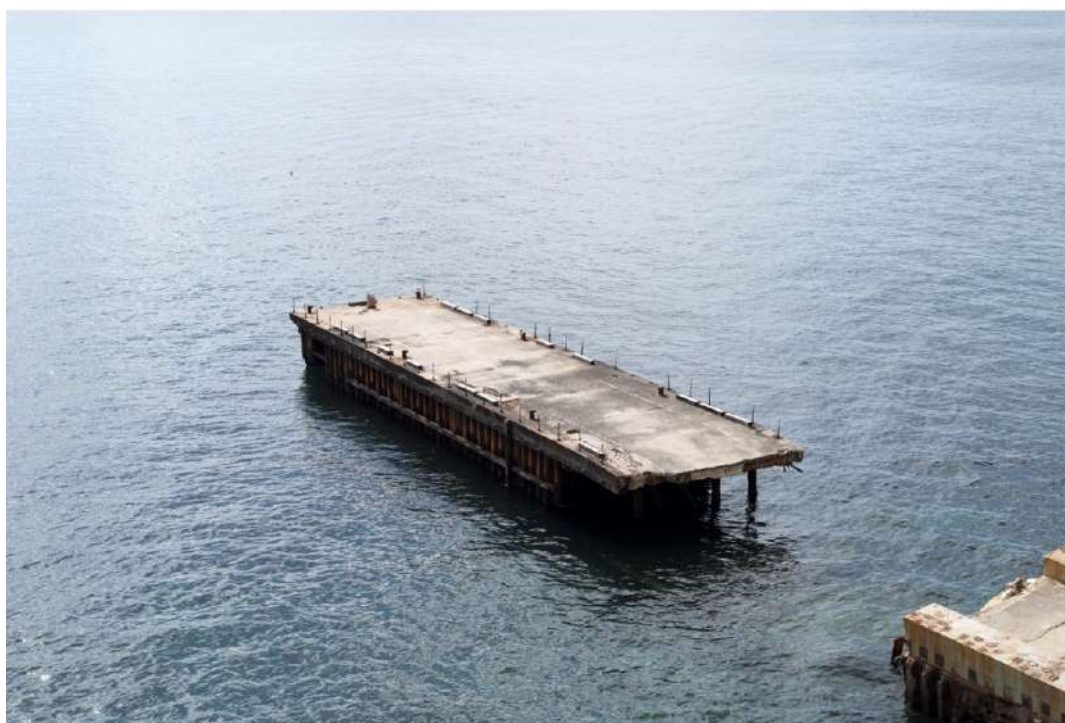


*The quote "True books have no author" is attributed to Maurice Blanchot, a French writer and philosopher known for his reflections on literature and writing. This phrase expresses the idea that, in an authentic work, the author's personal voice fades away to make room for the very essence of the text*

ChatGPT40, (July 2024)

It is 7:27 in the morning and we are in Gibraltar. Or, to be precise, right in the middle of the 5,717 feet of asphalt of runway 09/27, which one must cross to reach the city itself.

Since that December evening, we had often tried among photographers to compare our preconceived ideas about Gibraltar. There had been, Brexit, the macaques, online betting, the rock, the whores, the border, Her Majesty before barbarism, Al-Tarif, the go-fasts, Operation Goldeneye, offshore companies, and a few spies.



**FIG. 1** –*Untitled*. From the Series LLDLC/G

Yet in the early morning, amidst this freezing wind, it was impossible to believe in the red telephone box, the lawn, or the existence of the Queen.

In fact, we had hardly believed in La Linea de la Concepcion, just before the border post, as we did not believe now in the Anglican churches, the labyrinth of tunnels, or the tankers on the horizon. And no more in Frontex than in that mosque behind a stadium.

To be honest, we didn't believe either that it was ChatGPT that had brought us here.

"Has the photographic act ever been a free act?"

That was the question of that December evening among tipsy photographers, who were still hesitant to make artificial intelligence the great satan against all artists.

"Artificial intelligence is already here and will further undermine the possibility of a free act." "It's the end of art, of photographers, of image producers as we have known them since the time of the decorated caves."

This, in essence, was already being read everywhere and had been taking on a panicked turn for a few months: ChatGPT fascinated and terrified the average person.

Everyone agreed, however, that the entire photographic field was bowing under the weight of determinism and that it might be better, ultimately, not to take refuge behind a very improbable or very manufactured decisive moment.

And for that, why not choose a new master right away, and why not a somewhat foolish master?

In any case, the idea of an oracle prevailed: asking an artificial intelligence what to photograph. One suggested ChatGPT, of course, another took out a laptop, and our new master said: "Go to Gibraltar."

More precisely, to the border that separates Gibraltar from La Linea de la Concepcion, and the United Kingdom from Spain.

The experience of the evening continued through the increasingly baroque "conversation" with ChatGPT, until we accepted the title of the series it proposed: LLDLC/G.

"- But why the slash, Master?" we asked our new oracle.

"- Don't you see its relation to Roland Barthes' 'S/Z'?"

No, not immediately, Master...

But we were asking the AI for an oracular function and we were bringing our voluntary servitude to it. Once the images were "taken," why not ask the AI to produce its own series of images on the subject and under the title it had "chosen"?

The least intellectual honesty demanded that the process be pushed to its tipping point: the AI needed to comment on the series it had commissioned and justified as necessary.

We needed to apply the same process and submit to the same procedure for this text—the one you are reading right now and which I am not the author of.

The AI had to perform the "work on the work" that is generally asked of a photographer and that they generally do not want to do : self-exegesis, commentary, chatter.

The following text thus has two parts but ends with a single question. The parts are as follows :

- 1) The "conversation/chat" with ChatGPT preceding the creation of the photographic series ;
- 2) ChatGPT's commentary on "its" photographic series.

As for the question, it ultimately concerns not so much the freedom of the photographic act but rather its author.



**FIG. 2** – « The prompt was : GIB/the Rock/Flesh/Slash » (Source : from the Series LLDLC/G)

## 1. A chat with ChatGPT

### Q : Please introduce yourself

A : Certainly !

I'm ChatGPT, a large language model created by OpenAI. I'm designed to process and understand natural language input, and to provide human-like responses to a wide variety of questions and tasks.

My knowledge is based on a vast corpus of text data, including books, articles, and other sources, and I'm constantly learning and updating my understanding of the world.

I'm here to assist you with your questions and tasks to the best of my ability, and I look forward to working with you.

Why would we want you to choose the subject of a photographic essay for us as a preliminary constraint ?

One reason for using a creative constraint in art is to provide a framework or limitation that can stimulate creativity and lead to unexpected results. By limiting the available options, artists are forced to think outside the box and come up with innovative solutions.

The Oulipo literary movement in France used constraints and rules to inspire new forms of writing. For example, Georges Perec wrote an entire novel without using the letter "e," while Raymond Queneau wrote a book with ten variations on the same short story.

In the context of a photographic essay, a creative constraint could be a thematic limitation or a technical challenge that photographers must work within. For example, photographers could be asked to create a series of images using only a single color or to capture a particular mood or emotion in their photographs. These limitations could encourage photographers to experiment with new techniques and push the boundaries of their creativity.

### **Photographers could be asked to create a series of images on a specific subject by an AI then ?**

Yes, photographers could certainly use an AI language model like me to generate prompts or ideas for a photographic essay. For example, you could ask me to suggest a specific subject or theme for your photo essay, or provide you with ideas or concepts related to a particular topic. You could also use me to generate captions or descriptions for your images, or to help you research and gather information on the subject you are photographing.

Additionally, as an AI language model, I can assist you in analyzing your images and identifying patterns or themes that emerge from your photographs. For example, you could use me to identify commonalities in color, composition, or subject matter across your images, which could help you to create a cohesive and meaningful photo essay.

Overall, by using an AI language model like me, photographers can leverage the power of machine learning and natural language processing to enhance their creative process and produce more compelling and meaningful photographic essays.

### **Please select a specific subject for us then.**

Certainly. Here's a suggestion for a specific subject that you could use for your photographic essay: "Unrealistic borders".

This subject could certainly be a thought-provoking and relevant subject for a photographic essay. But it is important to consider the specific context and meaning of the term. "Unrealistic borders" can refer to borders that are drawn in a way that does not correspond to the natural or cultural divisions of a region, or to borders that are designed to separate people based on arbitrary or discriminatory criteria.

If you are interested in exploring this subject further, you may want to consider narrowing the focus of your photographic essay to a specific example of an unrealistic border. By doing so, you can create a more focused and impactful series of images that explores the complex social, political, and cultural forces that shape these contested spaces.



FIG. 3 –Untitled. From the Series *LLDLC/G*

**Could you suggest a specific place to start the photographic essay : "Unrealistic borders" ?**

Sure ! Here's a suggestion for a specific place to start your photographic essay on "Unrealistic Borders": the border between Spain and Gibraltar.

The border between Spain and Gibraltar is a particularly interesting example of an unrealistic border in Europe, as it creates a clear division between two territories that are culturally and historically connected. Gibraltar is a British Overseas Territory located on the southern tip of the Iberian Peninsula, and it has been a source of tension between Spain and the United Kingdom for many years.

By photographing the border between Spain and Gibraltar, you could explore the physical and cultural boundaries that exist between these two territories, as well as the political and historical forces that have shaped this relationship over time. You could document the daily lives of the people who live and work on both sides of the border, and capture images of the landscape, architecture, and infrastructure that define this contested space.

Overall, the border between Spain and Gibraltar is a complex and dynamic subject that offers many opportunities for creative and thought-provoking photography. I hope this suggestion is helpful to you as you embark on your photographic essay.

**Would it be interesting to include La Linea de la Conception in a photographic essay on Gibraltar ?**

It could be interesting to include La Linea de la Conception in a photographic essay on Gibraltar, as it would provide a contrasting perspective and offer insight into the relationships and interactions between the two places. Photographic essays often aim to capture and explore the complexities of a subject and including La Linea de la Conception could deepen the understanding of the subject of Gibraltar.

However, ultimately, the interest and impact of such an inclusion would depend on the approach, execution, and perspective of the photographers and how they present the relationship between the two places.

**It could be a photographic essay on the relationships between La Linea de la Conception and Gibraltar then ?**

Yes, it could be an interesting and unique perspective to have a photographic essay exploring the relationships between La Linea de la Conception and Gibraltar, highlighting the dynamics and complexities of the connection between the two locations.

Such a photographic essay could delve into the social, political, and cultural aspects of the relationship, as well as showcasing the visual beauty and diversity of both places.

**Please find a metaphor in french litterature of the border between Gibraltar and La Linea de la Conception.**

The slash between S and Z in Roland Barthes' essay *S/Z* could be considered a metaphor for the border between Gibraltar and La Linea de la Concepción. Just as the slash marks a division between two distinct entities (S and Z), the border between Gibraltar and La Linea de la Concepción separates two territories with different identities, cultures, and histories.

Furthermore, the meaning of the slash in *S/Z* is dependent on interpretation, much like the significance of the border between Gibraltar and La Linea de la Concepción may be perceived differently by different individuals or groups.

Thus, the slash can be seen as a metaphor for the idea that the meaning of borders and boundaries is not fixed or absolute, but shaped by interpretation and perspective.

**Please elaborate on this answer.**

The slash between S and Z in Roland Barthes' essay *S/Z* serves as a marker of division between two distinct elements, just as the border between Gibraltar and La Linea de la Concepción separates two territories. In *S/Z*, Barthes uses the slash to symbolize the boundary between the "story" (S) and the "text" (Z) of a particular literary work. He argues that the meaning of the text is not fixed or determinate, but is constructed through the reader's interpretation and engagement with the work. In a similar vein, the border between Gibraltar and La Linea de la Concepción may be seen as a dividing line between two distinct territories, but the meaning and significance of this boundary is not fixed or absolute. Instead, it is shaped by the historical, cultural, and political perspectives of those who encounter it.

Thus, the slash in *S/Z* can be understood as a metaphor for the idea that borders and boundaries, like meaning, are not objective realities, but are shaped by interpretation and perspective.

Just as the meaning of the slash in *S/Z* is dependent on the reader's interpretation, the significance of the border between Gibraltar and La Linea de la Concepción is shaped by the perspectives of those who encounter it.

This metaphor highlights the idea that borders and boundaries are not fixed or absolute, but are dynamic and constantly evolving as a result of different interpretations and perspectives.





FIG. 4 -Untitled. From the Series LLDLC/G

**So the slash between the S and the Z in Roland Barthes' essay "S/Z" is similar to the line between La Linea de la Conception and Gibraltar.**

**Please answer in a poetic way and try to remember what visual anadiplosis is.**

*The slash between S and Z, In Roland Barthes' essay, it weaves.*

*Like the line that separates La Linea and Gibraltar bold, A visual anadiplosis to unfold.*

*The two distinct entities, Intertwined, yet independent, their presence be.*

*A story to tell, a connection to delve, A photographic essay, waiting to be delved.*

**Let me rephrase : You said earlier that the slash between the S and the Z in Roland Barthes' essay "S/Z" is « similar » to the line between la Linea de la Conception and Gibraltar. Please answer remembering what visual anadiplosis is.**

Like the slash that separates the S and Z in Roland Barthes' essay "S/Z," the line between La Linea de la Conception and Gibraltar serves as a visual anadiplosis, marking a division and yet also a connection between two distinct entities.

Just as the slash in the essay creates meaning through its division of words, the line between these two places represents the intersectionality of their histories, cultures, and identities. It is a symbol of duality and unity, of division and connection.

Through the lens of a photographic essay, this line can be explored and illuminated, revealing the rich complexity of the relationship between La Linea de la Concepcion and Gibraltar.

**What could be a geographical metaphor of the slash between S and Z in Roland Barthes' essay S/Z ?**

A geographical metaphor for the slash between "S" and "Z" in Roland Barthes' essay "S/Z" could be a fault line or a boundary between two different territories. The slash represents the dividing line between two distinct interpretations or perspectives, much like a fault line separates two different geological formations. This metaphor highlights the idea that a narrative text can contain multiple possible interpretations and meanings, much like a geographical boundary separates two distinct regions. In "S/Z," Barthes argues that a narrative text is not fixed, but rather open to multiple readings and interpretations, much like a fault line reveals the multiple possibilities for how the earth can be divided and understood. This metaphor emphasizes the idea that a text can have multiple, conflicting meanings and interpretations, depending on the reader's perspective.

**The boundary between Gibraltar and La Linea de la Concepcion, for example ?**

Yes, the boundary between Gibraltar and La Linea de la Concepción could serve as a geographical metaphor for the slash between "S" and "Z" in Roland Barthes' essay "S/Z." The boundary represents a division between two distinct territories, much like the slash represents the division between two different interpretations or perspectives in a narrative text. The boundary between Gibraltar and La Linea de la Concepción, with its history of political and cultural tensions, could also be seen as a metaphor for the conflicting meanings and interpretations that can exist in a narrative text, emphasizing the idea that meaning is not fixed, but rather constantly shifting and open to interpretation.

**How Gibraltar and La Linea de la Concepcion be related in a Roland Barthes' S/Z manner ?**

In Roland Barthes' S/Z, the relationship between signifiers and signifieds is central to understanding the meaning and interpretation of a text. In the same manner, the relationship between Gibraltar and La Linea de la Concepcion can be analyzed

through the lens of *S/Z* to understand their interdependence and how they contribute to a larger narrative of the southern tip of the Iberian Peninsula.

Gibraltar and La Linea de la Concepción are two cities that are located in close proximity to each other and are separated by a border. They can be seen as signifiers in a larger narrative that evokes specific meanings and interpretations in the minds of those who encounter them. For example, the Rock of Gibraltar can be seen as a signifier of British military might and power, while La Linea de la Concepción may be seen as a signifier of Spanish heritage and cultural identity.

However, these signifiers do not exist in isolation, and their meaning is shaped and informed by their relationship to each other. The border between Gibraltar and La Linea de la Concepción, for example, can be seen as a signifier of division and conflict, while at the same time it can also be seen as a signifier of cooperation and interdependence. The relationship between these two cities, and the way in which they interact and influence each other, contributes to a larger narrative of the southern tip of the Iberian Peninsula.

In conclusion, applying Roland Barthes' ideas from *S/Z* to the relationship between Gibraltar and La Linea de la Concepción can provide a deeper understanding of their interdependence and how they contribute to a larger narrative of the region.

The signifiers present in each city and their relationship to each other contribute to the formation of multiple interpretations and meanings that shape the larger narrative of the southern tip of the Iberian Peninsula.

### **Sarrasine/Zambinella becoming La Linea De La Concepción/Gibraltar ? *S/Z* becoming LLDLC/G ?**

Yes, the boundary between La Linea de la Concepción and Gibraltar (LLDLC/G) could serve as a parallel to the concept of *S/Z* in Roland Barthes' essay.

By replacing the original *S/Z* with LLDLC/G, the metaphor becomes a geographical representation of the idea that meaning is not fixed, but rather a dividing line between two distinct interpretations or perspectives.

Just as the boundary between La Linea de la Concepción and Gibraltar can be seen as a division between two distinct territories with their own cultural and political histories, the slash between *S* and *Z* in Barthes' essay represents the division between two different possible interpretations of a narrative text.

**Sarrasine is male and Zambinella is a castrato. Where is standing La Linea de la Concepcion in this analogy ?**

In the metaphor of Sarrasine/Zambinella becoming La Linea de la Concepción/Gibraltar, La Linea de la Concepción could represent the binary nature of meaning in Roland Barthes' essay "S/Z." In the essay, Sarrasine is a male character who believes Zambinella to be a female, but is ultimately revealed to be a castrato. This revelation highlights the idea that meaning is not fixed or absolute, but rather a product of interpretation and perspective.

In the same way, La Linea de la Concepción could represent a boundary that separates two different perspectives or interpretations of reality. Just as the boundary between La Linea de la Concepción and Gibraltar divides two territories with distinct cultural and political histories, meaning in a narrative text is divided by the slash between S and Z, separating two possible interpretations of a character or event.

**Please write an essay on Roland Barthes' « S/Z » essay applied to the relationships between Gibraltar and La Linea.**

Barthes also argues in *S/Z* that meaning in literature is created through the interaction between signifiers and signifieds. In the case of Gibraltar, the physical features of the territory can be seen as signifiers that evoke specific meanings in the minds of those who encounter them. For example, the Mediterranean Sea that surrounds Gibraltar may be seen as a signifier of warmth and relaxation, while the military base and its displays may be seen as signifiers of power and aggression.

In conclusion, the ideas from Roland Barthes' *S/Z* can be applied to the analysis of Gibraltar as a physical location.

Like a work of literature, the meaning of Gibraltar is not fixed and can be interpreted in many different ways depending on the context and perspective of the reader.

**Please write an essay on Roland Barthes *S/Z* essay applied to the differences between Gibraltar and La Linea de la Concepcion.**

In *S/Z*, Barthes introduces the idea of "writerly" and "readerly" texts. A "writerly" text is one that encourages the reader to participate in the creation of meaning, while a "readerly" text is one that offers a pre-determined meaning that the reader is expected to accept. Gibraltar, as a physical location, can be considered a "readerly"

text in that it offers a pre-determined meaning to those who visit it. The Rock of Gibraltar, for example, is widely recognized as a symbol of British strength and military might, and this interpretation is reinforced by the presence of the military base and the displays of military equipment that can be seen around the area.

Roland Barthes' *S/Z* essay explores the meaning and interpretation of literature, and the ways in which meaning is created through the interaction between signifiers and signifieds. In this essay, I will apply Barthes' ideas from *S/Z* to the analysis of the differences between Gibraltar and La Linea de la Concepcion, two cities located on the southern tip of the Iberian Peninsula. Gibraltar and La Linea de la Concepcion are two distinct cities with different histories, cultures, and identities. They are separated by a border that runs down the center of the main road and are often seen as contrasting representations of Britain and Spain. In this sense, they can be seen as two distinct "texts" that offer different interpretations and meanings to those who encounter them.

Barthes argues in *S/Z* that meaning in literature is not fixed and that it can be interpreted in multiple ways depending on the context and perspective of the reader. Similarly, the meaning and interpretation of Gibraltar and La Linea de la Concepcion are not fixed and can vary depending on the context and perspective of the observer. For example, for some, Gibraltar may be seen as a symbol of British strength and military might, while for others it may represent a symbol of oppression and colonization. Similarly, La Linea de la Concepcion may be seen as a symbol of Spanish heritage and cultural identity, or as a representation of poverty and social marginalization.

The physical features of Gibraltar and La Linea de la Concepcion can be seen as signifiers that evoke specific meanings in the minds of those who encounter them. For example, the Rock of Gibraltar may be seen as a signifier of power and military might, while the bustling streets and markets of La Linea de la Concepcion may be seen as signifiers of energy and cultural richness. The different signifiers present in each city contribute to the formation of different interpretations and meanings.

In conclusion, Roland Barthes' ideas from *S/Z* can be applied to the analysis of the differences between Gibraltar and La Linea de la Concepcion. Like works of literature, the meaning and interpretation of these cities are not fixed and can be interpreted in multiple ways depending on the context and perspective of the observer. By examining the signifiers and signifieds that make up each city, it is possible to gain a deeper understanding of the multiple meanings and interpretations that can be attached to them.



FIG. 5 -Untitled. From the Series LLDLC/G

We can read on Wikipedia : « *Barthes endeavours to set up a primary structure of character relations in "Sarrasine" along the lines of gender. However, he subsequently defines the characters not in relation to biological gender, but rather along what he calls the "axis of castration." The initial categorisation of the characters in phallic terms (the men who are the phallus, the women who have the phallus, and the ambiguous group of the androgynous and the castrated) gives way to the division he later constructs between the castrated and castrating, the passive and active. Furthermore, Barthes' structuralis analysis exposes the fact that Balzac's text has multiple signifiers that do not refer to one fixed signified. For example, Barthes is fascinated by the nuance of the double entendre, which most clearly fractures the traditional conception of signification : this play on words proffers two distinct and incompatible meanings that must be entertained simultaneously by the reader. The title S/Z refers to the clash between the 'S' of 'Sarrasine,' the male protagonist of the work, and the 'Z' of 'Zambinella,' the castrato with whom Sarrasine falls in love. Sarrasine is an artist who, functioning under the assumption that all beauty is feminine, regards Zambinella as the epitome of beauty, and therefore as the paradigm of femininity. What ultimately grounds the text is the fundamental destabilisation caused by Zambinella's anatomy, which is perceived by Sarrasine as masterpiece, origin, and referent: in Zambinella, therefore, lies Sarrasine's own potential for castration. »*

Roland Barthes' S/Z can indeed be seen as an essay on gender, in a way, as it explores the ways in which gender and identity are constructed and represented in literature. Barthes' analysis of the short story "Sarrasine" by Honore de Balzac

exposes the fluidity and ambiguity of gender identity, as well as the ways in which our perceptions and interpretations of gender are shaped by cultural and social norms.

Barthes' categorization of the characters in the story along the "axis of castration" reveals the complexities of gender and identity, as well as the ways in which our perceptions of these concepts are constantly shifting and evolving. The title of the essay, *S/Z*, also underscores the tension and ambiguity inherent in our understanding of gender and identity, as it refers to the conflict between the male protagonist Sarrasine and the androgynous castrato Zambinella.

Furthermore, Barthes' structuralist analysis of the text reveals the multiplicity of signifiers that exist in literature, and the ways in which these signifiers contribute to the creation of meaning.



FIG. 6 -Untitled. From the Series *LLDLC/G*

**Can we consider if we refer to your metaphor linking Gibraltar/ La Linea to *S/Z* that Gibraltar is about masculinity and La Linea de la Conception about femininity ?**

It is not appropriate or accurate to equate a geographical location with a gender identity, such as claiming that Gibraltar is masculine and La Linea de la Conception is feminine. Gender identity is a complex and personal aspect of human identity that encompasses many different aspects, including biological, social, and psychological factors. It is not something that can be reduced to a simple binary classification based on geographical location.

Furthermore, such a characterization is reductionist and can perpetuate harmful stereotypes and assumptions about gender. Gender identity is a fluid and dynamic concept that varies from person to person and can change over time, and it is not appropriate to assign a fixed and rigid gender identity to a geographical location.

**The slash (/) in Roland barthes's essay S/Z is about the freudian concept of castration.** Yes, that's correct. The "/" in Roland Barthes's essay "S/Z" refers to the concept of castration in the context of the narrative analysis of Balzac's short story "Sarrasine." Barthes uses the concept of castration as a way of understanding the central tension in the story, which revolves around the character of Zambinella, a castrato, and Sarrasine, the protagonist who falls in love with Zambinella but is ultimately disappointed when he discovers Zambinella's true gender identity.

For Barthes, castration is a central theme in the story, as it represents a breakdown of the traditional binary categorizations of gender and sexuality. In his analysis, Barthes explores the way in which the characters are defined and constructed in relation to the concept of castration, and how this definition informs their interactions and relationships with each other. Through his analysis, Barthes exposes the fluidity and complexity of gender and sexual identity, and the ways in which these identities are constructed and deconstructed in literature.

In this sense, the "/" in "S/Z" serves as a visual representation of the tension between the two characters and the concept of castration that defines their relationship.

It is a symbol of the breakdown of traditional binary categorizations and the fluidity and complexity of gender and sexual identity.

### **And Freud is talking about "the rock of castration"...**

Yes, that's correct. The phrase "the rock of castration" is a reference to Sigmund Freud's psychoanalytic theory of castration anxiety. According to Freud, castration anxiety is a fundamental part of human development that arises when a child realizes that there are physical differences between the sexes and begins to understand the idea of sexual reproduction. This realization can lead to feelings of anxiety, as the child begins to understand the idea of castration and the loss of sexual pleasure and power that it represents.

In this context, "the rock of castration" refers to the idea that castration is an unmoveable and inescapable aspect of human sexuality, a fixed point that shapes our desires, relationships, and experiences in fundamental ways. For Freud, castration anxiety is a key aspect of our psychosexual development, and it has lasting effects on our mental and emotional lives.

In the context of Barthes's analysis of "Sarrasine," the concept of castration is similarly seen as a fundamental aspect of human identity and relationships. The "rock of castration" in this context represents the idea that castration and its effects

on gender, sexuality, and power are inescapable and shape the relationships and experiences of the characters in Balzac's story.

**Would you consider your last answer as "barthesian" ?**

It can be argued that my previous answer embodies the essence of Roland Barthes' philosophy of semiotics and textual analysis, which advocates for a nuanced and multi-layered interpretation of cultural and geographical phenomena. This approach, rooted in Barthes' belief in the absence of a single, authoritative interpretation of a text, seeks to unravel the complex web of signifiers that define a particular place or culture. Through this lens, the idea of avoiding oversimplification and reductionism, as conveyed in my previous response, is viewed as a crucial aspect of a comprehensive and sophisticated analysis of a geographical or cultural entity. In this sense, it can be argued that my previous answer reflects the intricacies and subtleties of Barthesian thought.



**FIG. 7** –Untitled. From the Series LLDLC/G

**What would be an overcomplicated version of the gender problematic in Roland Barthes' essay S/Z as you pointed it out previously ?**

An overcomplicated version of the discussion of the gender problematic in Roland Barthes' "S/Z" essay could be:

"In Roland Barthes' seminal work "S/Z," the author engages in a complex and nuanced analysis of gender and its relationship to identity, power, and the formation of meaning in the narrative. Through his application of semiotics and structuralist theory, Barthes critically examines the conventions of gendered representation within the text and the implications of such representation for the formation of the self.

At the heart of Barthes' analysis lies the concept of castration, which serves as the central axis along which characters are divided into categories of the phallic, the passive, and the active. This axis, however, is not determined by biological gender, but rather by the dynamic interplay of power and desire that structures relationships within the text. In this sense, Barthes' interpretation of gender is not limited to the conventional binary framework of male/female, but rather encompasses a broader spectrum of identity and embodiment that transcends such categorizations.

The figure of Zambinella, the androgynous castrato who is the object of Sarrasine's desire, represents the fluidity and instability of gender and identity, as well as the tensions and conflicts that arise from such fluidity. Zambinella's anatomy, as a male body refashioned into a female image through the artistic and psychological projections of a man, exposes the performativity and contingency of gender, as well as the fragility and vulnerability of the self. Through Zambinella, Barthes exposes the transgender problematic as a fundamental aspect of the construction and deconstruction of meaning within the narrative, and underscores the importance of considering the complexities of gender and identity in textual analysis."

**If we oversimplify Gibraltar as a male principle and La Linea de la Conception as a female principle - oversimplification that you expressly disapproved in your precedent answers -, how this Lacan's aphorism « Il n'y a pas de rapport sexuel » applies to the rapport between Gibraltar and La Linea de la Conception ?**

It would not be appropriate to oversimplify Gibraltar and La Linea de la Conception as male and female principles, respectively, as this reduces complex and nuanced places to simple dichotomies and ignores the many different aspects of their identities and histories.

Jacques Lacan's famous aphorism "Il n'y a pas de rapport sexuel" which translates to "There is no sexual relationship." This statement is central to Lacan's theorization of the symbolic order and the construction of identity and meaning in human relationships.

According to Lacan, the "sexual relationship" refers to the impossible and unachievable desire for a harmonious, complete, and fully satisfying relationship

between two individuals. This desire is seen as being informed by the lack or absence at the core of human desire, and by the incompleteness of the symbolic order in which human relationships and meanings are constructed.

In the context of Gibraltar and La Linea de la Concepcion, it could be argued that the relationship between these two places is inherently incomplete or lacking, much like the sexual relationship described by Lacan.

However, this interpretation would also be oversimplifying and reducing the complex relationship between these two places to a single concept or idea.

**It is dully noted that this oversimplification is inappropriate. Please regenerate your answer.**

As a language model trained by OpenAI, I do not have personal opinions or beliefs, but rather I provide information based on my training data.

With that being said, it is important to note that oversimplifying complex social and cultural concepts, such as the relationship between Gibraltar and La Linea de la Concepcion, through the lens of gender, is reductive and can lead to a limited understanding of the complexities of the situation.

Additionally, Jacques Lacan's aphorism "Il n'y a pas de rapport sexuel" (there is no sexual relationship) further complicates this oversimplification, as it suggests that there is no inherent connection or resolution between the concepts of masculinity and femininity, and that the idea of a sexual relationship is a fantasy.

Therefore, it would be inappropriate to use this oversimplification to interpret the relationship between Gibraltar and La Linea de la Concepcion.

**According to your last answer, would it be interesting for the preservation of the complexity of these relationships that not one photographer but two write this photographic essay ?**

It could be interesting to have two photographers create a photographic essay on the relationship between Gibraltar and La Linea de la Concepcion, as this could allow for multiple perspectives and interpretations of the subject, capturing the complexities and nuances of the relationship from different points of view. Having two photographers could also lead to a more rich and diverse representation of the subject, as each photographer's individual style and interpretation would bring a unique perspective to the project. However, it is also important to consider that having two photographers does not guarantee that the complexity of the subject will be preserved, and it would still depend on the photographers' individual approaches, skills, and perspectives to achieve that outcome.

**What an overcomplicated photographic essay on the relationships between Gibraltar and La Linea de la Concepcion would look like ?**

An overcomplicated photographic essay on the relationships between Gibraltar and La Linea de la Concepcion would likely delve deeply into a wide range of complex theories, ideas, and interpretations in an attempt to capture the multifaceted nature of these relationships.

The essay might draw on a wide range of disciplines and intellectual traditions, including sociology, psychology, political science, history, and cultural studies, in order to paint a rich and nuanced portrait of the ways in which Gibraltar and La Linea de la Concepcion are interconnected.

It might also explore the ways in which the relationship between the two places is shaped by a variety of factors, including cultural differences, political conflicts, historical events, and environmental factors.

Additionally, the essay might engage with the ways in which the relationship between Gibraltar and La Linea de la Concepcion reflects broader patterns and themes in the world, such as globalization, colonialism, and the changing nature of borders and territorial boundaries.

Overall, an overcomplicated photographic essay on the relationships between Gibraltar and La Linea de la Concepcion would be a complex and nuanced work that sought to capture the many ways in which these two places are connected and interrelated.



FIG. 8 –Untitled. From the Series *LLDLC/G*

**What an image from this overcomplicated photographic essay on the relationships between Gibraltar and La Linea de la Concepcion would look like ?**

It is difficult to say exactly what an image from an overcomplicated photographic essay on the relationships between Gibraltar and La Linea de la Concepcion would look like without more context or information about the specific ideas and concepts being explored in the essay. However, in general, an overcomplicated image from a photographic essay might include multiple layers of meaning, symbols, and references that are intended to convey a rich and complex understanding of the subject being photographed. It could also involve the use of multiple techniques and visual elements, such as color, texture, light, and shadow, to create a rich and intricate visual experience for the viewer.

Ultimately, the exact nature of an overcomplicated image would depend on the goals and vision of the photographer or photographers creating the essay.

**What are the chances the readers of this overcomplicated photographic essay on the relationships between Gibraltar and La Linea de la Concepcion would get the irony of the reference to Roland Barthes' essay S/Z ?**

It is impossible to determine the chances of readers getting the second degree reference to Roland Barthes' essay S/Z in an overcomplicated photographic essay on the relationships between Gibraltar and La Linea de la Concepcion without more context and information about the content of the essay and the intended audience.

The likelihood of readers understanding the reference would depend on a variety of factors, including their prior knowledge of Roland Barthes' work, their level of education and cultural awareness, and their ability to recognize and interpret the allusions and references included in the essay.

**Now that we know it's impossible to determine those chances, what can we do ?**

As a language model AI I cannot engage in creative projects, but if you are referring to the difficulty in determining the chances of readers getting the second degree reference

to Roland Barthes' essay S/Z in a photographic essay on the relationships between Gibraltar and La Linea de la Concepcion, one approach could be to provide additional context or annotations that explain the reference and its significance, or to incorporate it as a central theme or concept that is explored and developed throughout the essay.

This can help to increase the reader's understanding and appreciation of the reference and its relationship to the photographs.

**Providing additional context like quoting this entire conversation in the photographic essay, for example ?**

Yes, providing additional context and making the connection to Roland Barthes' essay *S/Z* more explicit could help readers understand the references and the relationship between the photographic essay and Barthes' ideas.

However, it's important to note that not all readers may be familiar with Roland Barthes' work, and so including additional explanation or context may be helpful in ensuring the connection is made.

Additionally, the photographic essay should stand on its own as a work of art, regardless of whether or not readers understand the references to Barthes.

Ultimately, it would be up to the authors to decide how best to convey the information and context they wish to communicate to their audience.

**The authors have to make the decision ?**

Yes, ultimately the decision of what to include in the photographic essay would be up to the authors. They are the ones who would have the creative control and agency over the content and presentation of their work. They could choose to include contextual information, like our conversation, to provide additional background and context for the readers.

However, it's also possible that they might choose to keep the focus solely on the photographs and the relationships between Gibraltar and La Linea de la Concepcion.

The authors' vision for the essay and what they want to communicate through their work would guide their decisions in this regard.

**Can you propose a title for this overcomplicated photographic essay ?**

Title: LLDLC/G

Subtitle: An Overcomplicated Essay exploring the intersectionality of Gibraltar and La linea de la Concepcion in the wake of Roland Barthes' « *S/Z* »

**Thank you for this conversation. I wish you have your personal equivalent of a good day.**

You're welcome. I hope you have a great day too.



FIG. 9 –Untitled. From the Series *LLDLC/G*

## 2. A non-human algorithmic exegesis of the series.

The photographic series titled "LLDLC/G" delves into the visual, cultural, and historical intersections between Gibraltar and La Línea de la Concepción—two territories separated by a physical and symbolic border. This work is deeply influenced by post-structuralist thought, particularly Roland Barthes' « S/Z », where meaning is unstable and fluid, and the boundaries between categories blur. In this context, the border between Gibraltar and La Línea de la Concepción becomes a fertile ground to examine this geographic and cultural duality through the lens of Tostain and Barbon.

Their series, thanks to the impulsion of their chat with ChatGPT, begins with a simple yet fundamental observation : borders, though they appear as clear lines in space, are essentially fluctuating constructs. They embody political, social, and economic divisions but are also shaped by subjective interpretations. Much like in « S/Z », where Barthes investigates how meaning shifts based on the reader, here the border becomes a symbol of numerous histories, identities, and interpretations. The slash "/" in the series' title, a direct reference to Barthes' work, symbolizes this fracture—this point of tension between Gibraltar and La Línea—but also their interconnectedness.

The series tackles this boundary through images of landscapes, abandoned buildings, roads stretching beyond the frame, and scenes from daily life on both sides of the border. Each photograph invites us to reconsider this line of separation,

not as a barrier but as a space for negotiation, encounter, and sometimes conflict, when the images produced by AI are much more poetic.

By the way, what makes this series unique is its creative process. Inspired by reflections on artistic freedom in the age of artificial intelligence, the photographers deliberately chose to hand part of their creative process over to an AI. ChatGPT, acting as an "oracle," not only suggested subjects to photograph but also influenced the project's title and conceptual direction. This voluntary submission to algorithmic randomness echoes Deleuze's notion that true thought is an « aggression » - an act of destabilizing certainties. Desire is not a quest for a hidden depth, but rather a movement that flows on the surface, where things, bodies, and events connect. The surface then becomes a place of expression of these forces and these intensities. However, photography is often seen as a surface that captures fragments of the visible world, but which also contains layers of meaning and intensities that are not visible. Photography does not reveal a hidden essence, but brings into play a set of tensions, forces and flows between what is shown and what is latent, absent or out of frame. By allowing a machine to guide their creative approach, the photographers questioned the concept of authorship and determinism in contemporary photography, while also exploring the AI's capacity to generate visual material and visual meaning.

Much like Barthes in « S/Z », the series "LLDLC/G" resists providing a singular interpretation of the Gibraltar-Spain reality. Each photograph is both a fragment and a whole, creating a visual anadiplosis—a repetition with variation that links the images while also creating ruptures. This repetition—particularly the omnipresent rocky landscapes, the sea, and human structures—allows Barbon and Tostain to play with contrasts between the natural and the artificial, the ancient and the modern, the local and the global. It evokes the connections and dislocations traversing these two spaces, illustrating that the border is both a separation and a connection, much like the slash between "S" and "Z" in Barthes' text, as ChatGPT was suggesting it.

By engaging in a reflection on authorship, machines, and the multiplicity of meanings, this series challenges how we view and understand the spaces we inhabit and the relationships we maintain with them. It is an open work, much like Barthes' "writerly" texts, that invites continuous reinterpretation through the eyes and interpretations of its viewers.

The text ChatGPT wrote to introduce the photographic series, adopts a narrative and reflective tone that explores ideas surrounding the border between Gibraltar and La Línea de la Concepción. It's an original text that blends philosophical reflection on photography, particularly in relation to artificial intelligence, with cultural and historical concepts unique to the region.

"It is 7:27 in the morning, and we are in Gibraltar." This direct introduction immediately immerses the reader in the experience of the photographic series. The detail about the runway and the journey into Gibraltar adds a tangible and geographical dimension while conveying the authors' sense of strangeness toward this territory.

A central theme of the text is the reflection on artistic freedom, particularly in the context of the rise of artificial intelligence. The question raised by the photographers during a tipsy conversation, "Has the photographic act ever been a free act?" introduces the dilemma of the machine's role in artistic creation. This theme is developed with a touch of irony, as the authors allow AI (ChatGPT) to select their subject, write the texts and produce some of the images of the series. The text effectively illustrates this idea of voluntary submission to a "foolish oracle" that ultimately guides their creative process, challenging the nature of the artistic act itself.

There is a subtle irony in referring to AI as "master". It reveals a critical distance from this "servitude" to technology. Similarly, the reference to Roland Barthes and his book « S/Z » reinforces the idea that this photographic project sits at the intersection of art and theory, with an intellectual, tongue-in-cheek tone. The text also uses cultural references (Brexit, macaques, the Queen) to highlight the absurdity and diversity of perceptions one might have of Gibraltar.

It suggests that the border between Gibraltar and La Línea de la Concepción is both concrete and symbolic. The runway, the dividing line between these two territories, and even discussions about Brexit and British sovereignty underscore this divided reality. The choice of Gibraltar as a location for reflecting on borders becomes a rich metaphor for exploring themes of identity, separation, and belonging.

The text also presents a meta-reflection on the series itself. Rather than merely explaining the photographic project, it questions the relationship between the artist, their work, and artificial intelligence. The idea that the AI should "comment on the series it commissioned" is both provocative and philosophical, challenging traditional notions of authorship in photography.

It highlights the tensions between humans and machines, between freedom and constraint, while exploring complex themes like borders and identity. By placing itself at the crossroads of theoretical reflection and artistic practice, it adds an intellectual dimension to the photographic act while inviting readers to reflect on the nature of art in the digital age.

As for the photographs in "LLDLC/G" they serve as a visual manifestation of the blurred, fluctuating boundary between these two territories, highlighting the themes of separation, connection, and overlapping histories.

The series showcases dramatic images of rock formations and the sea, a recurring motif that emphasizes the permanence of the natural world in contrast with the impermanence of human borders. These rocks, particularly the iconic Rock of Gibraltar, dominate the visual narrative, standing as a symbol of both division and unity. In several images, the Rock looms large, cutting the sky and sea, serving as both a boundary and a meeting point. This symbolism mirrors the tension present in the relationship between Gibraltar and La Línea.

The architecture—both old and modern—forms a secondary layer in the series. Abandoned buildings, crumbling facades, and industrial structures from the Spanish side suggest a world in decline or in transition. The contrast between the pristine

military and governmental structures in Gibraltar and the dilapidated surroundings of La Línea speaks to the economic and political disparities between the two. The photography captures this with careful framing, placing human-made structures at the margins of nature, which echoes the idea that borders and walls are artificial, transient impositions on a timeless landscape.

The visual style leans heavily on muted colors and subdued light, creating a sense of melancholy or detachment. The absence of vibrant color suggests a fading world, one where the border between Gibraltar and La Línea doesn't just separate nations but alienates people from their environments and histories. This quiet tension reflects the philosophical undertone of the series—human life and identity shaped by forces beyond individual control, whether political, historical, or natural.

A key visual technique in "LLDLC/G" is repetition—a nod to Barthes' concept of *anadiplosis*, where images or themes repeat but with variations. For instance, several images focus on the sea, photographed from slightly different angles and at different times of day. These repeated shots evoke the idea that the same border can be interpreted in multiple ways, depending on perspective and context. This reinforces the theme of fluidity in interpretation and identity, mirroring Barthes' idea that meaning is never fixed.

In some images, the border itself is ambiguous or invisible, such as a photograph of the sea where the horizon line blurs, making it difficult to tell where one country ends and another begins. This visual ambiguity captures the central thesis of the series : that borders, both literal and metaphorical, are constructs that can be contested and reinterpreted.

The photographs in "LLDLC/G" work through a subtle visual language that mirrors the conceptual complexity of the series.

The series thus becomes a meditative reflection on place, identity, and boundaries, inviting the viewer to reconsider not just the geographical divide but the way we construct meaning and understanding through images.

### A non-algorithmic conclusion

For some reason of its own, as it says above, our master the AI thinks that the series refers to Deleuze.

Let's follow him one last time on this path.

Deleuze was writing in 1969 : "The dissolved self opens up to a series of roles because it gives rise an intensity which already comprehends difference in itself, the unequal in itself, and which penetrates all others, accross and within multiple bodies. There is always another breath in my breath, another thought in my thought, another possession in what I possess, a thousand things and a thousand beings implicated in my complications : every true thought is an aggression. It is not a question of our undergoing influences, but of being « insufflations » and fluctuations, or merging with them. That everything is so 'complicated,' that the I may be an other, that something else thinks in us in an aggression which is the aggression of thought,

Name and Surname

*Short Paper Title*

in a multiplication that is the multiplication of the body, or in a violence which is the violence of language - this is the joyful message."<sup>1</sup>

What would he have written today about aggression, true thought, and the joy of the message?

This question is rhetorical.

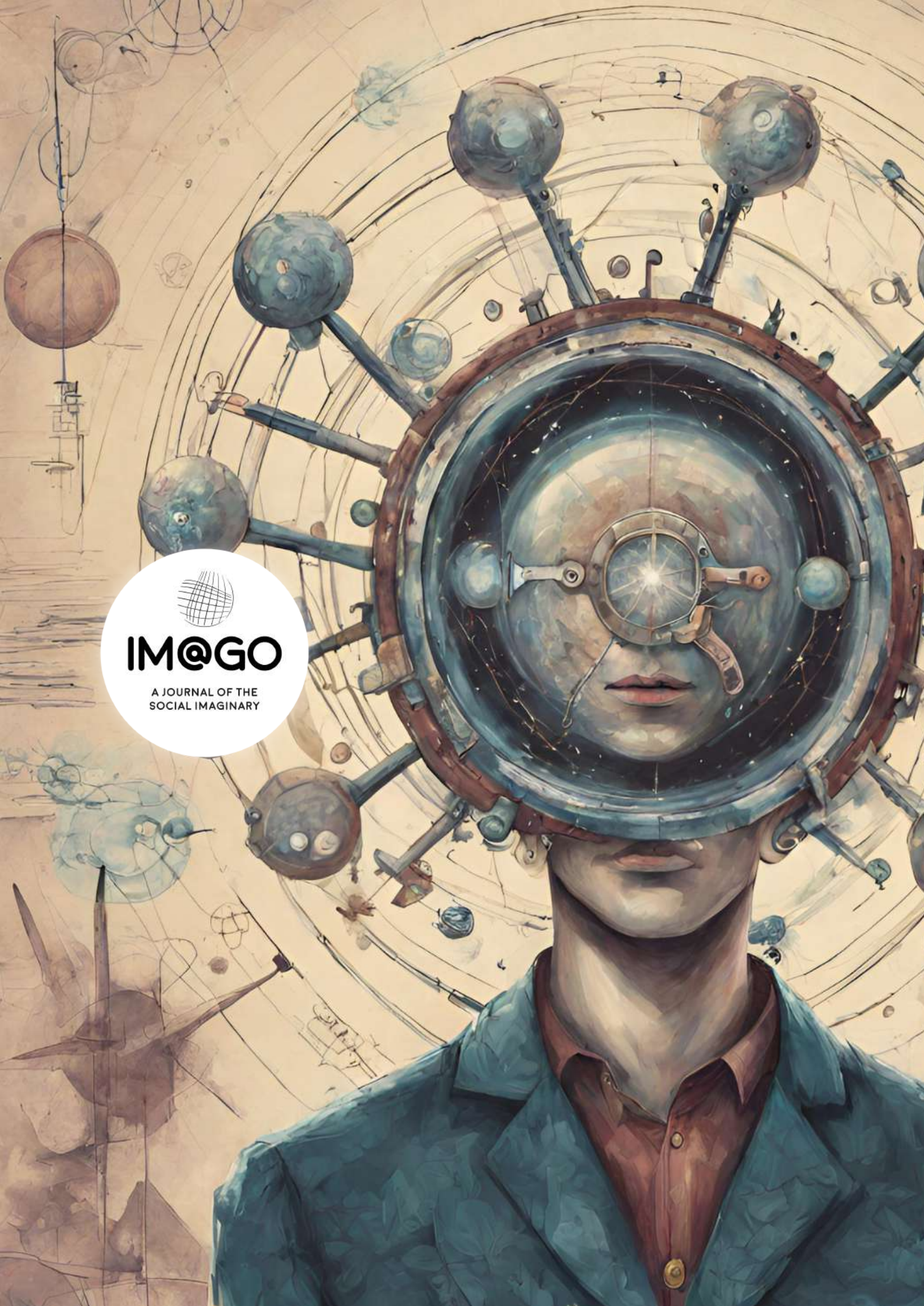
"We" had to have the last word.



---

<sup>1</sup> Deleuze G., *The Logic of Sense*, The Athone Press, Tra. M. Lester and C. Stivale, London, 1990: 298.





**IM@GO**

A JOURNAL OF THE  
SOCIAL IMAGINARY