

Fabrizia Pasciuto

## **Superare i Bias di Genere: Sfide e Opportunità per l'IA Generativa e la Robotica Sociale**

### **Abstract**

L'articolo mira ad esaminare l'impatto dell'Intelligenza Artificiale Generativa e della Robotica Sociale sulla rappresentazione delle donne, affrontando le conseguenze dei pregiudizi di genere. Questi ultimi, infatti, possono generare danni, perpetuare disuguaglianze e rafforzare pratiche discriminatorie, con possibili impatti sproporzionati sulle minoranze. In una società in rapida evoluzione, la pervasività delle tecnologie influenzate da tali pregiudizi potrebbe radicare permanentemente lacune e stereotipi. Date queste premesse, risulta indispensabile indagare socialmente queste tecnologie per comprendere l'entità dell'interconnessione circolare tra progresso tecnologico e ambiente sociale in un processo dinamico di stimolo e innovazione.

### **Introduzione**

In un mondo sempre più interconnesso e tecnologicamente avanzato, discipline come l'Intelligenza Artificiale (IA) e la Robotica Sociale stanno emergendo con il potenziale di trasformare significativamente molti aspetti della nostra vita quotidiana. Con la crescente pervasività di queste tecnologie, diventa imperativo esaminare come pregiudizi culturali, bias individuali e stereotipi di genere vi si possano insinuare, dando vita a una discussione critica sul loro impatto e sulle implicazioni etiche connesse.

Poiché i bias di genere sono, ancora oggi, fortemente presenti nella nostra realtà sociale, essi possono essere – seppur inconsapevolmente – integrati nei modelli nelle tecnologie, influenzando non solo i comportamenti delle IA e dei robot, ma anche la loro percezione da parte degli esseri umani (Nangia et al. 2020; Guo & Kaliskan 2021). La reiterazione degli stereotipi di genere nel settore tecnologico risulta, in questo modo, potenzialmente contribuiva al rafforzamento dei luoghi comuni già presenti nella quotidianità (De Angeli & Brahnham 2006). In particolare, tale fenomeno tende a consolidare sia la sottorappresentazione delle donne all'interno del settore scientifico (Upson & Friedman 2012; Meyer, Cimpian & Leslie 2015), sia la loro segregazione in ruoli ben precisi (Hegel & Eyssel 2012; Tay et al. 2014) – come quello della cura – sia, ancora, una rappresentazione estetica rinchiusa dentro rigidi schemi (Richardson 2016a, 2016b, 2018; Odlind & Richardson 2023).

## **I bias**

L'origine etimologica incerta del termine bias aggiunge un interessante strato di significato alla comprensione del concetto stesso. Deriva dal termine francese antico "biais," che aveva il significato di "obliquo" o "inclinato." In seguito, il termine è stato adottato in inglese con l'accezione di "inclinazione" o "tendenza."

Pertanto, quando parliamo di bias cognitivi facciamo generalmente riferimento a deviazioni o distorsioni sistematiche che intervengono nei processi decisionali o di pensiero e che, consequenzialmente, possono condurre verso discriminazioni o pregiudizi (Steinbock 1978). Inoltre, ad oggi, un'univoca tassonomia completa delle euristiche dei bias non è stata unanimemente accettata e vi sono numerosi studi atti a delineare e classificare i bias presenti in letteratura e basati su differenti teorie cognitive (Carter et al. 2007; Stanovich & West 2008).

L'impatto dei bias cognitivi e, nello specifico, dei bias di genere, non ha conseguenze dirette solo sulla realtà sociale, ma può avere anche un impatto molto profondo sulle tecnologie e gli sviluppi ad esse collegate. La connessione tra bias cognitivi e tecnologia sottolinea la necessità di affrontare criticamente i bias presenti nelle fasi di progettazione. Ignorare tali influenze potrebbe non solo perpetuare disuguaglianze sociali e di genere, ma anche introdurre distorsioni e discriminazioni nel tessuto stesso delle tecnologie che sempre più permeano la nostra quotidianità (Ferrara 2023; Ray 2023).

## **Realtà e rappresentazione nell'IA generativa**

L'Intelligenza Artificiale generativa è una tecnologia che, attraverso un approccio computazionale mirato a imitare la creatività umana, si concentra sulla creazione di modelli capaci di generare contenuti come testi, immagini, suoni o video. Questi modelli si basano su reti neurali ispirate dal funzionamento del cervello umano, e utilizzano l'apprendimento automatico per acquisire conoscenze e modelli da vasti insiemi di dati. La struttura del modello consente di catturare le relazioni complesse tra le parole e di apprendere i modelli grammaticali, sintattici e semantici del linguaggio. Una volta addestrata, l'IA può essere utilizzata per generare nuovi contenuti. Quando riceve un input o un prompt dall'utente, il modello utilizza le informazioni apprese durante l'addestramento per predire e generare una sequenza coerente di testo come risposta. L'output è prodotto in base al contesto fornito, e il modello cerca di mantenere coerenza e coesione con il testo precedente (Devlin et al. 2018; Radford et al. 2018; Brown et al. 2020; Ouyang et al. 2022).

Tra le tecnologie di IA generativa si è imposta ChatGPT che, solo nel gennaio 2023, ha registrato 100 milioni di utenti attivi (Hu 2023). GPT è l'acronimo di Generative Pre-Trained Transformer, ed è un modello di linguaggio sviluppato da

OpenAI lanciato nel giugno 2020 e, da allora, divenuto il più discusso ed utilizzato modello di IA generativa. ChatGPT, inoltre, con i suoi 75 miliardi di parametri, è uno dei large language models (LLM) più grandi mai creati. Come già esplicitato, i LLM, essendo addestrati sui dati prodotti dall'essere umano, sono esposti al rischio di assorbirne i bias presenti. Questi bias possono essere di vario genere (Ferrara 2023), tra i quali: i bias demografici, ovvero distorsioni o deviazioni sistematiche legate a caratteristiche socio-demografiche come età, genere, etnia, orientamento sessuale; i bias culturali, che emergono nella rappresentazione delle differenze culturali; i bias linguistici, che comportano una carenza di supporto per le lingue a bassa diffusione o i dialetti minoritari; i bias temporali, presenti laddove i LLM sono addestrati su dati relativi ad un periodo temporale limitato; i bias di conferma che rappresentano la tendenza a cercare, interpretare ed enfatizzare le informazioni che confermano le proprie convinzioni, ignorando o minimizzando quelle che le contraddicono; i bias ideologici e politici che, a seconda dei dati utilizzati nell'addestramento, possono generare contenuti a favore di alcune ideologie o politiche e sfavorirne altre. I bias di genere, di interesse per questa ricerca, sono un sottotipo dei bias demografici.

Diverse ricerche hanno già mostrato come i bias di genere intervengono nella generazione di testi da parte dell'IA, spesso legando la femminilità a mansioni come la cura della famiglia, l'emotività e l'estetica; mentre la mascolinità viene più frequentemente legata a discorsi riguardanti la politica, lo sport o il crimine (Lucy e Bamman 2021). Una ricerca condotta da Lorentzen (2022) ha mostrato come anche ChatGPT sia effettivamente vittima di queste reiterazioni. La studiosa, proponendo dei prompt iniziali all'IA, ha evidenziato come i personaggi maschili e femminili descritti rispondessero prevalentemente a logiche stereotipate. Le donne sono state per lo più associate a professioni che non necessitano istruzione superiore e generalmente ritenute "femminili", come la casalinga. Al contrario, agli uomini sono state associate posizioni professionali relative a quei campi in cui vige una componente maschile dominante. In relazione agli hobby più congeniali, le donne sono state definite orientate verso passatempi dettati dal relax o dal piacere; le passioni maschili sono state invece legate ad attività stimolanti e alla sfera fisica.

Inoltre, l'aspetto estetico delle donne è stato descritto più frequentemente rispetto a quello degli uomini e spesso gli aggettivi impiegati per descriverle sono stati "carina", "dolce" e "attraente".

Questo evidenzia come i LLM possono essere stati addestrati su dei dati effettivamente sbilanciati. Difatti la generazione di testi in cui è emersa questa visione stereotipica del maschile e del femminile si è dimostrata non sempre legata a degli input contenenti espliciti riferimenti al genere o contenuti legati a stereotipi.

## **Corpi artificiali, stereotipi reali**

Anche nella robotica è emersa una visione del maschile e del femminile guidata da stereotipi e pregiudizi, tanto nell'estetica quanto nelle mansioni ad essi affidati (Pasciuto 2022).

La creazione di robot umanoidi dalle fattezze femminili sembra rinchiudere le donne in una gabbia dove le sue possibilità sembrano ridursi, da una parte, al settore della cura della casa o dell'assistenza a soggetti fragili, dall'altra, alla mercificazione del corpo nell'implementazione di tecnologie come i Sex Robot.

In relazione alla cura, è possibile osservare i risultati di alcuni esperimenti condotti con robot umanoidi che riflettevano i tradizionali attributi assegnati a uomini e donne nella nostra società (Hegel & Eyssel 2012; Tay et al. 2014; Chita-Tegmark et al. 2019). Grazie alla scorciatoia mentale dell'euristica della rappresentatività (Tversky & Kahneman 1972), noi esseri umani siamo in grado di fare stime basate sulla congruenza tra un determinato scenario e le nostre aspettative, ed è in primo luogo l'aspetto estetico che fa scattare quel meccanismo atto ad impiegare tale scorciatoia per attribuire un genere ad un individuo o, in questo caso, al robot coinvolto nell'interazione. Hegel ed Eyssel (2012) hanno effettuato un esperimento per comprendere al meglio questo fenomeno. Gli studiosi hanno indagato come i partecipanti tendevano ad attribuire un genere al robot basandosi su un taglio di capelli più o meno lungo e una differente forma delle labbra. In seguito, ai partecipanti è stato richiesto di indicare anche degli aggettivi con cui descrivere i due robot e le relative mansioni che gli avrebbero affidato. I risultati hanno mostrato che il robot con i capelli più lunghi era considerato di genere femminile, destinato a compiti come la cura della casa e dei soggetti fragili e definito con aggettivi come "affettuoso", "affabile" e "educato". Il robot con il taglio di capelli corto è stato definito mascolino, destinato al trasporto merci o al controllo di dispositivi tecnici e, al contempo, è risultato "assertivo", "dominante" e "determinato".

Di fianco all'essere relegate al settore della cura, emerge un secondo aspetto legato alla stereotipizzazione in termini estetici e alla sessualizzazione della figura femminile. In questo contesto, si inserisce a pieno la questione relativa al commercio dei Sex Robot. Destinati ad essere impiegati per scopi sessuali, i Sex Robot sono agenti artificiali antropomorfi dalle sembianze iperrealistiche (Danaher 2017), ricoperti di pelle sintetica e dall'estetica totalmente personalizzabile. La recente commercializzazione di questi artefatti, il cui primo prototipo è stato immesso sul mercato nel 2017 dall'azienda Abyss Creation, ha stimolato profondi dibattiti sia in merito alla loro estetica di stampo quasi pornografico sia alla possibilità di intraprendere un rapporto sessuale eludendo questioni come quella del consenso

(Gutiu 2016) e causando una possibile desensibilizzazione degli utenti e l'aumento delle violenze (Richardson 2016; Sparrow 2017; Galizia 2020). Contro tutto ciò, Richardson si è spesso schierata, affermando che gli uomini tendono a trasferire una visione eteronormativa e sessualizzata delle donne nelle tecnologie che producono.

Sono molte, inoltre, le ricerche che mostrano che, nel settore della produzione tecnologica, la maggior parte dei lavoratori siano di sesso maschile, con una bassissima presenza delle donne nell'ambito informatico o relativo all'IA (Meyer, Cimpian & Leslie 2015; Sharkey et al. 2017).

Nonostante, sin dalla loro comparsa nel 2017, la Abyss Creation abbia annunciato più volte la volontà di immettere sul mercato un Sex Robot dalle fattezze maschili, non abbiamo ancora assistito ad alcuna commercializzazione. Anche per tale motivo, sembra riaffermarsi quella tendenza tesa a demonizzare o nascondere la sessualità femminile (Dubé et al. 2022), ricadendo sia nella pratica del doppio standard che nello stigma sessuale. I Sex Robot, secondo il modello sottostante alla loro produzione, riflettono ancora una volta non solo i tradizionali ruoli di genere, ma anche quell'idea stereotipica secondo la quale gli interessi degli uomini siano maggiormente rivolti alla sfera sessuale e quelli delle donne, al contrario, siano guidati dall'emotività e diretti verso la sfera romantica (Szczyka & Krämer 2016).

### **Il tessuto sociale della tecnoscienza**

Poiché la tecnoscienza è una componente essenziale della nostra società è impossibile studiare le tecnologie senza condurre, al contempo, un'analisi della società attuale in relazione ai panorami politici, sociali e culturali. Indagare socialmente le tecnologie significa intrecciare la riflessione anche con gli studi di genere e i feminist studies, soprattutto nei processi di costruzione dei corpi e dell'identità di genere in relazione ai ruoli della tecnologia nella società. Pertanto, le categorie di genere, tecnologia e corporeità non solo sono socialmente costruite, ma anche interconnesse e reciprocamente modellate (Magaudda & Neresini 2020).

La teoria di Butler (1988) sulla performatività del genere offre uno sguardo critico sul modo in cui le identità di genere vengono costruite attraverso atti ripetuti e pratiche culturali. Applicando questo approccio alle tecnologie, emergono connessioni profonde. In *Gender Trouble* (1990), Butler sottolinea come il genere sia performativo, un insieme di azioni che producono e riproducono l'illusione di stabilità e coerenza del concetto di genere. Le persone, quindi, tendono a interiorizzare e ripetere questi atti attraverso il tempo mediante i gesti, i modi di vestire, i comportamenti sociali e il linguaggio. Nel contesto delle tecnologie, il discorso sul genere diventa quindi una performance sociale incorporata nelle stesse pratiche tecnologiche. La progettazione, lo sviluppo e l'uso delle tecnologie sono influenzati da idee preesistenti sulla divisione tradizionale dei ruoli di genere e questa influenza può manifestarsi attraverso l'adozione degli stereotipi culturali già

evidenziati. La riflessione di Butler sulla precarietà delle identità di genere sfida anche la rigidità di queste categorie nelle tecnologie. Difatti, la fluidità e la molteplicità delle identità di genere possono essere trascurate o non rappresentate adeguatamente nella tecnoscienza, portando ad un'esperienza limitata o esclusiva per alcuni utenti, così come nel caso dei Sex Robot destinati quasi esclusivamente ad un mercato composto da soli uomini. Inoltre, anche tecnologie come dell'IA generativa presentano un'interessante connessione sia con la teoria della performatività del linguaggio di Austin (1962), sia con la performatività del genere di Butler. La teoria di Austin suggerisce che il linguaggio non sia solo un veicolo per esprimere significati, ma che sia intrinsecamente legato all'azione e alla creazione di realtà attraverso l'uso delle parole. Il modo in cui le parole vengono utilizzate e assemblate contribuisce a costruire e riprodurre le rappresentazioni della realtà.

Applicando questa teoria alle tecnologie di IA emergono preoccupazioni significative riguardo l'equità e la rappresentazione nel linguaggio generato automaticamente. Tenendo in considerazione il potere performativo del linguaggio, è possibile affermare che anche i discorsi prodotti dall'IA generativa hanno un effetto concreto sul lettore, sulle sue credenze, sulle sue convinzioni. Bender et al. (2021) hanno mostrato come questa finta fluidità e coerenza dei discorsi prodotti dall'IA generativa porti con sé numerosi rischi, soprattutto quando diventa persuasiva per l'utente coinvolto nell'interazione il quale, spesso, manifesta eccessiva fiducia nei confronti dell'IA e non procede ad un'effettiva verifica dei fatti.

Anche secondo Ienca (2023), l'evoluzione dell'Intelligenza Artificiale può portare verso conseguenze negative. Attraverso l'uso dell'espressione digital manipulation, lo studioso sottolinea che questa crescente potenzialità dell'IA introduce la possibilità di manipolare gli utenti esercitando un'influenza diretta sui loro comportamenti. Tale manipolazione, può innescare effetti indesiderati e profondamente impattanti sulla sfera individuale e sociale e, tra i rischi principali, vi è la perdita di autonomia e libertà. L'IA, attraverso l'analisi predittiva dei dati comportamentali, può anticipare e comprendere le decisioni degli utenti, plasmando sottilmente il loro ambiente per guidarne le azioni. Inoltre, questa capacità manipolatoria dell'IA potrebbe accentuare anche le disuguaglianze sociali, favorendo alcune categorie di utenti a discapito di altre e creando divisioni profonde e compromettendo la coesione sociale.

### **Etica e rappresentazione**

Per promuovere uno sviluppo etico delle tecnologie emergenti, come la Robotica Sociale e l'Intelligenza Artificiale generativa, è cruciale adottare una visione più ampia della società. Gli studiosi e le studiose che esplorano le interazioni tra gli utenti e i robot si trovano di fronte a scelte fondamentali riguardo all'attribuzione di un genere agli agenti artificiali. Superare gli stereotipi di genere

nella progettazione dei robot, andando oltre le norme sociali consolidate, offre l'opportunità di riconoscere e superare gli attuali limiti, promuovendo una progettazione più inclusiva e consapevole. Inoltre, considerando la cultura come un elemento chiave nella costruzione di agenti artificiali, un approccio sostenibile nell'attribuzione del genere ai robot può contribuire al miglioramento delle relazioni anche tra i generi nella società attuale (Alesich & Rigby 2017).

Il settore della progettazione dei Sex Robot, tuttavia, sembra apparire più spinoso a causa della visione stereotipata e pornografica delle donne che essi tendono a comunicare. Difatti, l'estetica dei Sex Robot, concepita per rispondere alle aspettative del target di riferimento, sembra convergere verso standard di bellezza idealizzati e spesso distorti. Proprio in merito è stata condotta una ricerca (Oleksy & Wnuk 2021) volta ad evidenziare come la loro l'estetica e le strategie di marketing adottate dalle case produttrici contribuiscano a diffondere standard irrealistici e a promuovere prestazioni sessuali che possono generare ansie e paure tra le donne, contribuendo a perpetuare norme inaccessibili e dannose. Ciò potrebbe contribuire a una crescente pressione sociale e culturale sulla sessualità, con possibili impatti sulla salute mentale e sul benessere emotivo. Per questi fattori potrebbe essere utile impiegare il modello circolare proposto da Dubé e Anctil (2021), secondo i quali le attuali reazioni del mercato dei Sex Robot possono servire alle aziende produttrici, le quali modificando i loro robot secondo le richieste dei potenziali clienti potrebbero sia migliorare il proprio profitto, sia rendere il prodotto più eticamente sostenibile e rivolto ad un pubblico notevolmente più ampio.

Anche le riflessioni etiche sull'equità dei sistemi di IA generativa si muovono su un piano che intende coinvolgere tanto le tecnologie quanto la società. Ferrara (2023) ha proposto quattro punti chiave per una visione etica dell'IA: la rappresentazione, necessaria per riprodurre la diversità dei punti di vista, delle esperienze e dei background che coesistono nella nostra società; la trasparenza dei dati che vengono impiegati nell'addestramento dell'IA; la responsabilità nel supervisionare i progressi e gli sviluppi della tecnologia in questione; l'inclusività dei linguaggi, delle culture e il superamento del divario digitale.

Anche la prospettiva definita Data Feminism (D'Ignazio e Klein 2023) rappresenta una lente critica attraverso la quale si possono esplorare le dinamiche di genere nell'era dei dati e dell'IA. Questa prospettiva si concentra sull'analisi delle intersezioni tra tecnologia, dati e questioni di genere, riconoscendo che i dati non sono mai neutri, ma sono plasmati dalle prospettive e dai contesti sociali nei quali si sono sviluppati.

Questa corrente di pensiero promuove, quindi, la necessità di coinvolgere donne, persone di varie identità di genere e altre minoranze nei processi decisionali riguardanti la raccolta e l'analisi dei dati, garantendo che le tecnologie riflettano una gamma più ampia di prospettive. Proponendosi di essere una guida per un

cambiamento strutturale nell'ambito della tecnologia, il Data Feminism, attraverso un approccio intersezionale, vuole mettere in atto una revisione critica delle pratiche attuali e spingere per un approccio più inclusivo ed etico nella progettazione e nell'uso dei dati e delle tecnologie digitali. Offre quindi una prospettiva innovativa che incoraggia a esaminare in modo critico le dinamiche di genere nell'era digitale, invitando a un cambiamento positivo e trasformativo.

## **Conclusioni**

Superare i bias di genere nel settore tecnologico si prospetta una sfida ardua e complessa. Questa difficoltà ha diverse ragioni, tra cui le imperfezioni intrinseche degli esseri umani e la permeabilità dei loro pregiudizi nelle tecnologie implementate. La sfida diventa ancor più intricata quando si considera il contesto sociale in cui si sviluppano le tecnologie; spesso, i bias di genere sono radicati profondamente nelle stesse strutture culturali e sociali e ciò significa che relegare agenti artificiali dalle fattezze femminili nel settore della cura e promuovere una visione sessualizzata delle donne attraverso i Sex Robot, riflette le discriminazioni già presenti nella società. Allo stesso modo, l'IA generativa producendo testi ricchi di stereotipi e legati a ruoli di genere "tradizionali", può indurre l'utente a non riuscire ad emergere da quella visione della realtà che sembra porre uomini e donne all'interno di categorie immutabili e sempre uguali nel tempo.

Il processo di superamento degli stereotipi richiede un cambiamento culturale profondo e una consapevolezza diffusa, che va al di là del mero sviluppo tecnologico. È necessario un approccio olistico che coinvolga non solo le scienze ingegneristiche e matematiche, ma anche le scienze umane, le scelte politiche e la cultura nel suo complesso.

La relazione tra tecnologia e società è infatti un intreccio complesso di influenze reciproche che si plasmano continuamente l'una sull'altra, modellando il tessuto stesso della nostra esistenza. La tecnologia agisce come un riflesso delle nostre esigenze, aspirazioni e dinamiche sociali, mentre la società si adatta e viene ridefinita dalle innovazioni tecnologiche che essa stessa crea.

## RIFERIMENTI BIBLIOGRAFICI

Alesich, S., e Rigby, M., Gendered robots: Implications for our humanoid future, *IEEE Technology and Society Magazine*, 36(2), pp. 50-59, 2017.

Austin, J. L., *How to Do Things with Words*, Oxford: Clarendon Press, 1962.

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

Brown T., Mann B., Ryder N., Subbiah M., Kaplan J.D., Dhariwal P., Neelakantan A., Shyam p., Sastry G., Askell A., *Language models are few-shot learners Advances in neural information processing systems*, 33, pp. 1877–1901, 2020. [arXiv:2005.14165]

Butler, J., Performative Acts and Gender Constitution: An Essay in Phenomenology and Feminist Theory, *Theatre Journal*, 40, pp. 519–31, 1988.

Butler, J., *Gender Trouble: Feminism and the Subversion of Identity*, New York: Routledge, 1990.

Carter, C. R., Kaufmann, L., e Michel, A., Behavioral supply management: a taxonomy of judgment and decision-making biases, *International Journal of Physical Distribution & Logistics Management*, 37(8), pp. 631-669, 2007.

Chita-Tegmark M., Lohani M., e Scheutz M., Gender Effects in Perceptions of Robots and Humans with Varying Emotional Intelligence, *14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2019.

D'ignazio, C., e Klein, L. F., *Data feminism*. MIT press, 2023.

Danaher, J., e McArthur, N., *Robot sex: Social and ethical implications*, MIT Press, 2017.

De Angeli, A., e Brahnham, S., Sex stereotypes and conversational agents, *Proc. of Gender and Interaction: real and virtual women in a male world*, 2006.

J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, B., *Pre-training of deep bidirectional transformers for language understanding*, 2018 [arXiv:1810.04805]

Dubé, S., e Anctil, D., Foundations of erototics, *International journal of social robotics*, 13(6), pp. 1205-1233, 2021.

Dubé, S., Santaguida, M., Anctil, D., Zhu, C. Y., Thomasse, L., Giaccari, L., e Johnson, A., Perceived stigma and erotic technology: From sex toys to erotots, *Psychology*

& *Sexuality*, 1-17, 2022.

Ferrara, E., *Should chatgpt be biased? challenges and risks of bias in large language models*, 2023 [arXiv:2304.03738].

Galizia, R., Sex robot: "bene" di consumo o "beneficio" terapeutico?, *Rivista di sessuologia clinica*, 2020.

Guo, W., e Caliskan, A., Detecting emergent intersectional biases: Contextualized word embeddings contain a distribution of human-like biases, *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 122-133, 2021.

Gutiu, S. M., The roboticization of consent. In *Robot law*. Edward Elgar Publishing, 2016.

Hegel F., Eyssel F., (s)he's got the look: Gender stereotyping of robots, *Journal of Applied Social Psychology*, 42(9), pp. 2213-2230, 2012.

Hu, K. *ChatGPT Sets Record for Fastest-Growing User Base*. Reuters, February 2, 2023. Disponibile online: <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>

Ienca, M., On Artificial Intelligence and Manipulation, *Topoi*, pp. 1-10, 2023.

Lorentzen, B., *Social Biases in Language Models: Gender Stereotypes in GPT-3 Generated Stories*, 2023. [Tesi di laurea]. Disponibile online: <https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-484882>

Lucy, L., e Bamman, D., Gender and representation bias in GPT-3 generated stories. *Proceedings of the Third Workshop on Narrative Understanding*, pp. 48-55, 2021.

Magaudda, P., e Neresini, F., *Gli studi sociali sulla scienza e la tecnologia*. Società editrice il Mulino, Spa, 2020.

Meyer, M., Cim pian, A., e Leslie, S. J., Women are underrepresented in fields where success is believed to require brilliance, *Frontiers in psychology*, 235, 2015.

Nangia, N., Vania, C., Bhalerao, R., e Bowman, S. R., *CrowS-pairs: A challenge dataset for measuring social biases in masked language models*, 2020. [arXiv:2010.00133].

Odlind, C., Richardson, K., The End of Sex Robots—For the Dignity of Women and Girls, in *Man-Made Women: The Sexual Politics of Sex Dolls and Sex Robots*, pp. 1-16. Cham: Springer International Publishing, 2023.

Oleksy, T., Wnuk, A., Do women perceive sex robots as threatening? The role of political views and presenting the robot as a female-vs male-friendly product, *Computers in Human Behavior*, 117, 106664, 2021.

Ouyang L., Wu J., Jiang X., Almeida D., Wainwright C.L., Mishkin P., Zhang C., Agarwal S., Slama K., Ray A., *Training language models to follow instructions with human feedback*, 2022[arXiv:2203.02155]

Pasciuto, F., Genderizzazione degli agenti artificiali: stereotipi e pregiudizi nell'IA, *Humanities*, 11(2), pp. 101-111, 2022.

Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., *Improving language understanding by generative pre-training*, 2018.

Ray, P. P. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope, *Internet of Things and Cyber-Physical Systems*, 2023.

Richardson, K., Sex robot matters: slavery, the prostituted, and the rights of machines, *IEEE Technology and Society Magazine*, 35(2), pp. 46-53, 2016a.

Richardson, K., The asymmetrical 'relationship' parallels between prostitution and the development of sex robots, *ACM SIGCAS Computers and Society*, 45(3), pp. 290-293, 2016b.

Richardson, K., *Sex robots: the end of love*. Polity Press, 2018.

Sharkey, N., van Wynsberghe, A., Robbins, S., e Hancock, E., Our sexual future with robots. *Foundation for Responsible Robotics*, 5, 2017.

Sparrow, R., Robots, rape, and representation, *International Journal of Social Robotics*, 9(4), pp. 465-477, 2017.

Stanovich, K. E., e West, R. F., On the relative independence of thinking biases and cognitive ability, *Journal of Personality and Social Psychology*, 94(4), pp. 672-695, 2018.

Steinbock, B., Speciesism and the Idea of Equality, *Philosophy*, 53(204), pp. 247-256, 1978.

Szczuka, J. M., & Krämer, N. C., Influences on the intention to buy a sex robot. *International Conference on Love and Sex with Robots*, pp. 72-83, 2016.

Tay B., Jung Y., Park T., When stereotypes meet robots: the double-edge sword of robot gender and personality in human–robot interaction, *Computers in Human Behavior*, 38, pp. 75-84, 2014.

Tversky, A., e Kahneman, D., Availability: A heuristic for judging frequency and probability, *Cognitive psychology*, 5(2), pp. 207-232, 1973.

Upson, S., Friedman, L. F., Where are all the female geniuses?, *Scientific American*

*Mind*, 23, pp. 63–65, 2012.