IM@GO
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

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.

ISSN: 2281-8138
N. 23 - Year XIII / July 2024 pp. 429-438
www.imagojournal.it

Introduction

Memory 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

ISSN: 2281-8138

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?"

ISSN: 2281-8138

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

ISSN: 2281-8138

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.

ISSN: 2281-8138

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.

N. 23 – Year XIII/ July 2024 pp. 429-438
imagojournal.it
ISSN: 2281-8138

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.

ISSN: 2281-8138

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?

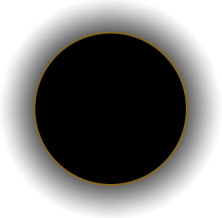
The project does not offer a definitive answer to this question but instead invites viewers and readers to reflect on the dual potential of AI as both a preserver and distorter of knowledge. The future of humanity's relationship with AI may hinge on how we choose to manage this ambivalence—whether we 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

ISSN: 2281-8138

role AI will play in shaping our collective memory and understanding of history in the digital age.



N. 23 – Year XIII/ July 2024 pp. 429-438
imagojournal.it
ISSN: 2281-8138

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.

N. 23 – Year XIII/ July 2024 pp. 429-438
imagojournal.it
ISSN: 2281-8138