xCoAx 2024
12th Conference on Computation, Communication, Aesthetics & X
Fabrica, Treviso, Italy
2024.xCoAx.org

Mimesis and Machinic Agency: An Exploration of Autonomous Image-Text Loops

Darío Negueruela del Castillo
dario.neguerueladelcastillo@uzh.ch
Digital Visual Studies, Zurich, Switzerland

Iacopo Neri
neri.iacopo@gmail.com
Digital Visual Studies, Zurich, Switzerland

Pepe Ballesteros Zapata
jose.ballesteroszapata@uzh.ch
Digital Visual Studies, Zurich, Switzerland

This article explores the possibility of speaking of art in the context of stacked multimodal machine learning models performing recursive translation loops between image and text. We introduce an experiment consisting of a generative autonomous loop that translates iteratively between textual prompts and imagerial outputs to depart independently from a given input. The evolution of the loop is analysed both computationally, through metrics tracking model convergence, and qualitatively, through critical interpretation of the generated outputs. We elaborate on the uncanny articulations produced through this mimetic process and discuss how they urge new debates on machinic agency and aesthetics.

1. Introduction

Recent advances in multimodal machine learning, especially in vision-language models like DALL-E, Imagen, and Stable Diffusion, have enabled unprecedented robotic creativity in the automatic generation of images from textual prompts. The resulting outputs exhibit a mystifying capacity not just for photorealistic rendering, but also for conveying metaphor, symbolism, and affect through creative recombination of visual tropes and icons. It is often commented that these models seem to make images that appear meaningful without possessing meaning. But what might it mean for an artificial system to "make meaning" in the first place? Can we meaningfully speak of concepts like creativity, imagination, and aesthetics in reference to machines? Such questions urgently warrant revisiting enduring philosophical debates on mimesis, agency, authorship, and interpretation; and have recently been heatedly discussed after the irruption of capable generative models.

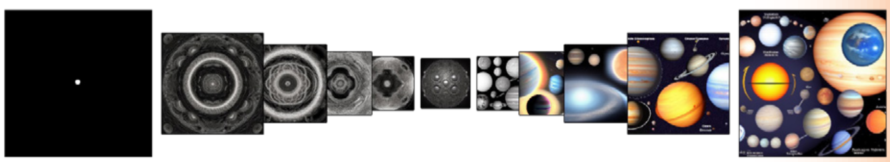

Fig. 1. Starting image and end image of the autonomous generative loop. The textual prompts are not featured.

Self-referentiality and mimetic processes of translation between modalities are perennial topics in discussions on the essence of novelty in art and creativity (Gebaure & Wulf 1996). With the latest developments in AI, recent works have also revisited and experimented with this theme[1]. In this article, we develop an experimental framework to

1. See for instance the 2017 video installation by Jake Elwes (2017), UNI_VERSE Studio's interactive installation "Recursive Reflections" (2023) and Papatheodorou and Dilaura's Visual Dialogues (2023)

Keywords GenAI, Mimesis, Artificial Intelligence, Stable Diffusion, CLiP.
DOI 10.34626/2024_xcoax_016

explore these issues through a simple but conceptually provocative exercise: constructing an autonomous loop between computer vision and natural language processing modules that recursively translates between textual and visual modalities. We detail the implementation of this system using state-of-the-art AI and analyse the results from both a computational and a cultural perspective.

Our aim is not to advance purely speculative claims about machine consciousness or creativity. Rather, by rendering visible the mimetic machinations of such a system, we hope to problematize reductive assumptions about generative AI and inspire further reflection on the complex liminal spaces emerging between humans and intelligent machines, and by extension, contribute to central and inexhaustible debates in visual culture and art history on the nature of art.

1.1. Theoretical Background

Theoretical investigations of mimesis have a long lineage within Western aesthetics, from Plato's condemnation of poetic imitation divorced from truth to Aristotle's rehabilitation of mimesis as an articulation of universal forms, to later figures like Vico, Adorno, and Benjamin (Potolsky 2006). A key tension between notions of mimesis as passive mirroring versus active recreation or reimagination of the world persists. Within theory on photography, for instance, this tension manifests in debates around indexicality and automation. From Barthes' notion of the photographic image as an emanation of the real (1977), to Flusser's conception of cameras as programmed and programming apparatuses[2], to more recent discussions of computational photography, there is rich disagreement around photographic mimesis as imprint, construction, or simulation (Cadava 1997).

The advent of AI generative models forces a resurgence of these debates. As Manovich observes, whereas earlier mimetic media like photography could only sample from existing reality, deep generative models can synthesise new realities (Manovich 2023). This collapsing of sampling and simulation summons Baudrillard's concept of the hyperreal, for whom the perfected simulacra of postmodern media no longer imitate or represent reality, but rather precede and generate reality through models and codes (Baudrillard 1994).

AI generative systems like DALL-E, Midjourney or Stable Diffusion, which 'imagine' images seemingly from thin air appear to realise Baudrillard's vision. But should we take the advertising rhetoric of "imagination" and "creativity" seriously in reference to machines? Critics argue such anthropomorphic terms misleadingly impart machinic processes with humanistic sensibilities (Salles, Evers, and Farisco 2020). Against this, others advocate for an "AI humanism" which genuinely grants intelligent systems creative agency (Lewis 2022).

These perspectives resonate in contemporary new materialist thought, which similarly critiques anthropocentric ontologies and advocates a "flat" ontology that connects the human and nonhuman within hybrid and distributed networks and relations (Bennet 2010). Some proponents of new materialism celebrate generative AI art as indicative of posthumanist distributed creativity. For example, Goriunova observes how robots trained on vast datasets intuitively remix existing cultural material to conjure affects through nonhuman associative logic (Goriunova 2023). On the other hand, some scholars advocate the

2. "The camera encodes the concepts programmed into it as images in order to program society to act as a feedback mechanism in the interests of progressive camera improvement" (Flusser 2000, p. 48)

3. With Transindividuation further elaborates on the concept of individuation coined by Gilbert de Simondon to refer to the process by which individual subjectivity emerges from collective cultural and symbiotic systems. For instance he "I" is composed of intergenerational accumulations of memory, technology, language, beliefs, etc. These diverse inheritances constitute a transindividual milieu. Transindividuation describes how the "I" is continuously transformed through its embeddedness in sociotechnical systems beyond itself. (Stiegler 1998)

idea that while art has always involved technology and artificial intelligences, many of the current developments and applications of AI for artistic practices and especially image generation, fall in the inane category of Candy Crush-like generators of spectacle without substance (Zylinska 2020). In other words, this position maintains that in lacking a subjective autonomous dimension, generative AI systems like the ones above-mentioned, actually refute posthumanist theories of materially embedded distributed cognition, instead of reflecting them. Thus, despite their apparent convergence, a closer look quickly reveals the tensions between theoretical posthumanism which decentralises the human, versus the practical instantiation of supposed posthuman intelligence in current AI.

The question of mimesis at play in generative AI is further articulated in the works of thinkers like Yuk Hui and Bernard Stiegler. For Hui, machine learning algorithms manifest immanent creativity, so framing AI within instrumental goals misconstrues its mimetic capacities. Generative AI, thus, does not impose external programs but instead aligns itself with the grain of things, according to Hui, inductively discerning its innate structures and articulating, by interpolation, novel - latent- versions of them. For Hui, these types of models bring to light the creative potentials already at play within the ontological flux of reality (Hui 2016). His cosmotechnics reimagines automation, creativity and cognition beyond anthropocentrism. On a different line on post anthropocentric mimesis, although not necessarily antithetical, Bernard Stiegler argues that AI lacks veritable open-ended human imagination. recognizes the potential dangers of automation to human culture and memory. He advocates reconceiving computational mimesis as a process of transindividuation[3] so that machinic mimesis amplifies, rather than attenuates, the long circuits of memory enabling collective significance.

Critical discussion of AI aesthetics must be situated within this broader discursive context around mimesis, creativity, agency, and posthumanism. With this conceptual scaffolding established, we now introduce some core ideas behind the AI systems used in our work as well as the experimental framework. Then, afterwards, we will proceed to discuss the experiment in casual friction with theories on the image, creativity and art.

2. The Autonomous Loop

At the core of our experiment lies a simple yet powerful loop between computer vision and natural language processing functions: a cycling process that gives rise to an intriguing space of opaque translation between image and text modalities. A space which becomes crucial, as it aims to provide a technical opportunity to capture the essence of imagination's inherent incommunicability and the autopoietic nature of artistic research. It is here that the true creative potential of our experiment unfolds, within the fractures of multimodal interpretation.

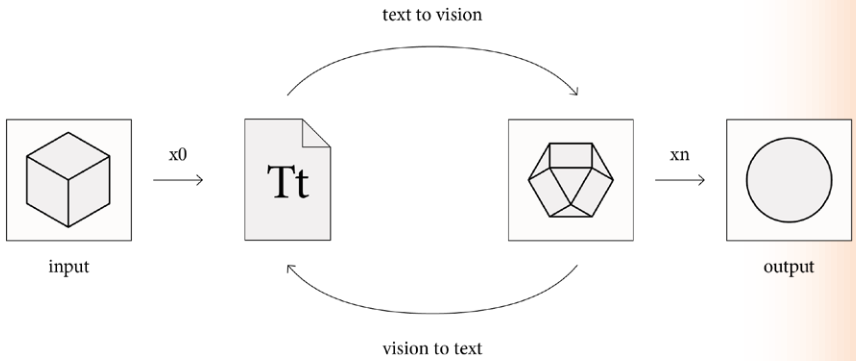

Fig. 2. Diagram of the autonomous loop.

Like in a game of broken telephones, two machine learning models interact with each other, exchanging data in turns, so that the output of one becomes the input for its counterpart. Text generates an image that later generates a text, and so on, until external conditions are met or an amount of iterations is reached (Figure 2). By compelling the generative models to produce not machine-readable embeddings but human-readable artefacts, we navigate beyond the confines of mere precision and into the realm of nuanced ambiguity.

2.1. Computational Pipeline

The loop is structured as follows:

1. Input an initial image
2. Pass the input through a vision to text model to generate a textual description
3. Pass this text through a text to vision model to generate a corresponding image
4. Repeat steps 2 and 3, using the latest output as input for the next iteration

This recurs until forcibly terminated. The key innovation here is the recursive chaining together of two translation modules to create an autonomous loop or feedback circuit. The system thus becomes generative, able to synthesise new semantic chains escaping the original input. It is important to notice that multiple runs of the loop with the same input might not trigger the same chain of translations. More precisely, the machine learning models involved are fundamentally stochastic, yet governed by a numeric value, or seed, that is responsible for the initial random conditions from which they begin their calculations. Under identical settings, using the same seed will produce the same results, so we deliberately avoid fixed seed values and foster a spirit of exploration in the generation of diverse and variable results. In conclusion, the pipeline implies a non-unique translation between a textual prompt and an image.

To run the process, two machine learning models are adopted: CLiP Interrogator[4] and Stable diffusion[5], respectively for the task of textual description and image generation. CLiP Interrogator is a model to infer a textual prompt from a given image in order to support representational exploration. As stated by its name, CLiP Interrogator is based on CLiP (Contrastive Language-Image Pretraining) which is the foundational model at the core of this whole set of products, responsible for

4. Source: https://github.com/pharmapsychotic/clip-interrogator

5. Source: https://huggingface.co/runwayml/stable-diffusion-v1-5

the translation between image and text pairs. What CLiP Interrogator does is try to unfold the image-to-text translation process by exposing a set of human readable prompts, ranging from details on the subject depicted, the settings, or the style and up to more abstract media-oriented keywords representative of online trends, to describe a given image. Technically, it compels the model to select within a vast range of terms, the most suitable keywords. Stable Diffusion is a state-of-the-art diffusion-based model known for its high-quality text-to-image generation and for having its code and weights publicly accessible. Parallel to generating images via a text description, it provides additional tools such as inpainting, which replace content within an image, outpainting, which extend an image out of its frame, and generating image-to-image translations. The latter steers the image generation directly through another image, which aims at compositional reproduction, therefore optimising the generation in order to obtain a visual twin of the input image. Three key elements behind such technologies are explained more in depth in the following.

2.2. Diffusion

From the rapid path that a meme follows through social media, to the way the ink spreads gradually into a glass of water, diffusion describes the process of transforming matter from concentrated to dispersed. Inspired by algorithms originally developed for physical simulation, the principle of diffusion is exploited in stable diffusion models to transform random pixel distributions – particles of visual information – into structured images. In this regard, it echoes the artistic process of a painter itself – starting with a white canvas and diffusing a visual image through layers of painting until the final image is composed. Only the learning methodologies are different, or rather opposed. Stable diffusion models learn through an inverse process, akin to a sculptor working backward from a finished piece to a raw block, effectively reversing the steps to create from randomness. The reason behind this inverse diffusion in training is simple: it is easier to go from an image to randomness, than from randomness to an image, because the former applies random pixel values to an image, and the latter needs to transform randomness into a known pixel distribution.

Far from merely revealing a technical curiosity, the way contemporary image generation models are trained have profound implications on the way we understand the act or artistic creation in the post-digital era. Traditional visual theories like those proposed by W.J.T. Mitchell or Walter Benjamin focus on the artefact itself – the finished painting, its reproduction, its spatial context. Diffusion models challenge the artefact-centric view by suggesting that the process of creation is not linear but cyclical, not just a journey from nothingness to completion but a continuous loop of creation and deconstruction.

2.3 CLIP

Contrastive Language-Image Pretraining (CLiP) is a foundational model architecture developed by OpenAI that marked a significant milestone in AI research upon its release on January 5, 2021[6]. Foundational models in AI serve as the core architectural components for larger and more complex models. What made CLiP the backbone of several mod-

6. See https://openai.com/research/clip

els working on visual and textual data, lies in its implicit multimodality. Compared to previous category-dependent models that excelled at understanding text or images respectively, CLiP learns jointly through image-text pairs and interprets the world in a way that arguably is more similar to how humans do. In the training process, the model not only learns the best image-text pair fit, but also discerns why the other descriptions do not align with the target image. This is known as "contrastive learning" strategy – hence the models' name. Finally, the model is able to move beyond learning simple tags for images and it has proven capable of interpreting nuanced descriptions of the visual world, raising consequent concerns on the role of interpretation and its constituent system of references.

2.4. Embeddings

The last concept needed to fully capture the backbone behind the proposed pipeline are embeddings, as opposed to human readable artefacts. In the context of CLiP, embeddings are a numerical translation of the visual and text information that capture the most important features of the data. These features allow us to extract superficial information such as colours, shapes, syntax, vocabulary; but also complex phenomena such as objects, patterns, context, or tone and mood from a text. Just like every painting or sculpture has unique characteristics that set it apart – colour, texture, or subject matter – an embedding captures these unique features but in a numerical form that a computer can understand. Contrary to the standard communication protocols via embeddings, the discussed experiment proposes the iterative exchange of human readable artefacts – images and texts – to explore the operationalization of the underlying CLiP model rather than its mere actualization.

3. Assessments

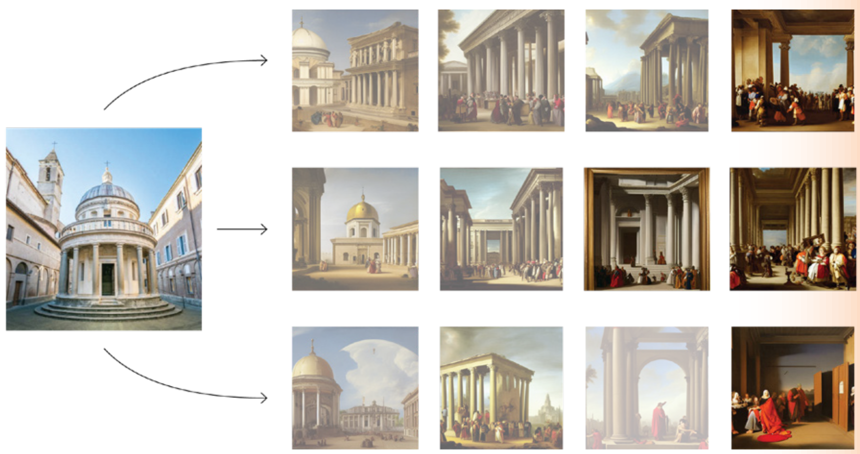

Fig. 3. Example of three image sequences highlighting the mentioned influence of colonnades in the image composition. In transparency, the switch in the point of view: from an outdoor setting to an indoor space.

We begin our evaluation of the pipeline and its outputs with a visual assessment of the images generated, their journeys and their descriptions, later supported by a more computationally-oriented approach. In the space of opaque translation between human readable texts and images, we discover the relevance of abandoning the conventional expectations of a one-to-one correspondence between input and output to embrace the underlying ambiguity and complexity of cultural cognition embedded in the models. This exercise, more than a simple artistic endeavour, aims at shedding light on the possible understandings, biases and tendencies of the models exploited. In this regard, by focusing not solely on the generated artefacts themselves but on the generative trend or trajectory, we draw parallels with mathematical analysis, where derivatives provide insights into the behaviour of functions rather than their individual values.

3.1. Visual Assessments

Complex narratives emerge as images and texts interplay, aligning with architectural and visual theories on elements and symbols. For instance, sequences generated from architectures of Renaissance and Baroque Rome (Figure 3) unsurprisingly depict urban scenes of plazas, markets and noblemen with a particular focus on classical architectural elements such as colonnades. The latter precisely, seems to possess a distinctive importance in the image generation as it often anticipates a later change in the composition of the image, with the point of view transitioning from an exterior setting to indoor spaces. In this sense, colonnades resonate with actions of movement and transition as it is well recognized in architectural history through their role as urban filters between public and private areas. Interestingly, a contrasting trend emerges when the generative process begins with an image of a skyscraper, which tends to produce images that adopt a bird's eye view perspective, focusing on cityscapes and panoramic views of urban landscapes (Figure 4).

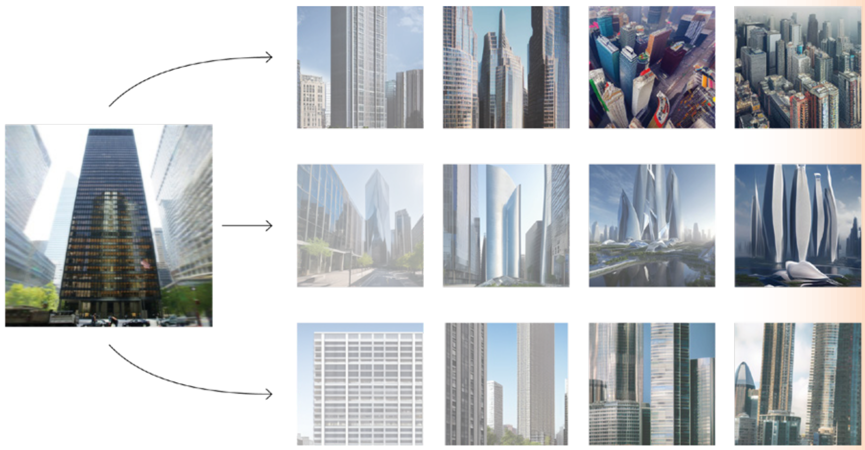

Fig. 4. Example of three image sequences highlighting the journey from a skyscraper input image. In transparency, the progressive distance of the point of view: from building to city.

3.1. Computational Assessments

The cyclical influence between text and image is the key element to the understanding of the loop's dynamics. Following an initial qualitative analysis of the generated images, the serendipity of the experiment – at times settling into a consistent theme while at others remaining widely oscillating between different ideas – motivated a quantitative comparison. Convergence plots were selected as quantitative lenses to visualise the stability of the generated content over time. They portray whether each generation "look and feel similar "to previous results (converging), or wildly different (diverging). Showing the model's behaviour when both aligning to a consolidated pattern as well as exploring new path-ways, convergence plots are intended in this research as representative of the models' 'machinic creativeness', or its creative rhythm.

Technically, image and text embeddings are collected to calculate how different they are in two separate iterations. The process of computing this difference is what we call the 'loss.' A high loss means the new image or text is very different from the previous one, while a low loss means it's significantly similar. Finally, plotting all loss calculations over time gives shape to the convergence plot.

Being a generic difference between embeddings, the loss can also be evaluated across several moments of the loop to elucidate on distinct model's behaviours. In this regard, losses are computed in a twofold manner: against the original input, and against a limited set of previous iterations. Respectively adopting the baseline or the rolling average of the losses, these two trends aim at depicting the capacity of the model to wander into undefined themes and its loosely fuzzy rhythm.

Whether the baseline evaluation clearly showcases the open-ended nature of the experiment as a simple comparison between the embeddings of the first generation and each iteration's embeddings, the detection of sudden changes of topic within the loop requires a more nuanced approach. Thus, a rolling average (a moving average iteratively calculated) was selected to represent the stability or instability of the loop. In other words, each generation is analysed together with its anticipating instances to begin a process of thematization, where consequent similar iterations are visually identified and separated from globally diverging ones. Technically, this rolling average convergence plot is computed by calculating the difference between the current loss – the difference between the current and previous generation, and the mean loss of the last N iterations (empirically, we selected the three previous iterations for this publication).

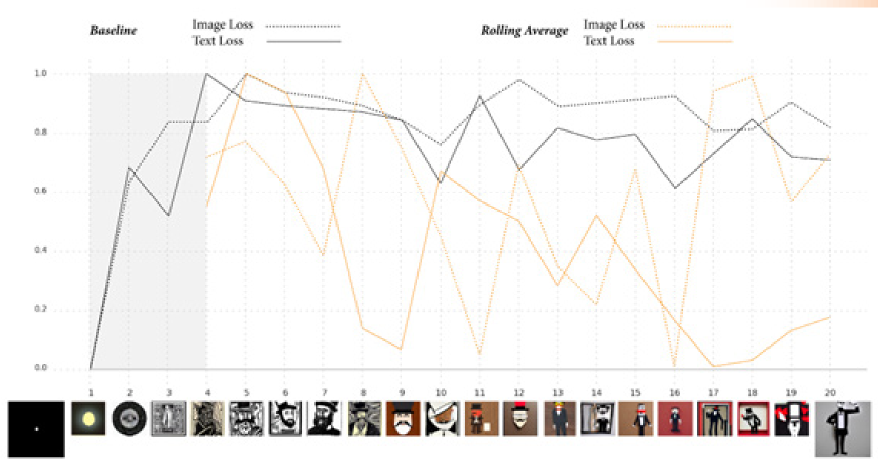

Fig. 5. Example of a convergence plot.

We begin testing the convergence-divergence pipeline using an input image of a white dot on a black background, favouring basic visual motifs over more complex ones. It is in our interest to test the capacity of the model to produce unexplored visual thematics and its intrinsic tendency toward the figurative. Simple images are easier to define thematically, while complex visuals can be challenging to describe. Taking inspiration from Durer's theoretical writings on painting[7], ideas come from abstraction, and the most abstract aspect of reality is the point. The point does not exist in reality and serves as the matter to all forms.

7. Fowler, Caroline O. Drawing and the Senses in Early Modern History. Turnhout, Brepols, 2016.

Far from our initial purpose, a glimpse into the extracted prompts from the proposed input image gives us an idea of the model's interpretative framework: "A rafted moon in the dark sky with a few clouds," "humpty dumpty in form of an egg," "The stone is round." It is clear that the model's strive for figuration automatically converts abstraction into a tapestry of routinary objects. Nevertheless, the simple yet abstract geometric forms allow an easy visual exploration of a variety of visual topics.

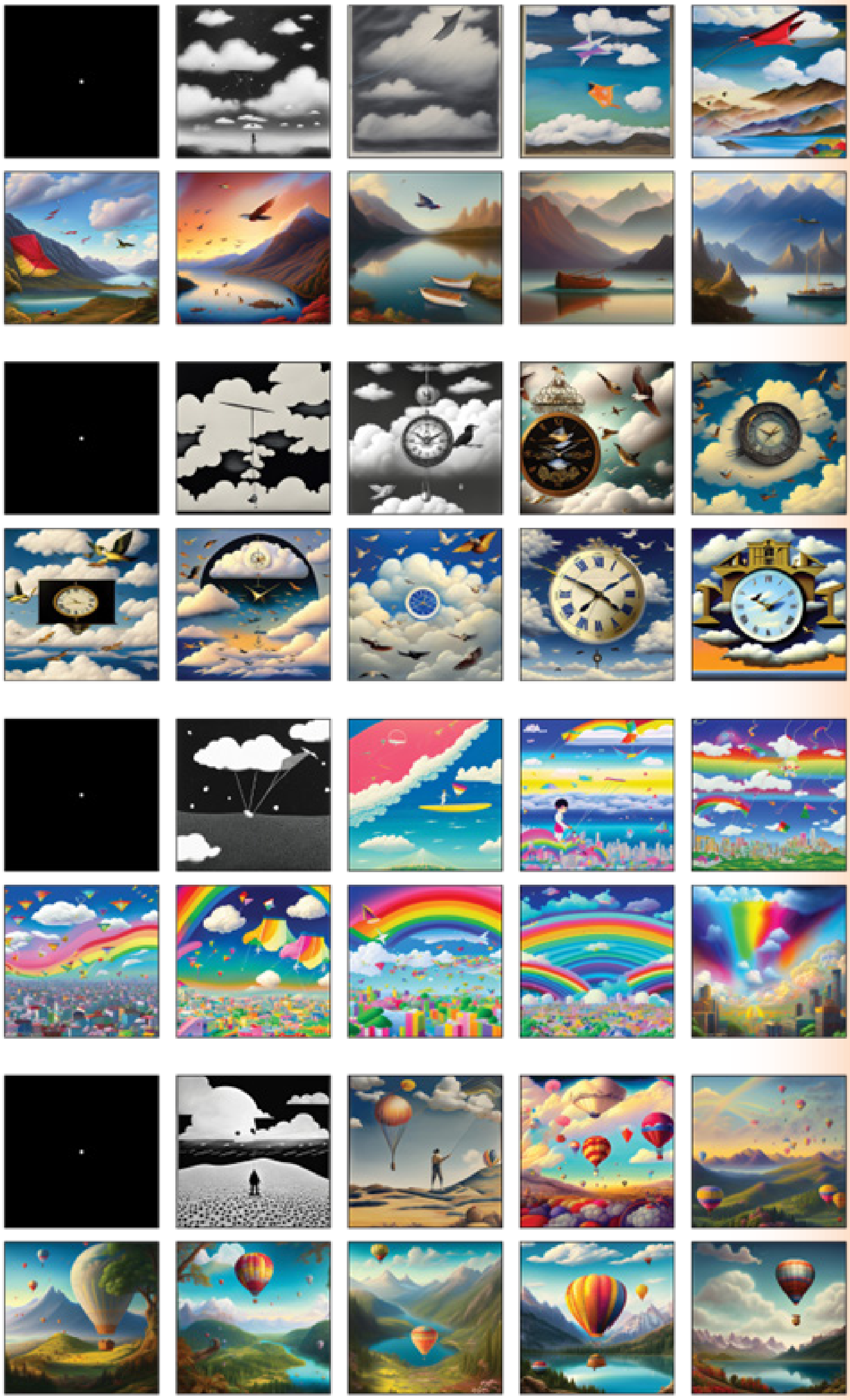

Fig. 6. Examples of generated images. Each loop can be read from left to right, top to bottom.

For what matters the baseline evaluation, numerous experiments begin with a rapid shift from the original input, as shown in Figure 5. This behaviour can be explained by the numerous artefacts the CLiP interrogator model introduces in interpreting the input image. Moreover, we see from the convergence plot that once the loop is running, the likelihood of the model returning to a state resembling the original input is exceedingly low, therefore tending towards the discovery of a new topic. This becomes a quantitative demonstration of the models' tendency towards figuration: the difference between the input image and the second image is large, and such difference is maintained in time until reaching convergence. The model is able to extract a variety of visual topics from a simple input image but will never generate a simple image from a complex visual composition (Figure 7).

On the other hand, spikes or sudden changes registered on the rolling average convergence plot can be interpreted as experimental moments where the machine deviates from its current topic. This is caused by unexpected wording in the extracted prompt – creative analysis of the output image, or by unexpected images created from similar prompts – creative visual generation. The interaction between the image and text loss serves as an indicator of how much each generation deviates from the average of the last N generations. When both the text loss and image loss move in the same direction, it suggests a coherent shift in both the visual and textual aspects of the generated content. When this coordinated direction moves upwards means that the model is exploring a new theme, exploring new territories in both visual and textual domains, while downwards suggests that the model may be settling into a more stable or repetitive pattern. While this is the case for some iterations, we see from the models' behaviour that synchronised movement is seldom found for a large number of iterations. It is a common scenario through the looping process to see a divergence between the visual and textual content. Hence, the most common pattern found within these plots is that the model might be generating images that are visually similar to the recent trend but are described differently in text, or vice versa. In short, models are not well-aligned in representing the same concept. In most cases, such a model's behaviour serves as a trigger to avoid complete convergence, simulating small regions of convergence, but always escaping the cycle towards new thematic paths. These non-linear dynamics found in convergence plots suggest that the model is prone to sudden changes in the embedding space, which ensures a tendency of ever-topic change when approaching infinity[8].

8. The loop is only disrupted when the generated image or text has the potential to exceed the ethical boundaries of model usage – paradoxically, it is the model itself that is generating explicit content that goes beyond what is permissible as input. We have surpassed this limitation by forcing the model to generate within the current iteration until the output is not characterised as "NSFW" (Not Safe For Work).

4. Conclusions

> *The situation is precisely the reverse: language and imagery are no longer what they promised to be (...), transparent media through which reality may be represented to the understanding. (...) language and imagery have become enigmas, problems to be explained, prison houses which lock the understanding away from the world. (Mitchell 1984)*

Beyond the computational morphodynamics at play in the models we have used for this experiment, it remains challenging to attempt a qualitative interpretation within the concepts and epistemological grounds of visual studies in the broader sense. However, it is also urgent to do so, and perhaps even fruitful to attempt to discuss this experiment in light of some relevant theories and insights on the nature of images. Without the ambition to sketch a comprehensive discussion, let us examine the unfolding of the loop in friction with different theoretical standpoints and interpretations of images and digital art.

In early iterations, outputs largely remain legible as mundane descriptions of ordinary objects, scenes, and actions, reflecting the models' grounded training (Figure 7). Soon we discover there is a particular difficulty with staying in the realm of the abstract, in terms of non-figuration. Several of the iterations of the loop start with simple shapes, such as a point, a cross or a square, and quickly shift towards shapes contextualised in figurative scenes. There is an inherent abhorrence of abstraction encoded in the pipeline, as everything is not simply a sign, but behaves like an index to a multiplicity of concrete associations.

Rather than a perfect mimetic representation, the system appears to articulate meaning through expressive resonance across signs. Forms gesture beyond themselves toward connotative associations from accumulated cultural exposure. The weight of the training dataset cannot be overlooked. The model that performs the textual guidance of the image generation -CLiP- was trained on millions of image-text pairs, from where it has learned generalizable concepts and their visual grounding. In this respect, for CLiP nothing is abstract as any word is linked to a series of visual representations.

What type of mimetic mechanisms are here at work, which allow for an open stochastic association? And what can this mean for our interpretations of these images, which never come alone, but belong to a network of fluctuating and enchained signifiers?

4.1. Precedents and Resonances of Autonomous Generative Loops

The endless translation loop we introduce, in which an image generates text that generates a new image recursively, evokes the concept of associative "trains of endless imagery" described by 19th-century Victorian British art critics. Authors like Archibald Alison argued viewing artworks triggered spontaneous chains of personal associations and emotions in the viewer's imagination. The autonomous cycling of the AI loop seems to parallel this theory of proliferating associations, yet beyond the sphere of individual appreciation. The loop propagates visual and textual mutations in a machinic errand, with each output forming an associative link with the next. Moreover, we must note that for associationists, significance arose from the imaginative process itself, not from mimetic fidelity (Craig 2007). In this sense, the loop manifests a core associationist notion that seems hardcoded in the software architecture itself: that meaning stems from the subjective (in this case machinic and stochastic) proliferation of associations, not from the inherent qualities of signs.

In addition, this combinatory visual logic aligns with surrealist techniques of radical juxtaposition. Breton's definition of surrealism as "pure psychic automatism" freed from conscious control to manifest latent desires (Breton 1924), seems uncannily apt here, In the process of inquiring into what drives the specific interpretational steps in the generative loop, we can ask ourselves if the loop's output can also be seen to channel what Benjamin called the "optical unconscious" of cinema and photography, which reveals elements of reality inaccessible to human vision alone (Benjamin 1931). The system surfaces subliminal patterns encoded from its training data and reconstitutes fragments into alien yet uncannily recognizable formats. Indeed, the autonomous generation of strangely evocative imagery and text performs a kind of automatic allegorical thinking. The system we put in place analogously appears to construct provisional meanings by materialising associations between disparate elements.

The technical intricacy of the generative process mirrors the complexity of allegorical interpretation. Just as decoding an allegory involves unravelling a constellation of symbols, understanding these images requires unpacking the computational operations linking them. This resonance affirms Benjamin's assertion that radical technical media like photography, and in this case generative AI, reveal the device-like nature of art itself. We appreciate that in the case of the loop, aesthetic aura appears demystified as the mechanical framework of the AI pipeline is highlighted, and meanings are conveyed through interconnected transparent networks of relations, in a pseudo-Benjaminian way (Benjamin 1935).

The type of operation we observe here at play may or may not be behind apparent syntactic and semantic density (Goodman 1976), but something makes us remember Panofsky's warning about the danger of images. Thus, these looped visual generations can also become our pharmakon, "the substance of the images that she (art historian) studies is a powerful substance, attractive but altering." (Didi-Huberman 2005), as they keep us fatally projecting their absences in the next iteration yet to come. Nothing surprising in the longer genealogy of contemporary internet's attention economy. Most of these models and platforms make their arrival once our eyes and hands have already been trained in endless doom-scrolling on social media. This networked visual economy seems suspect of changing the status of the image once more, bringing it closer to a sequence of flashes that literally hit us physically, affecting us in more immediate, bodily ways. The new status of the digital image in late capitalism is of course subject to in-depth critique from different perspectives by authors like Crary, Mirzoeff and Manovich, and it escapes the scope of our work here.

However, there is another unexpected pharmakon effect at play here. With the advent and rapid progress and popularisation of generative AI, the sheer amount of synthetic imagery poured onto the internet keeps growing at an exponential rate. And so far, there has been no widespread method for safe watermarking these images and therefore it will become difficult to filter them out from future training datasets that are based on huge web scraping. In fact, this can lead to the phenomenon of model collapse, which refers to when a model trained on its own outputs ceases to work properly as if clogged in a self-referential cacophony. This appears to point to a different status of whatever 'meaning' synthetic images have encoded or responded to. Perhaps the collapse has to do with a self-referential type of engendering as opposed to the outward and fertile polysemy of human-made images, fundamentally unstable and problematic in their very own non-synthetic elements of 'expression'.

Fig. 7. Initiating the loop from plain colours still ends up either in clear figuration or pattern-based compositions.

5. Open Questions: Mimesis Reimagined

Our experimental generative image-text loop highlights a range of philosophical tensions surrounding mimesis in the age of Generative AI. Notions of creative imagination, authorship, representation, and cognition are all troubled by this simple recursive system. We cannot comprehensively address all these issues in this article, but we have outlined some key problematics raised by the loop.

Firstly, can the loop's seemingly allegorical undercurrents be equated with human aesthetics? Its machinic rearrangements manifest a nonhuman logic, undermining assumptions that meaningful combination is unique to human creativity. However, this does not necessarily make the loop an ideal posthuman embodiment of distributed cognition, as it remains constrained by its disembodied precooked interpretation of 'culture' from the internet's flattening Common Crawl.

At the same time, by algorithmically recycling cultural symbols, the loop seems to instantiate at least some core elements of human imagination. However, this associative process is divorced from personal experience and voids it of a type of meaning that does not solely depend on our own ad-hoc projection. Thus, this ambivalent mimesis questions distinctions between human and machine creation. Indeed, the lack of human intentionality in the generation of images in this specific loop prompts us to ask ourselves if we can indeed speak of a kind of "asemic" yet self-referential image. To further complicate things, the fluctuating divergence of the loop resists binaries of either servile mimesis or random noise. and enacts a fluid promiscuity between copying and novelty. In addition, this unpredictability also challenges authorship and control, as the contingent machine processes at the core of the deep learning models shape the outputs as much as the initial prompts.

The experiment we have set up, while limited in its reach, serves as a valid first approach to the inquiry of machinic mimesis in generative AI. Here, the digital image emerges as a contradiction of the normative Platonic hierarchy and no single original serves as the origin of sense and meaning. Here, mimesis operates through dissemination, a processual and recurrent amalgamation in which the models' latent spaces are simultaneously the medium and the message, a self-referentiality ready to host any meaning we might want to project onto it. Representation gives way to endless morphogenesis as the image is deterritorialized through recursive loops of algorithmic generation to be finally re-territorialized. An endless landscape of liquified signs, where image and text emancipate from referents and attain autonomy, circulating through feedback loops in horizontal self-recreation.

References

Armand, Louis.
2023. *Entropology*. Anti-Oedipus Press.

Barthes, Roland.
2009. *Mythologies*. Translated by Annette Lavers. Rev. Vintage ed. Vintage Classics. London: Vintage Classics.

Barthes, Roland, and Stephen Heath.
1977. *Image, Music, Text: Essays*. 13. [Dr.]. London: Fontana.

Baudrillard, Jean.
1994. *Simulacra and Simulation*. The Body, in Theory. Ann Arbor: University of Michigan Press.

Bennet, Jane.
2010. 'A Vitalist Stopoover on the Way to New Capitalism'. In *New Materialisms: Ontology, Agency, and Politics*, edited by Diana Coole and Samantha Frost. Duke University Press.

Berger, John.
1997. *Ways of Seeing: Based on the BBC Television Series with John Berger; a Book Made*. 37. pr., 1. publ. 1972 by British Broadcasting Corp. and 1977 by Penguin Books. London: British Broadcasting Corp.

Boehm, Gottfried.
2022. *The Passion of Images*. Walther Koenig. https://www.amazon.co.uk/Gottfried-Boehm-Passion-Carolin-Meister/dp/3960987722.

Burnham, Jack W.
1987. *Beyond Modern Sculpture: The Effects of Science and Technology on the Sculpture of This Century*. 7. print. New York: Braziller.

Cadava, Eduardo.
1997. *Words of Light: Theses on the Photography of History*. Princeton, NJ: Princeton Univ. Press.

Craig, Cairns.
2007. *Associationism and the Literary Imagination: From the Phantasmal Chaos*. Edinburgh University Press. https://doi.org/10.3366/edinburgh/9780748609123.001.0001.

Didi-Huberman, Georges.
2005. *Confronting Images: Questioning the Ends of a Certain History of Art*. University Park, Pa.: Pennsylvania State University Press.

Eco, Umberto.
1989. *The Open Work*. Cambridge, Mass: Harvard University Press.

Elwes, Jake.
2017. *Closed Loop*. Jake Elwes' Website. Available online: https://www.jakeelwes.com/project-closedLoop.html

Flusser, Vilém.
2000. *Towards a Philosophy of Photography*, London: Reaktion Books,

Foucault, Michel.
1993. 'About the Beginning of the Hermeneutics of the Self: Two Lectures at Dartmouth'. *Political Theory* 21 (2): 198–227.

Fowler, Caroline O.
2016. *Drawing and the Senses in Early Modern History*. Turnhout, Brepols.

Galanter, Philip.
2019. *Artificial Intelligence and Problems in Generative Art Theory*. https://doi.org/10.14236/ewic/EVA2019.22.

Gebauer, Gunter, and Christoph Wulf.
1996. *Mimesis: Culture—Art—Society*. Translated by Don Reneau.

Goodman, Nelson.
1976. *Languages of Art: An Approach to a Theory of Symbols*. 2nd edition. Indianapolis, Ind.: Hackett Publishing Co, Inc.

Goriunova, Olga.
2023. 'The Automation of Creation: From Template Art to AI'. In *Technological Futures: Emerging Curatorial, Exhibitionary and Institutional Practices across AI and VR*, edited by Joasia Krysa and M Tyzlik-Carver. Vol. Data Browser V. Open Humanities Press.

Hertzmann, Aaron.
2022. 'When Machines Change Art'. Aaron Hertzmann's Blog. 17 December 2022.

Hertzmann, Aaron.
2023. 'What Is Creativity? Can Computers Be Creative?' Aaron Hertzmann's Blog. 27 September 2023.

Hui, Yuk.
2016. *On the Existence of Digital Objects*. Electronic Mediations 48. Minneapolis: University of Minnesota Press.

Kosuth, Joseph, and Gabriele Guercio.
1993. *Art after Philosophy and after: Collected Writings, 1966 - 1990*. Cambridge: The MIT Press.

Krauss, Rosalind.
1979. 'Sculpture in the Expanded Field'. *October 8*: 31–44. https://doi.org/10.2307/778224.

Lewitt, Sol.
1967. 'Paragraphs on Conceptual Art'. *Artforum*, 1967. https://www.artforum.com/print/196706/paragraphs-on-conceptual-art-36719.

Manovich, Lev.
2023. 'Seven Arguments about AI Images and Generative Media (Chapter 5 of Artificial Aesthetics'. *Artificial Aesthetics*, January.

Martinez, Gonzalo, Lauren Watson, Pedro Reviriego, José Alberto Hernández, Marc Juarez, and Rik Sarkar.
2023. 'Towards Understanding the Interplay of Generative Artificial Intelligence and the Internet'. arXiv. https://doi.org/10.48550/arXiv.2306.06130.

Massumi, Brian.
2019. *Architectures of the Unforeseen: Essays in the Occurrent Arts*. U of Minnesota Press.

Mitchell, W. J. T.
1984. 'What Is an Image?' *New Literary History* 15 (3): 503–37. https://doi.org/10.2307/468718.

Offert, Fabian.
2019. 'The Past, Present, and Future of AI Art'. *The Gradient*, 18 June 2019.

Papatheodorou, Theodoros, and Jack DiLaura.
2023. *Visual Dialogues: Doodles That Spark Conversations between Deep Learning Networks'*. https://doi.org/10.34626/XCOAX.2023.11TH.349.

Potolsky, Matthew.
2006. *Mimesis*. The New Critical Idiom. New York ; London: Routledge.

Raulet, Gérard.
1983. 'Structuralism and Post-Structuralism: An Interview with Michel Foucault'. *Telos* 1983 (55): 195–211.

Ruskin, John.
1903. *The Complete Works of John Ruskin*. Edited by E. T. COOK and Alexander WEDDERBURN. London: George Allen.

Salles, Arleen, Kathinka Evers, and Michele Farisco.
2020. 'Anthropomorphism in AI'. *AJOB Neuroscience* 11 (2): 88–95.

Sarin, Helena.
2018. 'Playing a Game of GANstruction'. *The Gradient*, 13 September 2018.

Shumailov, Ilia, Zakhar Shumaylov, Yiren Zhao, Yarin Gal, Nicolas Papernot, and Ross Anderson.
2023. 'The Curse of Recursion: Training on Generated Data Makes Models Forget'. arXiv.

Stiegler, Bernard.
1998. *The Fault of Epimetheus*. Technics and Time / Stiegler, Bernhard 1. Stanford, Calif: Stanford Univ. Pr.

UNI_VERSE.
2023. *Recursive Reflections*. Online. https://www.universestudio.xyz/work/uni-verse-recursive-reflections/

Zylinska, Joanna.
2020. *AI Art: Machine Visions and Warped Dreams*. Open Humanites Press.