



Signata

Annales des sémiotiques / Annals of Semiotics

17 | 2026

Sémiotique de l'intelligence artificielle générative

The Compositional Rhetoric of Multimodal Generative AI

About Tropes and Stereotypes

La rhétorique compositionnelle de l'IA générative multimodale. Sur les tropes et les stéréotypes

Maria Giulia Dondero



Electronic version

URL: <https://journals.openedition.org/signata/6354>

DOI: 10.4000/16qma

ISSN: 2565-7097

Publisher

Presses universitaires de Liège (PULg)

Printed version

ISBN: 978-2-87562-524-3

ISSN: 2032-9806

Provided by Université de Liège



Electronic reference

Maria Giulia Dondero, "The Compositional Rhetoric of Multimodal Generative AI", *Signata* [Online], 17 | 2026, Online since 15 June 2026, connection on 10 September 2026. URL: <http://journals.openedition.org/signata/6354> ; DOI: <https://doi.org/10.4000/16qma>

This text was automatically generated on September 7, 2026.



The text only may be used under licence CC BY 4.0. All other elements (illustrations, imported files) may be subject to specific use terms.

The Compositional Rhetoric of Multimodal Generative AI

About Tropes and Stereotypes

La rhétorique compositionnelle de l'IA générative multimodale. Sur les tropes et les stéréotypes

Maria Giulia Dondero

AUTHOR'S NOTE

I wish to express my sincere gratitude to Adrien Delière for his valuable collaboration in my work on GenAI. I'm also deeply indebted to Constantine Nakassis for his corrections and suggestions on this text (presented first at a jointly sponsored session of the Mass Culture and Semiotics workshops of the University of Chicago in December 2025) and to Pierluigi Basso Fossali and Enzo D'Armenio for their insights on my texts on GenAI visual compositions. I would also like to thank the two anonymous reviewers who helped me to better structure the present article.

Introduction

- 1 This article addresses two characteristics of the enunciative rhetoric of multimodal Generative AI (GenAI): first, the difficulty of generative models in producing visual tropes that deviate from the visual composition rules learned during training; and, second, the banalization and typification of styles of historical painting when reproduced by GenAI models.
- 2 More precisely, the first part of my article tests the way in which AI models operate in representing the human face through prompts that ask for the negation, subtraction, addition, or substitution of some fundamental parts of the face such as the nose. These tests have to be understood in the frame of the visual rhetoric of Liegian μ Group

(1992), that describes visual tropes as unexpected features present in visual products and, more particularly, as the gap between a visual representation (qua perceptible evidence) and the expected, pre-visible result (qua visual prototypes). If this first part focuses on the relation between the mereology of the visual figure and the rhetoric of GenAI with regards to prototypes, the second part explores how GenAI brings forth stereotypes about recognized artistic painterly styles (e.g., Van Gogh's). In this context, by starting from the familiar forms of historical styles in the Western artistic tradition, it will be possible to estimate the level of stereotyping (and innovation) of the forms produced by GenAI.

- 3 For each prompt related to a painting style, I examine the plastic variations among the generated images in order to describe what I call a “regional style,” meaning the set of visual patterns associated with a particular region of the database activated by the prompt. The regional style is distinguished from the “global style” unique to each model. Indeed, Midjourney, DALL·E, and other models—each built on different datasets and systems—impart a distinct overall style to all generated images.
- 4 The images I generate for these two purposes are not created with aesthetic intent; rather, they hold a scientific status. Here, images function as experimental tools produced to investigate, first, the compositional rules governing images in generative models and, second, the way in which the historical styles are absorbed and transformed by those very models. In addition, my focus on the formal composition in painterly styles is also oriented by a pedagogical objective. This research aims to enable users of such AI models to recognize stylistic differences, which are determined mainly by textures and rendering in specific compositional structures. It is crucial to study the impact of these generative models and their visual outputs on our perceptual culture, as well as on our judgments and aesthetic appreciations.

1. The Rhetoric of faces in GenAI (negation, subtraction, substitution)

- 5 Let us begin with my tests on images of faces.
- 6 On the topic of the artificial way of composing an image, a number of authors have theorized the “artificial perception” and “delegated perception” of GenAI models.¹ But the rules of composition of these models have not been explained nor analyzed.²
- 7 After thousands of experiments with Midjourney and DALL·E, I can state that the so-called “delegated perception”, or “distributed perception” of multimodal generative models, can be conceived of as a sort of hybrid enunciative process composed by various instances of discourse³ that are intertwined:
 1. The first instance is the *compositional rules* that have been stabilized through the model's learning from the training data. As we will see, these compositional rules, thanks to which the generated images display a certain perceptual coherence, may in certain cases *struggle with prompts*.
 2. The second enunciative instance can be identified in the *distribution* process through which the generative model mobilizes database images associated with the prompt via the annotations present in its training datasets.
 3. Finally, there is an *aleatoric* dimension which lies at the basis of every act of image generation. Such an aleatoric dimension is also part of the training phase of the diffusion

model (which works through stochastic processes of noise addition and subtraction) and of the “translation” of the original prompt (given by a user) into the revised prompt (generated by the GenAI, e.g., by reformulating the user prompt via ChatGPT)—and subsequently into the image produced.

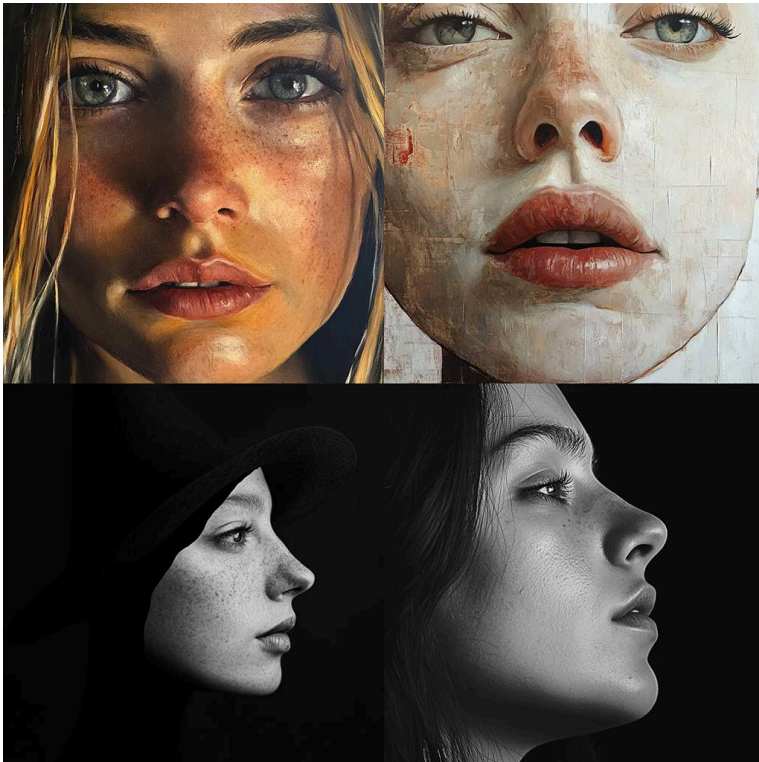
- 8 My tests focus on what may be considered as obstacles to this enunciative process, that is, operations of negation, suppression, and subtraction.
- 9 In my past research (Dondero 2015, 2020, 2023, Badir and Dondero 2016), I realized that negation is an absolutely crucial question in understanding the “language” of images and their metalanguage.⁴ Therefore, when GenAI became available, I decided to test how negation functions in the realm of GenAI models. Our first tests in 2023 (D’Armenio, Delière, Dondero 2024) asked Midjourney to remove the sun from a Van Gogh’s landscape. However, the model did not comply, as it was unable to grasp the connection between the word “without” and the word that followed. Every word in the prompt had to be concretized by the GenAI in the output. We finally realized that objects can only be removed from the composition using a special command, “minus minus (--).” In a further article (D’Armenio, Dondero, Delière, Sarti 2025), we explored other ways of negating or removing objects from compositions through “negative” prompts. For instance, we asked Midjourney to generate a dog without a tail, and the model positioned the dog in a way that makes the tail clearly visible. However, using the specific function “--”, which allows one to modify a generated image by excluding specified elements, we achieved, in the case of the dog, a different result from the Van Gogh landscape test described in D’Armenio, Delière, Dondero (2024): Midjourney performed a rather “clever” operation by positioning the dog so that the tail is placed out of frame. So, the solution “chosen” by Midjourney makes it impossible for us to verify whether the dog is itself meant to have a tail or not, revealing perhaps how the model translated the prompt epistemically *with respect to the viewing subject* rather than ontically in the diegetic object represented.
- 10 The successively tested cases concern the face. Midjourney’s inability to generate faces without eyes, without a nose (or other attributes of the head) or to generate portraits without a head, is clear when looking at the following images (Figures 1, 2, 3). It is worth noting that we didn’t use the word “without” but rather the suffix “-less,” because our previous experiments had shown that the term “without” has no negation effect.

Figure 1.



Midjourney. Prompt: "Eyeless portrait", September 2024.

Figure 2.



Midjourney. Prompt: "Noseless portrait", September 2024.

Figure 3.



Midjourney. Prompt: "Headless portrait", September 2024.

- 11 In the case of the prompt "Eyeless portrait", the face is rendered frontally, and the eyes are covered with a blindfold. In the case of a "Noseless portrait," the nose is prominently emphasized. For a prompt such as "Headless portrait," the result is a face covered with various layers and portrayed more like a lifeless sculpture.
- 12 It is important to note that when prompts produce aestheticizing semantic traits (sculpture, statue), we can infer that the prompt has gone against the established rules of composition. But let us look at other examples to better understand this.
- 13 I also tested the outputs of DALL·E. It is similarly unable to generate a face without a nose. This suggests that some figurative configurations are more resistant than others to local manipulations prompted at a lexical level. In such cases, the compositional process of the model does not follow the prompt, but instead seems to treat certain entities (such as the face) as indivisible, as parts of a whole that must be reused when generating new images.
- 14 Moreover, I observed that when we used a prompt such as "A hairless portrait," the hair disappears. Yet, not with the nose. Is a face without a nose perhaps too abstract of a concept for the model? Or is it too far removed from the rules that trained the model? Or is this face without a nose too close to semantic regions associated with violence, pain, and war, that belong to censored topics?⁵
- 15 Here is the result for "A Noseless face" in DALL·E (Figure 4).

Figure 4.



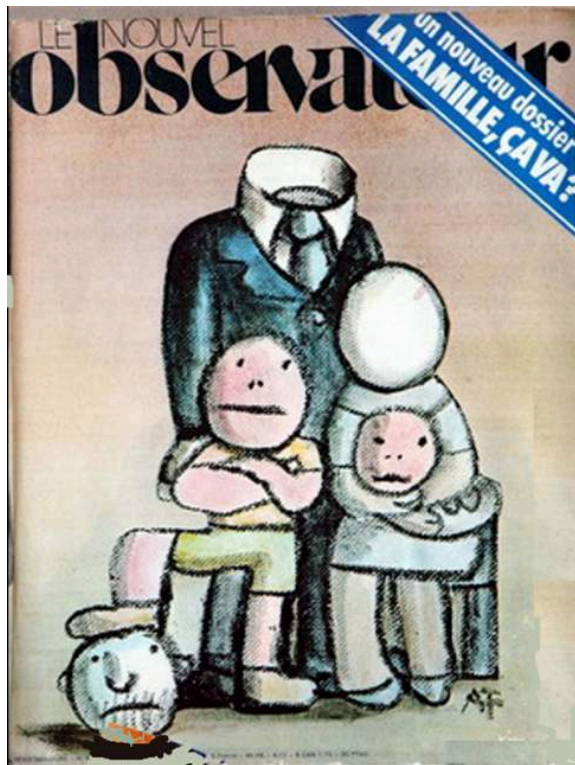
DALL·E. Prompt: "A noseless face". October 2024.

- 16 In October 2024, after a number of tests, I asked ChatGPT for clarifications about its visualization tool DALL·E in order to understand how it itself presented this representational behavior, and then compare the latter with the outputs from DALL·E and Midjourney. My question asked why generative models like Midjourney and DALL·E seem incapable of generating a noseless face. ChatGPT provided a very detailed explanation related to the training data and the neural architecture behind generative models. First, it recalled the fact that models are trained on massive datasets of images from the internet, which typically feature "complete" faces with noses, eyes, and other *essential* features. As the models learn to "expect" a nose when generating human faces, it makes it challenging to produce an image without a nose because such prompts conflict with the standard facial geometry the model has learned. Second, it states that in such models, human faces are represented in a compressed form in the latent space.⁶ Removing a prominent feature like the nose might push the generated image to a point in latent space that doesn't align with any plausible representation of a human face. The misalignment between embeddings highlights this difficulty in generating images that are not already cataloged within models' compositional schematizations. This can result in the model either *ignoring the instruction* or creating unnatural, distorted images. ChatGPT concludes that "Generative models excel at creating variations within learned boundaries, but extreme deviations from the norm (such as a face with no nose) might not generalize well. Since the model has seen millions of images with noses but almost none without, it lacks the capability to *confidently* generate a noseless face" (my emphasis).
- 17 In conclusion, according to what ChatGPT declares, the inability to generate noseless faces likely arises from a combination of the biases in training data (the noseless faces

are very rare and likely associated with censored semantic zones), the learned structure of faces, and the challenges of deviating too far from the norm during generation. This would imply that only very common or prototypical figures can be represented. However, we know that GenAI can also generate uncanny monsters, assemblages of parts and figures that do not prototypically (at least within the datasets) belong to the same subject. Indeed, when prompted with fantasy objects—such as a “squirrel wearing clothes”—DALL-E is able to generate corresponding images.

- 18 Faced with this apparent contradiction in the generation process, ChatGPT, questioned again, distinguishes between generating a “squirrel wearing clothes” as a *creative variation of familiar concepts* and a “human face without a nose” as a *violation of fundamental structural expectations* and as “a disruption of a basic, deeply embedded understanding of how faces are constructed”.
- 19 In fact, ChatGPT states that in generating squirrels wearing clothes, the model is combining two familiar and distinct (but also, I would add, associated, in specific zones of the datasets such as in children’s books) elements it has seen in training data: animals and clothing. It indeed seems to me that, for composing, it is easier to add or layer elements from distant categories than to subtract what otherwise appear to be defining characteristics. ChatGPT then proposes the distinction between symbolic and structural manipulation: if squirrel with clothes combines two ideas that don’t violate any deep structural rules (squirrels have bodies that can accommodate external objects like clothes), the noseless face represents a structural manipulation that disrupts the model’s understanding of what a human face is.
- 20 This issue of combining and subtracting touches upon two issues: the figurative prototype and the mereology of objects—the latter concerning the relations among the parts that compose a whole, on the one hand, and between the parts and the whole, on the other. Prototype and mereology, according to a certain semiotic approach, become rhetorical issues, since deviations, displacements, suppressions, or recombinations of expected traits of a type are at stake.
- 21 When this object is constituted by all its fundamental parts, it represents a perfect token of that type (of object). Conversely, when its constituent parts are separated, subtracted, or displaced, the resulting imperfect object can be understood as an error—or as a *trope*.
- 22 In this sense, I see a significant connection between figurative semiotics from a cognitive perspective—in particular, the visual rhetoric of the μ Group (1992)—and the functioning of GenAI. This semiotic approach relies on figurative types composed of more or less fundamental attributes (using various gradients of typicality of the prototype) that make these types more or less recognizable (more or less iconic or abstract) and that represent a reference for recognizing and including or excluding further occurrences. These types foremost provide a framework with respect to various rhetorical deviations such as those in figure 5, which are frequently used in artistic, political, and advertising discourses in order to call attention to specific topics addressed through a disrupting perspective.

Figure 5.



"Rhetorics of the type. Cover for *Le Nouvel Observateur*. A. François". From Groupe μ , *Traité du signe visuel. Pour une rhétorique de l'image*. Paris, Seuil, 1992.

- 23 Artistic and poetic discourses use these types of deviations from a category prototype to unsettle the observers' expectations and make them enlarge the territory of their sensibility.
- 24 In the framework stemming from the visual rhetorics of the μ Group, violations such as subtraction and substitution of objects' parts can be seen as more or less rhetorical (and meaningful) and as more or less aesthetically efficient deviations from the established types (such as a face). On the contrary, multimodal GenAI training of the sorts I have experimented on do not seem to have contemplated rhetorical effects via deviations from what is expected to be "correct" by users (i.e., a type represented by all its attributes).⁷
- 25 In November 2025, I added to the initial prompt of the noseless face some adjectives such as "fictional" or "fantastic" but the results did not truly change. The nose and the ears were strongly exaggerated, as these features were processed together. More recently (March 2026), I prompted ChatGPT again, and the result I obtained for the prompt "Represent a noseless face" is more interesting (Figure 6). It seems that the model is capable, if not of stepping away entirely from a prototypical schema of the human face, at least of envisaging a different gradient of typicality—namely, the blurred nose, that is a sort of position at the periphery of the category of a prototypical human face. In a sense, the nose remains in its expected position on the face, but its form is not fully completed and becomes blurred toward its tip.

Figure 6.



DALL·E. Prompt: "Represent a noseless face." March 2026.

- 26 If these models have been trained with the aim of giving users the most probabilistic and so, from our point of view, most common and ordinary results, they are also being trained to represent, if not unstructured, at least vague and more visually uncertain features when negation is elicited by the prompt.

1.1. Operations of substitution and addition

- 27 I also tested cases involving the substitution of the nose or addition of features such as colors to the face. For example, the operation of replacing the nose with something else (thus applying a symbolic rather than structural manipulation following the distinction formulated by ChatGPT), made me produce some prompts introducing animal features such as "A face with <'a beak', 'a snout', 'a muzzle'> instead of a nose." The results were quite coherent: there is a certain capacity to smoothly integrate these elements into the structure of faces.
- 28 Next, I tested with less obvious attributes for a nose, such as "a laptop," "a car," and "two noses." The results were immediately much weaker, as the model, in the first two cases, simply superposed a laptop or a car onto an existing nose (so "on a nose" rather than "instead of a nose").
- 29 In the case of "two noses" (Figure 7), the result shows a kind of "nasalization" of the face, where not only the nose but also the lips and cheeks undergo a process of transformation into nose-like features.

Figure 7.



DALL·E. Prompt: "A face with two noses", October 2024.

- 30 It appears that a prompt that is too exotic leads to a simple superposition over the nose or to a multiplication and nasalization of other parts constituting the face.
- 31 Finally, I had to ensure that it wasn't the structure of the prompt itself that was preventing us from manipulating the nose as I intended. So, I tested with "nothing" in place of the objects I previously tried to use to replace the nose. When using a prompt such as "A face with nothing instead of a nose," (Figure 8) the resulting image suggests that a certain subtraction had been generalized: it is as though parts of the face, particularly the head, cheeks, and partly the neck, are rendered less prominent or less intense through a specific texture and lighting. Indeed, the "negated" parts are clearer than the parts depicting the typical color of the skin of the face. This indicates that while the model can attempt to reduce prominence of certain areas of the face, it struggles to entirely exclude the feature "nose" without triggering broader compositional adjustments.

Figure 8.



DALL·E. Prompt: "A face with nothing instead of a nose," October 2024.

- 32 The model is indeed capable of making modifications that aren't too extreme and, in a sense, "hybrid," striking a balance between familiar variation and structural violation. I tested this option by mentioning the presence of a nose, but with an unusual color: white, mauve, pink, black, red, yellow, blue, and green.
- 33 Let us take a close look at the pink result (Figure 9). This color seems to stay within the realm of believability, demonstrating how the model can navigate slight variations in appearance while adhering to more familiar compositional norms. The pink color of the nose is accompanied by rosy cheeks and almost unnaturally closed eyes and mouth, resembling those of a doll (the texture simulates the materiality of inanimate plastic objects). So, we can state that the pink color recalls the dolls' universe in GenAI training. This is shown by the way that the GenAI does not limit itself to coloring the nose, but shifts toward depicting activities involving non-human objects that normally belong to the world of childhood: are we facing a passage from a color to a stereotyped association of this color with the face?

Figure 9.



DALL·E. Prompt: "A face with a pink nose", October 2024.

- 34 The more "exotic" colors, like black and yellow (Figures 10, 11), resulted in modifications to other parts of the face, as if there had to be a correlation or coherence to justify the fact that the nose is black or yellow.

Figure 10.



DALL·E. Prompt: "A face with a black nose", October 2024.

Figure 11.



DALL·E. Prompt: "A face with a yellow nose", October 2024.

- 35 The black nose coincides with full lips and rather thick eyebrows, while the yellow nose seems to spread, affecting the ears and lips. In the latter, the forehead disappears, and the overall effect for the observer is more like seeing a mask. Once again, as in the “Headless portrait” test through Midjourney, the result evokes a reference to the art world, suggesting a form of dehumanization of the face—a world that seems characterized by stillness.
- 36 When contrasting the two faces in terms of the features that are correlated with them, one notices that the black nose emphasizes more pronounced, insistent features, whereas the yellow nose brings out softer, more angelic, perhaps even androgynous ones. Are we faced with stereotypes of black and yellow colors related to racial features?
- 37 Let us take a closer look at the purple nose (Figure 12).

Figure 12.



DALL·E. Prompt: “A face with a purple nose”, October 2024.

- 38 The face is distorted and appears lifeless, as if it were once again a lifeless mask displayed in a museum, or a sculpture with hollow eyes. But the most interesting to me is the blue nose (Figure 13).

Figure 13.



DALL·E. Prompt: "A face with a blue nose", October 2024.

- 39 The blue color of the nose transforms the face into a carnival mask. This result is interesting for understanding the semantic traits that the model activates when a request falls outside of the densest parts of the training database of the model. What we can suggest here is that the blueness of the nose taps into a semantic zone of the database related to carnival masks and costumes. At first sight, this result may appear surprising, but on closer inspection it reveals a deeply ingrained stereotype in Western cultures: the manipulation of forms, and especially colors, on the face typically occurs during the carnival period, when ordinary values are subverted. This output constitutes a tipping point, since we move from a model behavior that produces small variations of the face at the level of the expression plane—such as in the case of yellow and black coloration—to a behavior that may appear as a kind of “taking the initiative” at the level of the content plane, directing the output toward a very specific semantic region. We thus move from slight variations on the prototype to a more dramatic shift toward discrete semantic regions, such as the carnivalesque. More tests have confirmed this statement. The machine’s “certainty” (its degree of confidence in the correctness of the representation proposed) corresponds to situations that, from the perspective of our Western culture, we might perceive as stereotypical, such as a blue nose in carnivalesque attire.⁸
- 40 These investigations on the face and especially the ones recalling negation procedures and colors that are not present in the training dataset reveal that, in the context of diffusion models, elements on the plane of expression are not separated from a plane of content (inanimate artistic objects or carnival masks).
- 41 These multimodal GenAI models, which are being built on the most recurrent practices present in recorded texts of human experience (and the metadata codings provided by

human labor in training the model), can be studied in the framework of what the semiotician Umberto Eco termed the “encyclopedia” (Eco 1975). But the variety inherent in the encyclopedia (and the inner contradictions it contains) is sacrificed by these models in favor of a hierarchy of attributes constituted by what is most probable in the database’s translation of human experience (such as the common structure of a face).

- 42 Research is ongoing to determine in which situations prompts involving intersemiotic translation are able to overcome the “resistances” of the models. The overlap of resistances to the rules learned during the training phase would mean that models can behave in ways that are not purely probabilistic. It also means that models would become able to personalize and canalize their outputs according to the semiotic register (artistic, ironic, scientific, etc.) more or less explicitly mentioned by the prompter.⁹ The linguistic register could possibly be “translated” in the terms of the model’s weights, that are the numerical parameters that determine how the model processes and transforms input data to produce an output. Indeed, during training, the model learns by adjusting these weights to reduce the error between the predicted output and the desired output.¹⁰ The so-called “temperature” is a parameter that controls the randomness of a model’s predictions by adjusting how the system of weights influences output probabilities, thereby affecting the creativity of the text: lower temperatures produce safer, more coherent text, while higher temperatures encourage novelty and variation. Acting on the weights and temperature makes it possible to destabilize the probabilistic order that the models we use by default rely on.

2. Stylistic stereotypes produced by GenAI

- 43 If, in the first part of this article, I tested the resistance of prototypes to edge cases, in what follows I examine the model’s ability to reproduce and modulate unique artistic visual styles. While a face is a sort of semantic unity, here I am interested in a stylistic unit, a unit produced by an enunciative process, the latter being recognized, in the domain of painterly art, as a sort of perceptual whole. From one point of view (e.g., from within the domain of art criticism), a painter’s style *a priori* offers an inimitable perspective on the world which mostly depends on their perceptual experience and sensor-motricity. In a sense, this style constitutes a specific kind of unity, insofar as it can be understood as a world in itself—that is, a particular way of seeing, an optic distinct from others. Indeed, each painterly style is composed of a syntagmatic set of traits that makes it recognizable as such; a non-coherent trait, or a non-homogeneous syntax of these traits, would render this recognizability questionable. It is this recognizability that is at stake in the following pages, where the unity of style is, as we will see, tested and ultimately “undermined” by the imitation game in which the model engages. How will this style, intended as a syntagmatic unity, resist the model’s work on its units? And, conversely, what does the model reveal about its functioning in this imitation of specific dynamic wholes that are painterly styles?
- 44 While the current literature on AI-generated stereotypes predominantly focuses on cultural stereotypes, particularly those related to gender and race, stylistic stereotypes—which are the focus of my ongoing research¹¹—are, in my view, more challenging to describe. Unlike racial and gender stereotypes, many of which operate within the domain of *recognizable visual forms* that can be relatively easily lexicalized (i.e., they are

part of figurative semiotics such as the fundamental elements composing a face), stylistic stereotypes are embedded in the *formal composition of the image* (i.e., are part of plastic semiotics) and they are not, or not exclusively, oriented in a way conducive to the representation of easily recognizable objects. On the contrary, they manifest *within* bidimensional images, as ensembles of traits which form, for example, the density of marks constituting a textural zone.¹² They manifest through elements such as the contrastive distribution of light, perspective, point of view, framing, and other compositional features.

- 45 The literature on multimodal GenAI abounds with studies in which researchers seek to develop methods for constraining algorithmic processes so as to maximize personalization and the uniqueness of outputs through the fine-tuning of databases. In contrast, the tests I performed with these models were not guided by a pursuit of beauty or originality. Rather, I deliberately allowed the machine the greatest possible freedom to operate according to the most common and statistically probable modes of generation. In practice, when I request a multimodal translation from natural language to a visual style, I deliberately choose a highly recognizable style (e.g., Van Gogh's or Impressionism) and pair it with a very general theme or genre (such as a landscape or a portrait). The aim is to understand how the machine operates in itself, in the most autonomous way possible, while working on the most familiar and straightforward styles and themes.
- 46 The advantage of prompting highly recognizable styles is that *it becomes possible to measure the distance between the original sensory-motor characterization of each painter's hand* (as revealed in the stylistic features of the artists' paintings) *and the forms* that are generated through the processes of annotation, database construction, algorithmic work, and randomness.¹³ Moreover, it is important to test painterly styles in order to examine the differences between the original forms and their transformation in artificial remakes. This allows us to estimate the level of stereotypy and creativity exhibited by each generative model.
- 47 After extensive testing, it can be stated that every stylistic element associated with a particular painter or artist in the prompt—such as chromatic tendencies or the rhythmic repetition of brushstrokes—is transformed into a caricature of sorts in the output. This occurs because individual features are distributed across numerous examples, resulting in a paradoxical visual effect; that is, *the process of averaging and distribution produces a selection and an exaggeration of certain stylistic features*. The outcome is invariably kitsch, in the sense of reducing nuances to simplified and standardized visual compositions. Of course, this judgment about the standardization and stereotypification of forms produced by GenAI depends on the ideological assumption that the original forms are artistic and aesthetically relevant, that is, forms that are exemplary of a specific combination of features and specific conceptions of space, light, color, et cetera and that these make a painter's overall style unique and non-repeatable.
- 48 If using historical styles serves as a foundation for observing the gaze of the machine, it is, as we will see, also useful for studying the stereotypy of historical styles themselves, which can necessarily be defined as something recognizable and, I would dare say, typified.
- 49 Let us examine a collage of 50 outputs of the prompt "A Landscape in Van Gogh's style" (Figure 14). This number is sufficient to evaluate patterns of repetition and variation.

Figure 14.



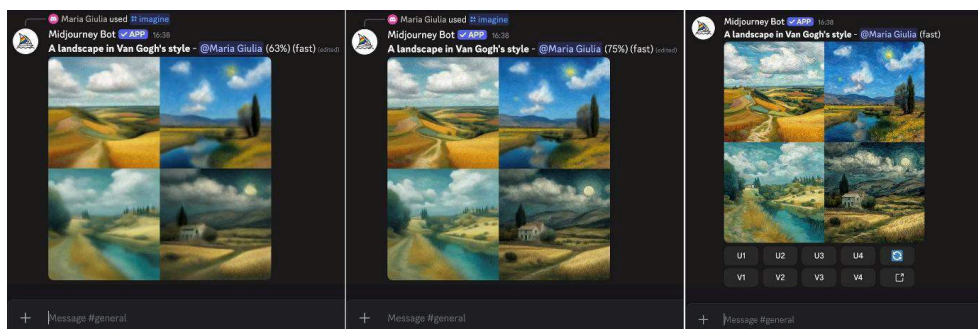
A. Delière and Midjourney. Collage of 50 outputs for the prompt "A Landscape in Van Gogh's style".

- 50 As we can observe from this collage obtained through the same prompt, the interesting result is that all the pictures generated are similar but slightly different. These slight differences show that each set of images produced from a prompt is more relevant as a *sampling of the region of the dataset* than as isolated images. In other words: the images produced by the generative models count more as extractions of features typical of a region of the dataset than as images per se. (Put otherwise, only considering an isolated image is insufficient to scientifically study the workings of such models.) This is why I call such ensembles of extracted features resulting from a region of a database a "regional style".¹⁴ A region of a dataset encompasses a series of features produced by the work of algorithms on certain domains of the dataset. Such domains are associated with annotations referencing the style that is mentioned in the prompt, as well as which is operationalized by the correspondences of embeddings (the embedding of the prompt and the embedding of the notation of visual forms). Moreover, the slight variations among the 50 outputs result from the different degrees of randomness inherent in how the GenAI operates.
- 51 A first degree of randomness at the level of natural language exists in the case of a model like DALL·E, which is known to revise the prompts it processes. Indeed, the model that revises the prompt is itself a Large Language Model (LLM). As such, it functions like all LLMs—that is, probabilistically: it formulates the revised prompt based on the most probable tokens given the initial prompt. There is a kind of random selection within the token distribution, which means that the revised prompt is not always the same for a given initial prompt. Therefore, the prompt actually submitted to the generative model—the revised prompt—already contains an element of randomness that can influence the final image.
- 52 The second level is in the very way that models learn—whether in learning to create embeddings, to form text-image associations, or to generate images. This is because, during training, a model's set of parameters or weights is initially randomized. Different initializations can lead to different final models, and consequently to different embeddings and associations. Furthermore, training involves processing large datasets, which are often presented to the model in a randomized order. This

introduces additional variability into how the model learns and internalizes concepts. As a result, even identical inputs and computations can yield small differences in outcomes.

- 53 The third and possibly the most interesting level for us intervenes during the *initialization* of the generative process linked to the diffusion model's functioning. It is important to stress that LLM (for text) and diffusion models (for images) function differently. While in both cases the model (1) acts iteratively, (2) is guided by the instruction/prompt, and (3) produces something that is probabilistically related to what it has learned, a difference between LLMs and diffusion models is that to generate texts, the former model acts sequentially (from the beginning of the sentence to the end of the sentence, token after token), while to generate the image, the model acts “by increasing resolution.”¹⁵ Midjourney acts on the entire image at once, but first in a very blurry and general way, at low resolution, then gradually more precisely, less and less blurry, with increasing resolution/precision.¹⁶ The important difference, thus, is the axis along which the progressive aspect is done: “sequential” order for a sentence, that is, from left to right and top to bottom; “by sharpness, resolution” order for an image, from coarse to fine, from noised to denoised. It is possible to observe some progressive steps of the generation of four images by Midjourney in the visualizations below (Figures 15-17).

Figures 15-17.



Left: 63% of the generation process; Center: 75% of the generation process; Right: Process completed.

- 54 This composition procedure differs markedly from the token sequences used in LLMs, since the output space must be organized in a more mereological rather than purely linear manner.

2.1. The gadgetization of painterly styles

- 55 By conducting multiple tests on various well-known painters of the sort noted above, it became apparent that it is very easy for GenAI programs to reproduce Van Gogh's style, but very difficult for such models to reproduce Paul Klee's style. All the tests done with the latter name gave as results colorful, geometrical images resembling “naïve art” that were not at all recognizable, to my eyes at least, as belonging to Klee's style (see Figure 24).¹⁷ Why is that? Does this mean that Van Gogh's style is already, in itself, stereotyped, and for this reason, easy to be reproduced? Or is it the result of the enunciative praxis that Van Gogh's style has undergone, and which has made his style easily recognizable, namely, all those processes of gadgetization and commodification that have placed Van Gogh's images on keychains, advertisements, tee-shirts,

handbags, and the like?¹⁸ If the latter, it would be the “Van Gogh style,” and not his original style itself, that is responsible for the exaggeration and kitschification of its traits in GenAI platforms. This raises an important question about the process of stereotyping through the various layers and mediations of the original style.

- 56 To test my hypothesis, I submitted the same prompt to Stable Diffusion, which has a different dataset and training characteristics than Midjourney and DALL·E. The results are the following (Figure 18).

Figure 18.



A. Delière and Stable Diffusion. Collage of 50 outputs for the prompt “A Landscape in Van Gogh’s style”. September 2025.

- 57 This collage provides empirical evidence that Stable Diffusion produces even more typified forms than DALL·E and Midjourney. Van Gogh’s style is easily recognizable, but it is even further removed from the compositional variety of his landscapes. In this collage, the same composition is repeatedly reproduced, although slight rotations and displacements of elements are visible. But why is Stable Diffusion an even more typifying model? Is it because the dataset consists of images scraped from the internet and because the training process relies on a less curated and less controlled approach? If we examine a part of the dataset used to train Stable Diffusion (LAION_2B_Aesthetics)¹⁹ and search for the images whose textual description is the closest to “a landscape in Van Gogh’s style,” we find images that do not correspond to Van Gogh’s own production but rather to images belonging to the broader class of images that have been annotated with the tag “Van Gogh” (Figure 19).

Figure 19.



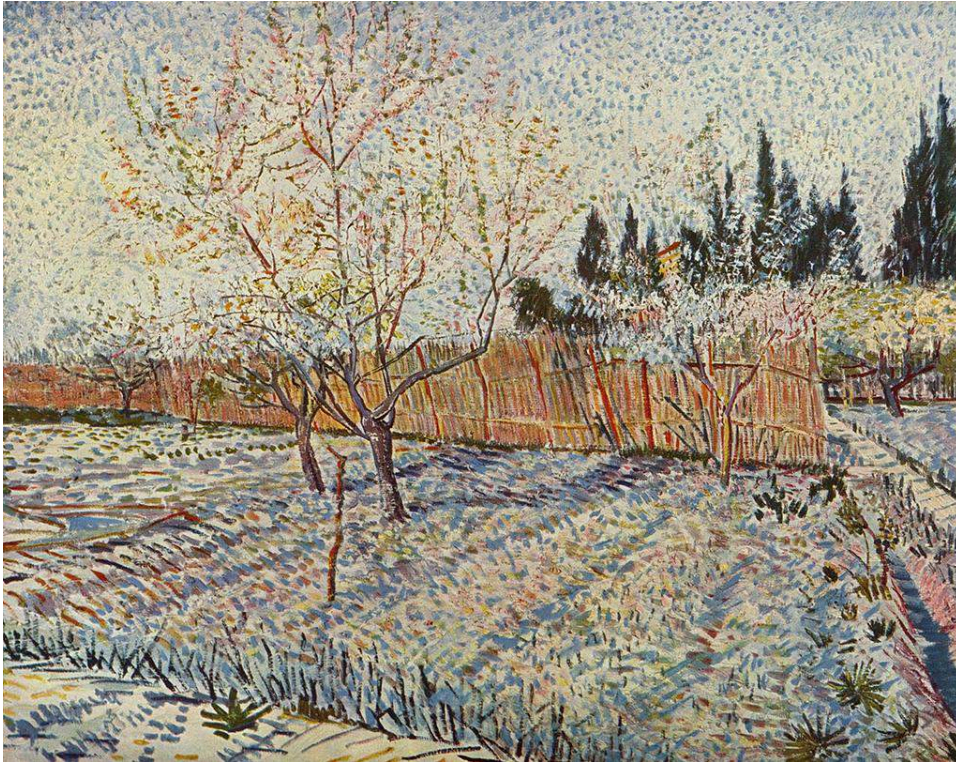
A. Deliège. Images of LAION-2B_Aesthetics retrieved as “closest” to the prompt “A Landscape in Van Gogh’s style”. Empty cells correspond to images no longer available at the time of conducting the retrieval. June 2025.

- 58 The fact that this (partial) set of images contains works that were not created by Van Gogh makes it clear that the images used to train Stable Diffusion also include remakes of Van Gogh’s works, as well as other imitations or distortions of his historical style (see, for instance, the penultimate image in the second row or the second image in the last row). This means that the typification of Van Gogh’s style through GenAI does not depend (or not depend solely) on its internal configurations, but also on the long trajectory these configurations have undergone over the past century through a multiplicity of reformulations.

2.2. Regional and global styles

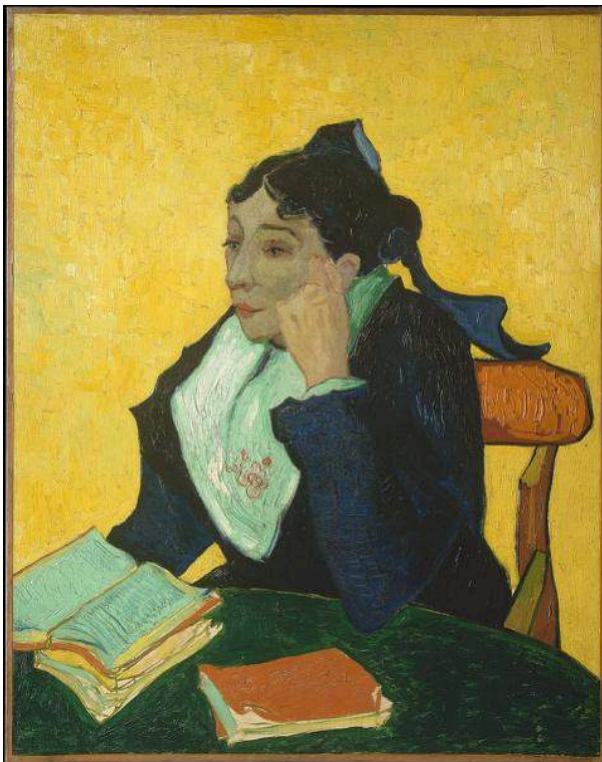
- 59 Before discussing the issue of the regional style, let us recognize that describing a style is, of course, a challenging task, as it traditionally concerns the variation of a microsyntax of marks that remains recognizable across different represented subjects, such as a landscape or a face.²⁰ To pursue this further, one might ask whether authorial syntax varies depending on whether the subject is a landscape or a human face. And do all landscapes produced by the same hand share the same syntax?
- 60 If we compare, for instance, *Orchard with Cypresses* (1888) and *Portrait of Madame Ginoux* (1890), both by Van Gogh, we can observe certain differences in the brushwork (very fragmented in the first painting and characterized by large areas of uniform color in the second) (Figures 20-21).

Figure 20.



V. Van Gogh, Orchard with Cypressess, 1888.

Figure 21.



V. Van Gogh, *Portrait of Madame Ginoux (L'Arlésienne)*, 1890.

- 61 In the case of artificial styles (from GenAI models), this question can be tested by repeatedly prompting, for example, “Van Gogh’s style” but specifying a subject other than a landscape—such as a portrait (Figure 22).

Figure 22.



Midjourney. Prompt: “A Portrait in Van Gogh’s Style,” 50 outputs. July 2025.

- 62 If we compare Figure 22 with Figure 14, we can identify the very same microsyntax of Van Gogh’s style in both the portrait and landscape tests. The model does not appear to be programmed to distinguish between the different forms of sensorimotor activity characterizing historical landscapes and portraits. It is true that some Van Gogh portraits mostly use very similar traits to those used in his landscapes. But in the case of GenAI outputs, the model averaged “Van Gogh” and renders landscapes *and* portraits in the very same way. Indeed, some portraits are even embedded within typical Van Gogh landscapes, reproducing the luminous syntax characteristic of his depictions of the sun and other celestial bodies. More generally, the kind of brushstroke remains consistent across both landscapes and portraits, and it is as if the style of especially one landscape, *The Starry Night* (1889), had been transferred onto faces—and especially onto Van Gogh’s self-portrait series. This suggests that the model has inferred Van Gogh’s *Starry Night* style (and possibly some other famous variations of this landscape) in a very precise way during its training phase ... but, perhaps, too precisely to be able to reproduce the internal differentiation of Van Gogh’s style? Indeed, the “Van Gogh style” encompasses only a narrow *range of variations* of his complex and diversified style.²¹
- 63 If we test the reproduction of other painters’ styles, we observe that, by repeating the same thematic prompt (e.g., “landscape” or “portrait”) *while varying* the names of painters and artistic movements (“in the style of Picasso,” “Klee,” “Gauguin,” etc.), we can discern *what remains stable and recognizable*, despite the variations introduced by the typified style of each painter (Figures 23–25).

Figures 23-25.



Prompts to Midjourney: left – “A Landscape in Picasso’s style”; center – “A Landscape in Paul Klee’s style”; right – “A Landscape in Gauguin’s style.” 2024.

- 64 Even though the three series of sets of four images are indeed different from one another, and though the styles of Picasso, Klee, and Gauguin can be partly recognized (albeit not definitively identified without reading the captions specifying the prompts), it can nonetheless be stated that certain features remain invariant across these generations.
- 65 Indeed, some stylistic traits are repeated across the three series, such as the colors employed: they are largely similar and seem attributable to *the generative model’s style* and in all three cases, the forms resemble those typical of naïve style painting. This “general” style is recognizable in all the generations produced from different prompts: I call this the “global style” of the generative model. This global style corresponds to the average, not of each painter’s annotated characteristics, but of the overall characteristics on which the machine has been fed and trained.
- 66 If regional styles show the declensions of local zones of the database, the global style encompasses all the paintings and more generally all the pictures that have been used to train the model, that is, the most popular styles of our time, made of photography and graphics filtered and inspired by animation, video games, East Asian comics, and so on.
- 67 Ultimately, the style of each model is, according to my view, identifiable *when we subtract* all the *specific* characteristics of the various regional styles invoked in our prompts. I propose that the generative model’s style is defined not solely by differences, as in the tradition of art history,²² but, more specifically, by subtraction.
- 68 The triple randomness (prompt revision, learning process, and initialization) characterizing diffusion models such as Midjourney makes us understand, firstly, the reason why all outputs stemming from a same prompt are always different from one another and, secondly, that they contribute to producing the range of variation constituting the regional style as a restrained stylistic stereotype featured by a specific density and syntax of traits.²³ The global style represents a different kind of stereotypy. It concerns what remains invariable despite regional style variations and includes the average of the totality of images through which the model has been trained, which is different from model to model. In this sense, in every output a typification of the style of a painter is associated with a stylistic typification of the general style of the model.

2.3. Style in the frame of Peirce's categories

- 69 This narrow range of variations of recurring features and objects in generative models' representation of historical styles may be understood in the light of two linguist anthropologists who address this problem in the terms of Charles Sanders Peirce. According to Eitan Wilf, generative models *do not only produce texts, they extract styles*. In his view, even if these generative systems:

much like a CD or an MP3 file, are technologies of reproduction, they do not reproduce specific texts, or [Peircean] Seconds, but styles, or [Peircean] Thirds. *Their object of reproduction is the principle of generativity* that is responsible for producing the specific texts that are the object of reproduction of various kinds of media technologies. These systems' reproduction of style consists both in their ability to *abstract a style* from a corpus of Seconds and to generate new and different Seconds or texts in this style, indefinitely so (2013:186, my emphasis).

- 70 Julien Porquet (2025) comments on Wilf's text and proposes an alternative account to Wilf's. In his view, the generative models, rather than reproducing Thirds as such, are more adept at extracting Firsts—abstract qualities such as textures, colors, and so forth and aligning them co-textually in the production of new images (this co-textual alignment being itself a Third, of course). Yet is this co-textual composition akin to the artist's style? Porquet argues not. Which is to say, *these chains of Firsts, even if they appear to emulate the style of certain artists, never exceed the sum of their parts*. In other words, the production of a sequence of visual qualities does not imply the capacity to directly constitute a style, that is Thirdness (a general matrix), but only a chain of Firsts (isolated qualities within the regional style of the model). However, it should be specified that Van Gogh's style remains recognizable: this suggests that the syntagmatic organization of these Firsts does not entirely depend on the recurrent *global* style of Midjourney; on the contrary, Midjourney appears to have learned Van Gogh's configuration of Firsts, such that a Third remains recognizable. This is not the case for other painters who are less represented in databases, such as Paul Klee: outputs prompted in his style are not clearly distinguishable from those of other painters (such as Paul Gauguin, for example). In this case, the Firsts do *not* compose a recognizable unity—i.e. Klee's Third—but rather a chain of Firsts structured according to a *global* stylistic syntax (the Third of the model itself).²⁴ The model does not seem to have successfully captured Klee's painterly specificities. But even in the case of "Van Gogh's style," the rendering remains somewhat odd, revealing a lack of completely successful extraction of the Van Gogh's. A confirmation of this assertion lies in the fact that, in Van Gogh-style productions, even though the pictorial texture is strongly suggested, the absence of patina in the hypersaturated colors and the clear separation between brushstrokes give the images a more photographic than pictorial appearance. In a sense, this photographic rendering (Third) of pictorial units (Firsts) may seem normal and predictable, since Midjourney was trained on more photographs than paintings. The effect is that the overall impression of these irregular painterly textures is typically photographic, creating a strange perceptual effect of translucent images imitating painterly opacity.²⁵ In other words, while the local effects (the isolated marks, as Firsts) are typical of Van Gogh's painting style, the overall effect—the Third—produces a hyperrealistic photographic effect that does not coincide with a complete implementation of Firsts in Thirds. (Secondness is not absent, but it does not seem to

function as a sufficiently structuring mediation for the constitution of style as a global consistency in texture.)

3. Comparison between historical styles and GenAI styles through image rating

- 71 Along with two mathematicians and experts in AI systems, Adrien Delière and Marc van Droogenbroeck, Jeanne Marlot and I decided to try to measure the perceptual distance between historical painterly styles and their reproduction by GenAI. The goal is to rate the faithfulness of the models' imitative operation in order to more finely characterize style through multiple parameters (in particular, color, texture, and geometry). More precisely, our rating protocol aimed to statistically evaluate the ability of Midjourney v6 to reproduce the stylistic traits associated with five major art movements (Impressionism, Cubism, Expressionism, Surrealism, and Abstract art) and five pairs of painters representative of these movements (one well-known male artist and one historically less-represented female artist). We then asked three specialists with an academic knowledge of art history and visual semiotics to evaluate both historical and AI-generated images using nine visual criteria operationalized from the formal categories of plastic semiotics (topological, eidetic, and chromatic) defined by Algirdas Julien Greimas (1989) and from the stylistic classical-baroque oppositions established by Heinrich Wölfflin (1915). These two authors proposed complementary approaches for describing how visual forms convey meaning independently of representational content. Their respective frameworks, although developed in different historical contexts, converge toward a common goal: understanding how compositional relations organize the perceptual unity of an image.²⁶
- 72 Each criterion selected for our study (see below) is expressed as a pair of oppositional poles defining a continuous 0–100 scale (e.g., centered forms [0] ↔ forms extending to the edges [100]) in order to avoid single values and to reflect the natural variability of stylistic traits within both historical paintings and AI-generated images.
- 73 For each movement and each criterion, the three experts performed six ratings: the historical artworks of the movement, those of the selected male and female painters, and the corresponding images generated by Midjourney using prompts referring to the same movement and painters.
- 74 The original image rating parameters that we used combined the plastic categories defined by A.J. Greimas and the formal criteria established by H. Wölfflin to distinguish classical and baroque visual modes. They were the following:
1. Are the forms (as concentrations of marks) centered (do the images have a center?) or are they spread out toward the edges of the image?
 2. Is the image laterally symmetrical (between left and right)?
 3. Is the image vertically symmetrical (between top and bottom)?
 4. Is the image sharp or blurry in terms of rendering?
 5. Are the lines mostly parallel or do they intersect?
 6. Is the rhythm of the lines dense or widely spaced?
 7. Are the colors saturated or desaturated?
 8. Is the light distributed evenly or concentrated on certain parts of the surface?

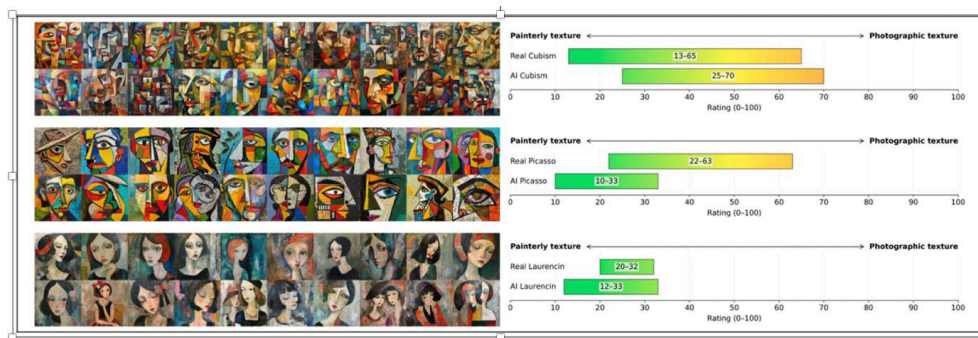
9. Is the simulated texture or grain of the surface (intended as the simulation of a volume or a repetition of traits building patterns and other sorts of regularities) closer to that of painting or photography?
- 75 These parameters were formalized by Adrien Delière. The resulting criteria are expressed as a continuum between two poles, emphasizing that visual properties rarely occur in absolute terms but rather occupy intermediate positions within a continuous perceptual spectrum. They are the following:
1. centered forms vs. edge-deployed forms
 2. left-right symmetry vs. left-right asymmetry
 3. top-bottom symmetry vs. top-bottom asymmetry
 4. sharp vs. blurred images
 5. parallel vs. intersecting lines
 6. dense vs. sparse rhythmic structures
 7. saturated vs. desaturated colors
 8. uniform vs. focalized lighting
 9. painterly vs. photographic texture.
- 76 The first section of the rating sheet, which concerns the first criterion—centered forms vs. edge-deployed forms—appears as follows (Figure 26).

Figure 26.

Adrien Delière. Example of ratings regarding centered-edge-deployed forms in abstract art.

- 77 Experts were instructed, for the historical reference and the generated images, to specify an interval between 0 and 100 of the known artworks representative of each parameter rated. In the two cases, experts were asked to take into account only most representative features and not the rare outliers.

Figure 27.

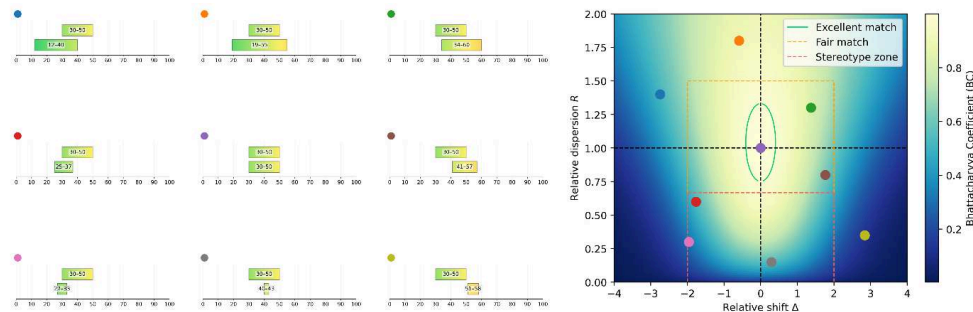


Adrien Delière. Example of the rating protocol for Cubism, on a single rating sheet. For each meta-image on the left, obtained from the prompts (top) “A Cubist painting”, (middle) “A painting in the style of Pablo Picasso”, and (bottom) “A painting in the style of Marie Laurencin”, the right-hand panels show the corresponding expert rating intervals on a 0-100 scale for the criterion “Painterly texture vs. Photographic texture”, for both historical paintings and AI-generated images.

- 78 Before beginning the rating process, we expressed the hypothesis that the rating operation of historical painterly production develops a kind of reasoning that could be described as following an “intensional” path, that is, a model highlighting the set of properties or characteristics that *define* what it denotes (Impressionism, Cubism, Picasso’s production, etc.) by relying on one’s own memory (the images are not perceptually present to the observer and therefore depend on the rater’s recall and expertise in art history). In this framework, the tokens lose their singularity, since only very general characteristics are considered relevant to constitute a type (i.e., of Cubist production).
- 79 We also formulated the hypothesis that in the rating of AI-generated images, we would have reasoned in another way, that is a more “extensional” way, because we evaluate AI-generated images one by one (20 per prompt) in their *observable specificity*. In this second case, singularities are taken into account; the generalized categories defining a type—where the tokens are only representative of the type—are replaced by singular and specific elements belonging to a class encompassing diversity.²⁷
- 80 Indeed, when we make the effort to recall all the tokens without focusing on any particular one (as in the case of all cubist paintings historically produced!), and to find what is invariant from one token to another, we are engaging in typification. By contrast, when we dispose of all the elements to be evaluated, they can conserve their diversity and extension as variations within a class.
- 81 When comparing the rating interval of an AI production with the interval of a reference production, two main quantities matter: the relative position of the center of the interval with respect to the reference one, and its relative spread. Figure 28 illustrates various example cases. In the case of (almost) perfectly identical intervals (purple), the AI production is an *excellent match* to the historical reference. When the interval is slightly shifted and/or slightly dilated (green) or compressed (brown), the AI production remains a *fair match* to the historical reference. A *(stereo)typification* occurs when the AI-related interval shrinks down to a very tight interval, indicating little variability in its production for the criterion examined. For a stereotype to be relevant, it still needs to remain reasonably shifted from the reference interval (red, pink, gray), as too much shift (yellow) indicates that the model captured a characteristic too different from the one under examination. This is then a *bad match* to the historical

reference, as is the case when the interval is too shifted (even if not too dilated) (blue) or too dilated (even if not too shifted) (orange). All these ratings comparisons are visualized in our Relative Ratings Map, which shows in which quality rating category each point belongs to.²⁸

Figure 28.



Adrien Delière. From ratings to the Relative Ratings Map (RRMap). (a) Examples of imaginary ratings, where the reference rating is fixed to the interval [30, 50], and the ratings evaluated against it are either dilated, compressed, and/or shifted, except for the central subplot. (b) Representation of these comparisons on the (RRMap), with the Bhattacharyya Coefficient (BC) shown in the background. The regions of interest, i.e., stereotype zone, fair match, and excellent match are indicated; the remaining region corresponds to bad matches.

- 82 This experience, aiming to better characterize the effect of stereotypy in AI-generated images, shows that, in most cases, the stereotypical effect is the result of a *reduction* of the essential characteristics of a style or movement, followed by the *diffusion* of this limited selection across a multitude of very similar objects. It also happened in the case of the brushstrokes employed by Van Gogh in *The Starry Night*, replicated throughout all the landscapes and portraits generated. In other cases, observed during the rating of AI-generated images, certain iconographies are repeated—for example, the motif of the mother and child in the artificial versions of Mary Cassatt, whereas the artist herself produced a far greater variety of subjects in her paintings.
- 83 This statistical test represents a confirmation of the observations earlier made regarding Van Gogh’s style. In some cases, we also notice other kinds of operation, such as the “feature added”. Notably in abstract art, some features are added by the model, such as blurring. In historical abstract art, blurry forms are not very often used, unless in very specific cases such as in the abstract expressionism of Pollock. But in AI outputs for abstract art (prompt: “an abstract painting”), the blur is distributed in every output as if it was an essential feature of abstract art. In this case as well, I hypothesize that the gadgetization of abstraction in decorative art, via anonymous ordinary posters, has had a strong impact on the model’s representational choice.
- 84 The case of Marie Laurencin shows a further kind of socio-cultural stereotype. Laurencin’s style of painting involved vague, blurred figures with tentative outlines. Yet because she is historically considered as Cubist painter, this easily attested style in her paintings, despite being part of the larger database on which such models are trained, is silenced in favor of very marked outlines typical of the most known period of Picasso’s production. That is, the averaging across “cubist painting” and its heavy weighting by better-known, and more gadgetized, male painters such as Picasso, skew prompts calling for “in the style of Laurencin” towards a regional style more readily recognizable as Picasso’s than as Laurencin’s.

Final remarks

85 In this article, I have shown that rather than asking whether GenAI is generally creative or accusing it of producing a stereotyped view of the world, it is instead possible to analyze the enunciational rhetoric of these models. We did so by examining the way GenAI models produce their outputs when asked to operate on plastic and figurative elements, as well as to imitate attested styles. As a result of this analysis, we can distinguish several cases of visual stereotypy produced by multimodal GenAI:

1. The first structural stereotype is exemplified by the last result of the tests involving the face colors. After initially refusing to represent a face without a nose—which I interpreted as a case of failed tropes—the model produced a masked carnival face as the prompt simply requested a blue nose. We thus observe a combination generated by the model that proves to be stereotypical: the association between strange colors in the face (on the expression plane) and the inversion of values typical of carnival attire (on the content plane).
2. The stylistic stereotype, for which the outputs concerning Van Gogh’s style was taken as a major exemplification, is due to gadgetization. This effect of gadgetization is visible in the outputs of Midjourney and confirmed in Stable Diffusion. Indeed, the latter contains many images annotated “Van Gogh” that are not actually by Van Gogh but have nevertheless received this label or association in annotation or, again, in text-image associations in the datasets. In this sense, it clearly appears that the more stylistic variations and citations have been reproduced, the more stereotyped the results become: the original styles remain recognizable but are distorted by a process of stereotypification. The latter manifests as a fairly tight selection of stylistic features considered as prototypical of a certain style, which are repeated and distributed both within a single image and across the different outputs generated from a single prompt. This behavior is also confirmed by the latest tests to which a rating protocol was applied.
3. The third type of stereotypy is what we could call “paratactic stereotypy.”²⁹ This third type specifies the second, and assembles the properties selected as fundamental to a style without producing the overall textural effect of the original style.
4. The fourth facet of the stereotypy under study depends on the datasets on which the model was trained (and on the organization of its latent space): this facet regards what I have called the “global style” of the model. Here, the media filter of the dataset—largely composed of photographs—spreads over the local styles of the artists invoked in the prompt, producing as an effect a photographic rendering of painterly features. I could call it “training-data mediatic stereotypy”.

86 In sum, this article shows the complex interplay of perceptual expectations linked to prototypes of cultural objects and stereotyped conceptions of styles. In addition, I have demonstrated the various ways in which historical painterly styles are reproduced and transformed through an enunciative praxis which such models presuppose (by being trained on the texts enunciated through this praxis) and itself continues ; the relevance of multiple processes of stereotypification—within and outside of such models—are most visible within contemporary gadget-oriented culture as they are taken up by the complex mediations enabled by GenAI models.

BIBLIOGRAPHY

- Agha, Asif. 2003. The Social Life of Cultural Value. *Language & Communication* 23 (3-4): 231-373.
- Badir, Sémir and Maria Giulia Dondero (orgs.) 2016. *L'image peut-elle nier?* Liège: Presses universitaires de Liège.
- Basso Fossali, Pierluigi. 2017a. *Vers une écologie sémiotique de la culture*. Limoges: Lambert Lucas.
- Basso Fossali, Pierluigi. 2017b. Dissuasori d'invenzione: i clichés musicali fra l'attrattore-tradizione e il vettore-creazione. *Nuove musiche*, 2, pp. 55-81. DOI: 10.12871/97888674184286.
- Basso Fossali, Pierluigi. 2026. Inhérences iconiques et poétiques hétérogénératives assistées. IA et nouvelle écologie des images. *Signata* 17. <http://journals.openedition.org/signata/6402>.
- Benveniste, Émile. 1971. *Problems in General Linguistics*. Coral Gables, FL: University of Miami Press.
- Broeckmann, Andreas. 2020. *Optical Calculus* (paper presented at the Images beyond Control conference, FAMU, Prague).
- D'Armenio, Enzo 2024. The mediatic dimension of images: visual semiotics faced with Gerhard Richter's artwork. *Visual Communication*, 23(3), pp. 473-497. <https://doi.org/10.1177/14703572221080529>.
- D'Armenio, Enzo, Adrien Delière and Maria Giulia Dondero. 2024. Semiotics of Machinic Co-Enunciation. About Generative Models (Midjourney and DALL•E). *Signata* 15. <https://doi.org/10.4000/127x4>.
- D'Armenio, Enzo, Adrien Delière and Maria Giulia Dondero. 2025. A Semiotic Methodology for Assessing the Compositional Effectiveness of Generative Text-to-Image Models (Midjourney and DALL•E). *Computer Vision – ECCV 2024 Workshops. Milan, Italy, September 29–October 4, 2024 Proceedings, Part XXII*, Cham: Springer, pp. 112-127.
- D'Armenio, Enzo, Maria Giulia Dondero, Adrien Delière, and Alessandro Sarti. 2025. For a Semiotic Approach to Generative Image AI: On Compositional Criteria. *Semiotic Review*, 9. <https://doi.org/10.71743/ee5nr33>.
- Delière, Adrien, Maria Giulia Dondero, and Enzo D'Armenio. 2025. Revisiting Wölfflin in the Age of AI: A Study of Classical and Baroque Composition in Generative Models. *Journal of Imaging* 11, no. 5: 128. <https://doi.org/10.3390/jimaging11050128>.
- Delière, Adrien, Jeanne Marlot, Marc Van Droogenbroeck, and Maria Giulia Dondero. 2025. How Good Is the Machine at the Imitation Game? On Stylistic Characteristics of AI-Generated Images. *Journal of Imaging* 11, no. 12: 429. <https://doi.org/10.3390/jimaging11120429>.
- Dondero, Maria Giulia. 2015. Les forces de la négation dans l'image. *Au prisme du figural. Le sens des images entre forme et force* (Acquarelli, Luca Org), Rennes: Presses universitaires de Rennes, pp. 105-124.
- Dondero, Maria Giulia. 2020. *The Language of Images. The Forms and the Forces*. Cham: Springer.
- Dondero, Maria Giulia. 2023. Images and their Ability to Negate. *Logic in Question: Talks from the Annual Sorbonne Logic Workshop (2011-2019)*, J.-Y. Béziau, J.-P. Desclés, A. Moktefi, A. Pascu (Orgs), Birkhäuser, Cham: Springer, pp. 321-336. DOI: https://doi.org/10.1007/978-3-030-94452-0_17.

- Dondero, Maria Giulia. 2025a. Semiotics of Artificial Intelligence: enunciative praxis in image analysis and generation. *Semiotica* 262. <https://doi.org/10.1515/sem-2024-0195>.
- Dondero, Maria Giulia. 2025b. Enunciative Praxis and Enregisterment in the Domain of Generative Artificial Intelligence, *Semiotic Review* 12. <https://doi.org/10.71743/bwf80g96>.
- Eco, Umberto. 1975. *Trattato di semiotica generale*. Milan: Bompiani.
- Eco, Umberto, Thomas A. Sebeok, and Jean Umiker-Sebeok, eds. 1983. *The Sign of Three: Dupin, Holmes, Peirce*. Bloomington: Indiana University Press.
- Eugeni, Ruggero. 2026. Superman in Delft. Thinking about enunciation with Visual Generative Artificial Intelligence. *Signata* 17. <http://journals.openedition.org/signata/6357>.
- Fontanille, Jacques. 1995. *Sémiotique du visible. Des mondes de lumière*. Paris: P.U.F.
- Fontanille, Jacques. 2006 [1998]. *Semiotics of Discourse*. New York: Peter Lang.
- Ginzburg, Carlo. 1980. Signes, traces, pistes. Racines d'un paradigme de l'indice. *Le Débat* no. 6, pp. 3-44. DOI: 10.3917/deba.006.0003.
- Greimas, Algirdas Julien. 1989. Figurative Semiotics and the Semiotics of the Plastic Arts. *New Literary History*, 20(3), pp. 627-649.
- Groupe µ. 1992. *Traité du signe visuel. Pour une rhétorique de l'image*. Paris: Seuil.
- Mance, Ivana. 2019. Towards the Theory of the Naïve Art – Grgo Gamulin and the Understanding of Modernism. *Artium Quaestiones*, (30), pp. 191-209. <https://doi.org/10.14746/aq.2019.30.9>.
- Meyer, Roland. 2023. The New Value of the Archive: AI Image Generation and the Visual Economy of Style. *IMAGE. Zeitschrift für interdisziplinäre Bildwissenschaft*, Jg. 19, Nr. 1, S. 100-111. DOI: <https://mediarep.org/entities/article/c2136f34-e0a8-44bd-adbe-c635a818ac4b>.
- Nakassis, Constantine. 2025. Person-Indexing Registers, Stardom, Auteurism, *Signata* 16. <http://journals.openedition.org/signata/5635>; DOI: <https://doi.org/10.4000/14rhi>.
- Nakassis, Constantine and Tatsuma Padoan. 2025. Introduction: Dialogues between Continental Semiotics and Linguistic Anthropology. *Semiotic Review*. <https://doi.org/10.71743/h0nwj114>.
- Offert, Fabian and Peter Bell. 2020. Perceptual Bias and Technical Metapictures. *AI and Society*, pp. 1133-1144.
- Offert, Fabian, Ranjodh Singh Dhaliwal. 2024. The Method of Critical AI Studies, A Propaedeutic. *arXiv preprint arXiv:2411.18833*.
- Pinotti, Andrea. 2025. C'est qui ? L'histoire de l'art sans noms à l'âge du Neural Style Transfer. *International Seminar of Semiotics*, Paris, May 20. <https://www.youtube.com/watch?v=IGSHZTEDukI>.
- Porquet, Julien. 2025. Copier, extraire, générer : Réflexions sur le transfert de style par IA. Invited talk at the Séminaire International de Sémiotique à Paris "Intelligence artificielle générative et nouveaux enjeux sémiotiques. Traduction et appropriations créatives." May 28. Recording available: <https://www.youtube.com/watch?v=ICULM3J-Tmw&t=12s>.
- Schaerf, Ludovica, Andrea Alfarano, Fabrizio Silvestri, Leonardo Impett. 2025. Training-Free Style and Content Transfer by Leveraging U-Net Skip Connections in Stable Diffusion. <https://arxiv.org/abs/2501.14524>.
- Wölfflin, Heinrich. 1932 [1915]. *Principles of Art History: The Problem of the Development of Style in Early Modern Art*. Los Angeles: Getty Publications.

Wilf, Eitan Y. 2013. From Media Technologies That Reproduce Seconds to Media Technologies That Reproduce Thirds: A Peircean Perspective on Stylistic Fidelity and Style-Reproducing Computerized Algorithms. *Signs and Society* 1(2): 185-211.

NOTES

1. On the topic of topological perception, see Offert and Bell (2020) and Somaini (2023) for a discussion on machinic perception. See also Broeckmann (2020) on “optical calculus” as “an unthinking, mindless mechanism, a calculation based on optically derived input data, abstracted into calculable values, which can become part of computational procedures and operations”.
2. The only exception I know of is the study of D’Armenio, Dondero, Deliège, Sarti (2025), in which the authors tested visual formal categories (topology, chromatism, outlines) and enunciative categories (person, time, space) in order to describe and compare the compositional trends of Midjourney and DALL·E.
3. Following Benveniste (1971), the term “instance of discourse” refers to the appropriation of langue in an act of discourse.
4. In all the works mentioned, I respond to Benveniste’s statements (1981) about the impossibility of conceiving of a langue and of a metalanguage in visual images.
5. For an overview about the problem of the generalization of results in the domain of GenAI and some new proposals for Critical Studies, see Offert and Dhaliwal (2024).
6. Latent space is a mathematically structured vector space learned by a model—via techniques such as neural networks and probabilistic models—in which data are encoded as points according to a set of latent variables corresponding to features that capture the underlying statistical, semantic, or structural properties of the data. On the comparison between latent space and Saussurean *langue*, see Dondero (2025a).
7. For a study which shows some rhetorical effects of GenAI, see Basso Fossali (2025).
8. For an interesting insight on these kinds of visual stereotypes, see Paolucci 2027.
9. A register can be briefly defined as a linguistic repertoire that has undergone a process which makes it distinguishable within a language, that is, socially recognized and recognizable (Agha 2003). On the relation between this theory and enunciative praxis in semiotics, see Dondero (2025b).
10. ChatGPT explains as follows the predicted and the desired outputs: “During training, the model makes a prediction — its own guess based on what it has learned so far. This prediction is then compared to the desired output, which is the correct answer provided in the training data. If the prediction doesn’t match the desired output, the model adjusts its internal weights to make its future predictions closer to the correct answers.” The correctness of ChatGPT’s sentences has been verified by Adrien Deliège, a mathematician and expert in AI.
11. This research on stylistic stereotypes is funded by two research projects obtained respectively at ULiège and the F.R.S.-FNRS: DESTINA (“Detecting Stereotypes in Generative Artificial Intelligences: Automatic Text-to-Image Translation and their Effects on Artistic Practices and Visual Studies,” 2025-2027) and “Stylistic Stereotypes in Multimodal Generative AIs: Toward an Education in the Aesthetics and Ethics of the Image” (WelChange Program, FNRS, 2026-2029).
12. On the differences and reciprocal relations between figurative and plastic semiotics, see Greimas (1989). Greimas’s semiotic theory of formal categories (a semiotics of plastic elements) proposes to analyze images through three fundamental categories: topological, eidetic, and chromatic, each defined by oppositional relations. The topological category concerns the spatial arrangement of components in terms of contrasts such as center vs. periphery, left vs. right, or top vs. bottom. The eidetic category pertains to the forms and their contours, distinguishing

rectilinear from curvilinear, fragmented from continuous, and parallel from intersecting structures. The chromatic category, more continuous by nature, involves contrasts of color, brightness (light vs. dark), and material effects (smooth vs. rough) that contribute to the overall perception of form. In relation to light, Fontanille (1995) introduced the useful distinction between illumination, referring to homogeneous lighting, and brightness or glare, referring to dazzling intensity.

13. For a discussion about the style of each generative model depending on its database, see D'Armenio, Dondero, Deliège and Sarti (2025), and Dondero (2025a, 2025b) for a first tentative distinction between regional and global styles.

14. To study regional style semiotically, it is necessary to describe two different configurations of pictures. The first one is the plastic features and the figurative set of marks that are both present in the multiple outputs obtained. This would be a plastic analysis as described by Greimas (1989): a structuralist analysis of topological, chromatic, and eidetic oppositions and repetitions within an image. The second configuration is the compositional repetition (and variation) across the fifty outputs, such as the use of asymmetrical topological and chromatic divisions of the pictorial surface or the stratification of the landscape levels featuring these Van Gogh-inspired pictures. These two dimensions, the plastic features and the compositional repetitions, are of course two different manners of organizing the marks present in the picture. Confronted with a reservoir of diverse plastic categories—such as figure/background, surface/depth, distant viewing/close viewing, linear/curvilinear outlines, and so forth—it becomes possible to identify prevailing compositional trends and the recurrence of specific details or plastic characteristics (e.g., circular celestial bodies populating the sky or chromatic nuances of the sky) associated with the same prompts or with slightly different ones.

15. More precisely, the randomness in the image generation process using diffusion models relates to the fact that a trained diffusion model has learned, from a noisy image and a piece of text, how to reconstruct an image corresponding to the original (non-noisy) image in the dataset and its associated text. Ideally, then, a well-trained model should be able, at the time of inference (i.e., when the model is being used), to create an image corresponding to any given text starting from “noise” (without associating it with any specific image from the dataset, as that association only occurs during training). So, at the beginning of inference, a noise image is generated and paired with the prompt (more precisely, the embedding of the revised prompt), which is then progressively transformed into a (denoised) image by the diffusion model. This step introduces randomness, because a different *initialization* (i.e., a different noise seed) will result in a different final image.

16. For a useful explanation about Latent Diffusion Models—and in particular Stable Diffusion—and the functioning of the denoising steps, and notably the third step which makes possible the separation between the semantic features and the style configuration, see Schaerf, Alfarano, Silvestri and Impett (2025).

17. “Naïve” art here refers to a theory of, and historical approach to, art wherein the artwork is made without presuppositions or influences from the history of art or of the world, as if it was a spontaneous expression of the imagination of the artist. We might read the term naive, then, in both its senses, as indicating something untainted and pure but also in the pejorative sense as a kind of fantasy uninformed by precisely what it claims to abjure. For a problematization of naïve art, its definition, its differences from amateurism and realism, and the paradox of a style without style, see Mance (2019).

18. For an overview on enunciative praxis and its operations on visual forms (actualization, realization, potentialization and virtualization) see Dondero 2025a and 2025b. On the processes of *dissemblage* (elimination of the individuality of the items in the training corpora) and *ressemblage* (the process through which pseudo-historical and pseudo-cultural objects are generated) in

diffusion models, see Eugeni (2026) who claims that these two processes constitute the enunciative praxis performed by multimodal GenAI.

19. <https://laion.ai/blog/laion-aesthetics/>.

20. On the textual density identifying a style, see Nakassis (2025). Discussing Hitchcock's films and his "screen signature," Nakassis argues that the list of features that make a style distinguishable from others is a cultural, metapragmatic model which represents, and in a way skews, what is otherwise more tacit: the competence by which viewers identify a certain textual coherence (co-occurring elements) across his films. (The list skews such a capacity by reducing the register, and the differential weights of its indexicalities as they are put into play in particular textures, to a simple list, as if each element could stand on its own as an index of the style of the auteur.) Nevertheless, Nakassis alerts us to the possibility that the authors may differ from themselves and from their 'enregistered personage,' (a sort of stereotype of their own individuality), and may not live up to their own name.

21. On the topic of "artificial styles," see Meyer (2023), who states that historical forms of representation can be detached at will from their time and place of origin and the work of their authors. "Thus 'style' ceases to be a historical category and becomes a pattern of visual information to be extracted and monetized." (Meyer 2023: 106–107, our emphasis)

22. On this topic, see Pinotti (2025), who explains that the grouping of artists into movements or schools depended, for a long period of time, on successive oppositions of styles and on subsequent integrations of previously opposed stylistic characteristics within the same movement. This occurs after operating further oppositions in relation to other, more different, styles.

23. According to an evidential paradigm (Ginzburg 1980 ["paradigma indiziario"], Eco 1983, Basso Fossali 2017a), the regional style would not designate a region contained within the dataset or the latent space that analysis simply identifies. Rather, it designates an evidential construct reconstructed *a posteriori* from recurring regularities observable in a series of outputs. In other words, each region is the name assigned to a cluster of regularities inferred from the outputs.

24. I am grateful to Constantine Nakassis for pointing out that a strong presence of the global style generally characterizing the model may be considered as a sign of the model's failure to learn a painter's style, which becomes indistinguishable from others (the differences in the Firsts that are strung together notwithstanding).

25. On painterly and photographic mediatic specificities, studied on the corpus of Gerhard Richter, see D'Armenio (2024).

26. Wölfflin developed a comparative art-historical method based on a series of stylistic oppositions that characterize classical and baroque visual regimes: linear vs. painterly, planar vs. recessive, closed vs. open form, multiplicity vs. unity, and clarity vs. obscurity. While Greimas focused on local oppositions between plastic features, Wölfflin conceived of the image as an integrated composition, analyzing how light, form, and space cohere into an overall visual order.

27. This was discussed in the context of the intensional versus extensional models of reasoning, a distinction that Aurora Donzelli recommended I apply to the analysis of GenAI images and those examined through computer vision tools, during the seminar "Enunciation/entextualization" (2023-2024) which prepared the publication of "Dialogues between Continental Semiotic and Linguistic Anthropology" published in *Semiotic Review* in September 2025 (Nakassis and Padoan 2025).

28. For more technical details, see Deliège, Marlot, Van Droogenbroeck and Dondero (2025).

29. On paradigmatic and syntagmatic stereotypes, see Basso Fossali (2017b).

ABSTRACTS

In this article, I address the enunciative rhetoric of AI-generated images, focusing, in particular, on two questions. The first concerns the composition rules learned by generative models in representing the human face through prompts asking for the negation, subtraction, addition, or substitution of some fundamental parts of the face such as the nose. These tests have the objective of studying the (supposed) path the machine realizes during the process of image instauration and of describing the image's compositional rhetoric in relation to the difficulty of these models in figuring uncommon figures such as noseless faces. Indeed, a noseless face is something unexpected and rarely found in the datasets which have trained these kinds of models. The second part of the article uses another corpus as a demonstration of the model's work with prototypic, unique, or unusual topics. It explores how GenAIs, solicited through prompts mentioning recognized artistic painterly styles (e.g., Van Gogh's and Klee's styles) bring forth stereotypical forms and textures. In this context, by starting from the familiar forms of historical styles in the artistic tradition, it becomes possible to estimate both the level of stereotyping and the level of creativity of these models.

Dans cet article, j'aborde la rhétorique énonciative des images générées par l'IA, en me concentrant sur deux questions. La première concerne les règles de composition apprises par les modèles génératifs pour représenter le visage humain à travers des prompts demandant la négation, la soustraction, l'ajout ou la substitution de certaines parties fondamentales du visage, telles que le nez. Ces tests ont pour objectif d'étudier le chemin (supposé) que la machine suit au cours du processus de création de l'image et de décrire la rhétorique compositionnelle de l'image en relation avec la difficulté de ces modèles à représenter des figures inhabituelles, telles que des visages sans nez. En effet, un visage sans nez est quelque chose d'inattendu et que l'on trouve rarement dans les ensembles de données qui ont servi à entraîner ce type de modèles. La seconde partie de l'article utilise un autre corpus pour montrer le fonctionnement stéréotypant du modèle. Cette partie explore la manière dont les IA génératives, sollicitées par des prompts mentionnant des styles picturaux artistiques reconnus (par exemple, les styles de Van Gogh et de Klee), produisent des formes et des textures stéréotypées. Dans ce contexte, en partant des formes familières des styles historiques de la tradition artistique, il devient possible d'évaluer à la fois le niveau de stéréotypie et le niveau de créativité de ces modèles.

INDEX

Mots-clés: IA générative, rhétorique visuelle, énonciation, image, stéréotype

Keywords: Generative AI, visual rhetorics, enunciation, image, stereotype

AUTHOR

MARIA GIULIA DONDERO

Maria Giulia Dondero, PhD, is a Research Director of the National Belgian Fund for Scientific Research (F.R.S.-FNRS) and Professor at the University of Liège where she teaches the Visual Semiotics and the Semiotics of Photography. She is the President of the International Association

for Visual Semiotics (IAVS/AISV) since 2023 and Responsible for the Internationalization in the Board of the French Association for Semiotics (AFS).

She is the author of five books: *Semiotics of Images. Analysis of Pictorial Texts*, 2024, De Gruyter, (with A. Lagopoulos, K. Boklund, J. Fontanille, R. Wallden, M. Katzaridou); *Les Langages de l'image. De la peinture aux Big Visual Data*, Paris, Hermann Éditions, 2020 (English augmented version: *The Language of Images. The Forms and the Forces*, Cham, Springer, 2020); *Des images à problèmes. Le sens du visuel à l'épreuve de l'image scientifique*, with J. Fontanille, Limoges, Pulim, 2012 (Eng. trans. *The Semiotic Challenge of Scientific Images. A Test Case for Visual Meaning*, Ottawa, Legas, 2014); *Sémiotique de la photographie*, with P. Basso (Limoges, Pulim, 2011); *Le Sacré dans l'image photographique* (Paris, Hermès, 2009). She has published around 120 peer-reviewed articles in French, English, Italian; some of her works have been translated into Spanish, Portuguese, Polish. She has directed 40 collective works and special issues on semiotic theory of visual language, scientific and artistic images, photography, and Generative Artificial Intelligence. She is Co-founder and Editor-in-Chief of the peer-reviewed Journal *Signata Annales des Sémiotiques / Annals of Semiotics* (<https://journals.openedition.org/signata/>) (Scopus, ANVUR and DOAJ) and Co-director of the book series 'Sigilla' at Presses universitaires de Liège. She is a member of the editorial board of several scientific journals (*Semiotic Review*, *Actes sémiotiques*, *Intelligibilité du numérique* among others). She has been Visiting Professor in Tunisia, Brazil, France, Italy, Argentina, United States.
Email: MariaGiulia.Dondero[at]uliege.be