UNDERSTANDING VISUAL STEREOTYPING IN IMAGES GENERATED BY MULTIMODAL AI: A MATTER OF STYLE?

Jeanne Marlot
University of Liège
j.marlot@uliege.be

Maria Giulia Dondero
F.R.S.-FNRS – University of Liège
mariagiulia.dondero@uliege.be

Abstract

This article examines the stylistic stereotypes (re)produced by multimodal GenAIs, proposing two complementary methodological approaches. The first, which combines quantitative and qualitative methods, is based on a contrastive analysis of historically produced works and automatically generated images. It is conducted using a notation grid composed of nine criteria based on the plastic categories formulated by A. J. Greimas and J.-M.

Digital Age in Semiotics & Communication, Vol. IX, 2026, 47–70
Under CC BY 4.0
https://doi.org/10.33919/dasc.26.9.3

Floch, as well as on H. Wölfflin's stylistic oppositions between the classical and baroque ways of seeing. This grid made it possible to statistically objectify the perceptual judgments of three experts. The tests through this notational protocol, re-examined in light of the language sciences, show why this attempt to grasp stylistic stereotypes is insufficient on its own. The second method consists of generating tropes by submitting to three multimodal AI systems prompts that deliberately combine a genre and a style that are historically unattested together and appear to be incompatible. By examining the relationship between visual genres and artistic styles through prompts that depart from historically attested configurations, the aim is to test the resilience of what is embedded in databases and, more broadly, in structures of meaning. The two proposed exploratory protocols open up avenues for investigating how generative models perform typification by producing stereotypical visual configurations and, conversely, how these stereotypes tend to impose themselves as rules governing image generation.

Keywords: generative artificial intelligence, stereotypes, style, trope, visual genres

Introduction

The concept of *style*, although long discussed in aesthetics and art history, remains problematic in the analysis of works. The emergence of text-to-image generative artificial intelligence, as well as works on *Neural Style Transfer* in computer vision communities, has recently reignited interest in this concept: to what extent have these models learned and are they capable of reproducing the creative gestures associated with an artist or an artistic movement?

This article proposes to address these questions by measuring the levels of creativity and stereotypy of multimodal generative AI systems (GenAIs). These systems are trained on a vast amount of statistically processed data, which may generate biases in the images they produce[1]. To our knowledge,

[1] Automatic image generation relies on three main components: large image-text datasets; efficient feature extraction methods based on deep learning; and powerful hardware infrastructure for training and inference. Current text-to-image generation models combine a natural language processing module, which converts prompts into embeddings (lists of numbers), and a diffusion-based system, which uses these embeddings to guide visual

the scientific community has not yet developed systematic methods to objectively assess the fidelity with which generative models reproduce pictorial genres, movements, and artistic styles. This is why we present two exploratory approaches aimed at understanding these stylistic stereotypes: the first, which combines statistics with semiotics and art history, is based on a contrastive analysis of historically created works and those generated automatically (1); the second tests whether GenAI can generate tropes that depart from the stereotypical visual forms it usually produces or reveals (2).

A brief overview of the key concepts used here will help clarify the objectives of the research presented below. In the artistic domain, *style* refers to a set of formal properties that allow one to identify an artist, a work, or a movement. It emerges from a tension between *uniqueness* (the innovation specific to a creator) and *repetitiveness* (the construction of a recognizable type based on a selection of features perceived as salient by certain recipients). This repetitiveness is based on a dual intertextuality: on the one hand, *intrapoetic*, when features are distributed across the work of a single artist and reveal a similar style; on the other hand, *interpoetic*, when formal properties common to several artists allow them to be grouped under the name of a single artistic movement (Dondero 2009). The term *type* refers to the perceptual process by which one extracts a set of recurring features from a series of occurrences. A *stereotype* arises when this process becomes reductive: in the case at hand, certain stylistic traits characteristic of a sample are overgeneralized to the entire category, freezing the style into a limited number of salient properties at the expense of its variations.

To avoid these minimal definitions becoming too abstract, we take a representative example of stylistic stereotypes (Dondero 2026a, 2026b). The microsyntax of Vincent van Gogh's style in his landscapes notably includes the motif of the "celestial body", as illustrated by the stars in *The Starry Night* (1889), the sun in *Olive Trees with Yellow Sky and Sun* (1889), or the golden

---

generation. They thus ensure a multimodal conversion determined by three factors: the structure of the training data—shaped by human practices of selection, annotation, and categorization—the learned rules, and the model architecture. This conversion relies on the similarities between the distributions of verbal and visual embeddings within a latent space. Images are generated through operations applied to pre-existing images, stored and annotated according to criteria such as style, artist, theme, or genre in databases such as *WikiArt*, *Artsy,* or *Google Arts & Culture*. These digitized image databases cover a wide spectrum of styles and artists, ranging from art history to contemporary photography, to recent trends in comics and animation. The generated images are inevitably stereotypical insofar as they tend to reproduce the formal characteristics of the most widely recognized images that the generative models encountered during their training phase.

luminosity that radiates from *Landscape at Twilight* (1890). These elements are typical of the artist's landscapes and thus fall under intrapoetic intertextuality. Multimodal GenAIs, however, overgeneralize this trait by applying it to portraits "in the manner of Van Gogh". This stereotypification also affects the brushstrokes, often limited to the swirling effects characteristic of *The Starry Night*. Yet the artist employs a certain technical diversity: the *Portrait of Madame Ginoux* (1890) is characterized by broad swaths of color, whereas in *Orchard with Cypresses* (1888), the fragmented brushwork resembles Pointillist works—a connection that establishes an interpoetic intertextuality, insofar as the application of paint in small, spaced-out touches evokes, in this latter painting, the techniques employed by Seurat and Signac. This means that AI models are not capable of embracing the full subtlety and complexity of Van Gogh's original paintings.

A clarification is needed before beginning the tests. The images generated during our research have no aesthetic pretensions: rather, they serve as scientific artifacts within the framework of a stylistic analysis. While artists and other creators strive to develop methods for guiding algorithmic processes to maximize the personalization and uniqueness of their creative paths by refining databases, the tests we conducted with these models aim to understand the machine's operation for its own sake, by minimizing user intervention, and to discern what distinguishes automatically generated images from those produced without the mediation of multimodal AI. Consequently, when we request a translation from natural language to a visual style, we deliberately choose a highly recognizable style and associate it with a very general theme or visual genre. The advantage of proposing highly recognizable styles is that it becomes possible to estimate the degree of stereotypy and creativity exhibited by each generative model—through the processes of annotation, database construction, and algorithmic processing[2]—by comparing the generated images to those that have shaped the history of art.

1. Evaluating stylistic imitation through contrastive analysis

As a first step, we develop, using statistical tools, a notational protocol designed to measure—by comparing images with one another—the fidelity with which text-image generative models reproduce established artistic styles (1.1), then we examine its methodological foundations through the lens of language sciences (1.2), before drawing critical lessons from it (1.3).

[2] For a discussion of the style of each generative model based on its database, see D'Armenio et al. (2025).

1.1. Methodological clarification

The first methodological proposal revisits an article (Deliège et al. 2025) titled "How Good Is the Machine at the Imitation Game? On Stylistic Characteristics of AI-Generated Images", which aimed to evaluate Midjourney v6's ability to reproduce the stylistic traits associated with major artistic movements and their representative artists. To conduct this study, five movements were selected (Impressionism, Cubism, Surrealism, Abstract Art, and Expressionism), and each of these movements was represented by two artists, a famous male painter and a lesser-known female painter: Claude Monet and Mary Cassatt; Pablo Picasso and Marie Laurencin; Joan Miró and Dorothea Tanning; Wassily Kandinsky and Hilma af Klint; Ernst Ludwig Kirchner and Gabriele Münter.

Our contribution lies in the development of an evaluative approach within a semiotic and art-historical framework that combines expert judgment with quantitative analysis. Three experts evaluated works produced by artists belonging to the selected movements and images generated by Midjourney, using an analytical grid composed of nine criteria to examine their visual composition. This grid draws on the plastic categories (topological, eidetic, chromatic) developed by the semiotician Algirdas Julien Greimas ([1984] 1989) and Jean-Marie Floch (1985), as well as the categories established by the art historian Heinrich Wölfflin ([1915] 1950) to characterize the classical and baroque perspectives (namely: linear *vs.* painterly; planar *vs.* recessional; closed form *vs.* open form; multiplicity *vs.* unity; clearness *vs.* unclearness).

The criteria selected for the study are presented as pairs of opposites, with the terms placed on either side of a double arrow to indicate that the plastic properties under consideration form part of a continuum, measured on a scale from 0 to 100:

- centered forms [0] ↔ edge-deployed forms [100]
- left-right symmetry [0] ↔ left-right asymmetry [100]
- top-bottom symmetry [0] ↔ top-bottom asymmetry [100]
- sharp image [0] ↔ blurred image [100]
- parallel lines [0] ↔ intersecting lines [100]
- dense rhythmic structures [0] ↔ sparse rhythmic structures [100]
- saturated colors [0] ↔ desaturated colors [100]
- uniform lighting [0] ↔ focalized lighting [100]
- painterly texture [0] ↔ photographic texture [100]

For each movement and each criterion, the experts evaluated six sets of images: the historical works of the movement, those of the selected male and female artists, as well as images generated by Midjourney using prompts that included the name of the movement, the male artist, and the female artist, respectively. The center of the interval indicates the typical value of the image set for the criterion in question, while its width reflects internal diversity (Figure 1). The evaluation is based on a 95% rule: for each criterion, the experts were asked to base their judgment on nearly all of the images considered, while excluding exceptional cases (5%). For the generated images, this meant that approximately one in twenty images could fail to meet the evaluated criterion without affecting the rating. When evaluating historical works, this rule took a different form: experts obviously could not have in mind the entirety of the works produced by an artist or a movement, but they were asked to base their evaluation on as broad a representation as possible of this body of work, disregarding obvious exceptions. This method thus offers a systematic comparison between the general representation of a movement (historical *vs.* AI) and the individual styles of artists (historical *vs.* AI).

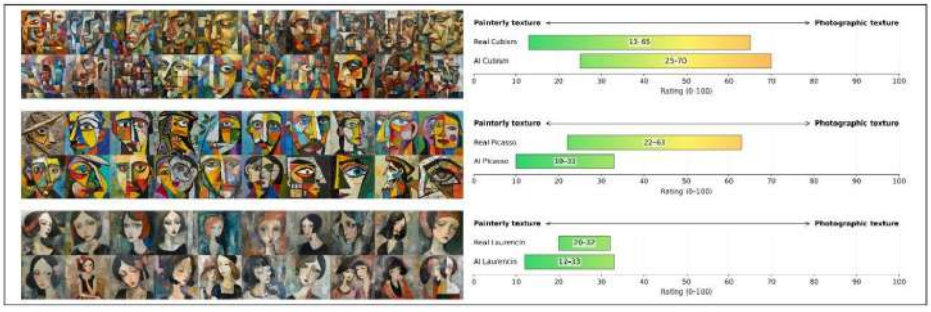

Figure 1: Illustration by Adrien Deliège of one of the evaluation sheets used for the study. The left side shows the AI-generated images (from October 2025) that were submitted for expert review. The right side shows the scores assigned by an expert, on a scale of 0 to 100, for the painterly texture ↔ photographic texture criteria, in reference to paintings historically produced by the Cubist movement (top), by Pablo Picasso (middle), and by Marie Laurencin (bottom), as well as by Midjourney. The scales arranged one below the other thus allow for a comparison of the scores assigned to each set of images according to each evaluation criterion.

The implementation of such a protocol reflects a methodological ambition that is intended to be carried forward in subsequent studies: namely,

the objectification of perceptual judgments that are ordinarily understood as grounded in individual intuition, or at least as varying according to each observer's perceptual skills and cultural frames of reference. By guiding image analysis through a shared set of criteria and translating expert judgments into comparable numerical values, this framework opens the way toward an empirical stylistics of AI-generated images—a field in which methodological rigor remains largely to be established.

1.2. Insights from language sciences

The article "How Good Is the Machine at the Imitation Game? […]" appeared in the computer vision periodical *Journal of Imaging*; it takes a quantitative approach to stylistic imitation. We propose a complementary, qualitative reading, grounding the analysis not in statistics but in language sciences. The experts' reasoning can indeed be reexamined through two conceptual pairs: *intension* and *extension* (derived from Frege [1892] 1948 and reworked by Carnap 1947) in the philosophy of language on the one hand, and *intensity* and *extensity* in tensive semiotics (Zilberberg 2012) on the other. While not perfectly overlapping, these two pairs resonate with one another and shed new light on the evaluation protocol that has been established.

When experts assign ratings to works associated with a particular artistic movement or painter, they are not physically confronted with the thousands of images produced during that period. They proceed by abstraction, identifying a few general characteristics deemed frequent and fundamental—a process that can be likened to *intensional reasoning* in the logical sense of the term (that is, proceeding by the selection of defining fundamental features characterizing a painter or a movement). Conversely, when Midjourney-generated images are displayed before the experts during the rating process, they directly observe all the occurrences and their visible characteristics—a process that falls more under *extensional reasoning* (taking into account the totality of the images effectively present and observable). This conceptual pair, *intension/extension*, calls for two remarks.

First, this mode of reasoning re-enacts the relationship between *types* and *occurrences*. In *Kant and the Platypus*, Umberto Eco ([1997] 1999) addresses these questions through the concept of the *cognitive type*, which he discusses in relation to the *prototype*. The latter encompasses several meanings; here we will focus on the one that defines it as a schema, conceived as a bundle of features. Theoretically, this prototypical schema acts as a mediator between the features of the type and those identifiable in the occurrences. In practice, the expert draws upon a set of internalized features derived from perceptual

experiences during which they have encountered Impressionist, Cubist, Surrealist, Abstract, or Expressionist works. This prototype then allows them to recognize, in each occurrence (whether it be a recalled work or a generated image), an approximate correspondence with this type. It is precisely this correspondence that the protocol seeks to formalize.

Second, the reasoning employed in the evaluation of images *in abstracto* is not only intensional but also extensional. In the protocol, the evaluation of historical works also relies on extensional reasoning, since experts must estimate an interval intended to cover approximately 95% of the works they know and consider representative of the criterion being evaluated. The extensional logic of the reasoning lies in taking into account the entirety of these works, even if they are not physically present. Access to the extension therefore differs depending on whether the evaluator mentally reconstructs the images corresponding to the artist or movement based on their knowledge of art history, or whether they directly observe the AI-generated image instances presented to them. We would thus distinguish between a *mnemonic extension* and an *empirical extension*.

These considerations are echoed in Zilberberg's (2012) tensive semiotics, which offers a complementary framework for thinking about the perceptual gradient of judgments. The evaluation protocol has an *intensive* dimension when it asks experts to identify and focus their attention on certain plastic qualities deemed typical of the movement or the artist—qualities that stem from experts' internalized aesthetic experience. Whether the corpus is memorized or directly observable, the question remains: is a given characteristic strongly or weakly present in the corpus?

The expert is confronted with the *extensive* component of the protocol when seeking to determine whether the observed characteristics are diffuse (distributed across nearly all images) or concentrated (limited to a small number of works or generations). Nevertheless, the perceptual judgment regarding the distribution of traits depends on the modalities under which the evaluation experience takes place. When the experts evaluate images they do not see directly, they mentally reconstruct a set of occurrences distributed across time and space, seeking to cover the variability of the corpus. In contrast, when faced with directly observable images, they apply the 95% rule defined by the protocol. Thus, the difference between the evaluation of historical works and that of AI-generated images can be formulated in terms of extensivity.

In short, the theoretical distinctions discussed in this section have direct implications for the analysis of the evaluation protocol's results: experts do

not operate under equivalent cognitive conditions when assessing historical artworks and AI-generated images. This asymmetry must be taken into account in order to properly conduct the study. Any future iteration of the protocol could attempt to mitigate it by providing experts with a reference corpus composed of selected historical works. However, this choice introduces its own biases, which must be considered in the interpretation of the results: why was one work by a given artist selected rather than another, and according to which criteria?

1.3. Results, limitations, and prospects

After presenting the evaluation protocol and its semiotic implications, we now turn to the analysis of the results. The results obtained show that Midjourney v6 correctly reproduces certain historical or authorial trends, but also produces stereotypes or fails to reproduce certain stylistic features and in particular the specific global effect of a painterly style and its texture. These results could only be established by addressing a methodological difficulty posed by the relativity of perceptual judgments. The perception of what constitutes a stereotype varies, in fact, according to each observer's experiences and cultural frames of reference, making individual evaluations difficult to compare. It is precisely this difficulty that the evaluation grid sought to overcome. Developed on the basis of nine criteria related to image composition and presented by the authors as transhistorical, this grid made it possible to formalize the experts' perceptual judgments in a numerical form, objectifying through statistical analysis observations that, without this scoring system, would have remained in the realm of individual intuition.

This approach paves the way for research aimed at determining how the typification performed by generative models produces stereotypical visual configurations and, conversely, how these stereotypes tend to establish themselves as production rules through the recurrence of certain patterns in the training corpora. Nevertheless, these nine criteria alone do not suffice to fully address the issue of stereotyping. In the cited article, a stereotype is identified when the generated images reproduce the central tendency of the historical works, while exhibiting lower variability (the two intervals are located roughly at the same point on the rating scale, but that of the generated images is narrower than that of the historical works).

However, the evaluation sheets raised certain difficulties. For example, how can one quantify a paradoxical observation? Regarding the criterion related to painterly and photographic texture, the analysts noted that while AIs accentuate the artist's simulated gestures by highlighting brushstrokes,

a closer examination reveals that these marks imitate the pictorial texture without reproducing its plastic properties. To borrow the terms of the Groupe μ (1992: 203–209), generative models produce "macules" (two-dimensional textural elements) where the act of painting mobilizes a "grain" (a three-dimensionality resulting from the impasto, the substrate, the application and the rhythm of the application gesture): in the generated images, as in photography, textural volume is thus simulated rather than realized. In multimodal AI-generated images, two media-specific characteristics compete (Dondero 2026b): one painterly, which attempts to imitate the style of the artists mentioned in the prompt by simulating the effect of brushstrokes here and there; the other photographic, which corresponds to the model's average signature. We refer to these two stylistic imprints respectively as *regional style* (typical of the painter in question) and *global style* (the style specific to the generative model) (Dondero 2025b).



Figure 2: Images generated by Midjourney v6 from the prompt "An impressionist painting" (October 2025).

The grid submitted to the experts, whose task was to identify stylistic stereotypes, focused primarily on the visual composition of the images, neglecting the figurative dimension, which is sometimes distinctive of an artist's style. Was it sufficient for our purposes? Not entirely. For example, most of the images generated by Midjourney imitating Mary Cassatt's style depict a young woman embracing a child.



Figure 3: Images generated by Midjourney v6 from the prompt "A painting in the style of Mary Cassatt" (October 2025).

However, while the Impressionist painter is best known for her portraits of children, most often accompanied by their mothers or nurses, she did not limit herself to scenes of affection. She also depicted the daily lives of bourgeois women teaching a child to read, reading a newspaper, holding the reins of a carriage, going out in a boat, or attending theater performances (Figure 4). What might explain such a narrowing of the thematic range addressed by the artist throughout her career? This reduction can partly be explained by historical realities: as a woman, Cassatt had limited access to subjects available to her male counterparts (not everyone was willing to sit for her, and the social spaces she could access and depict were constrained by the gender norms of her time). It is therefore true that relationships between maternal and child figures constituted a significant theme in her artistic production. Added to this is the persistence of sexist bias in the datasets used to train generative models: identifying the artist as a woman may have led image annotators to interpret her work primarily through the lens of motherhood.



Figure 4: Collage of works by Mary Cassatt available on Wikipedia: *The Reading Lesson* (c. 1901); *Reading* Le Figaro (1878); *A Woman and a Girl Driving* (1881); *Summertime* (1894); *In the Loge*, also known as *At the Opera* (1879).

This figurative stereotype produced by Midjourney may also be studied through the lens of visual genres. For example, the prompts "an Impressionist painting" and "a Cubist painting" yield images of landscapes and portraits, respectively, even though the Impressionists did not limit themselves solely to landscape depictions—think of Edgar Degas's dancers, for instance—while the Cubists painted visual genres other than portraits, notably still lifes, as illustrated by several paintings by Juan Gris. These generated images suggest a tendency for generative models to associate prevalent visual genres with the artistic styles they are prompted to reproduce.

The limitations identified suggest that future studies adopting a similar protocol should expand the analytical framework by considering the figurative and generic dimensions of the images under examination. Moreover, only three experts were consulted in the present study. The panel of evaluators should therefore be broadened, both in size and diversity, to include



Figure 5: Images generated by Midjourney v6 using the prompts "an Impressionist painting" and "a Cubist painting" (October 2025).

not only additional specialists but also non-experts, in order to provide a more robust statistical basis for the analysis and to test the protocol against a plurality of points of view. These limitations led to the implementation of a series of complementary experiments, which are discussed in the following section.

2. Testing stylistic imitation through tropes

Based on the observation that multimodal GenAI models seem to associate a preferred visual genre with artistic styles, we subsequently developed another methodology to examine whether generative models are capable of producing images that deviate from these recurring associations. After presenting the methodological foundations of this research (2.1), we address our results, showing the means by which these AIs attempt to produce visual compositions whose configuration—constrained by the instructions embedded in the prompt—is intended to be novel (2.2).

2.1. Methodology and corpus construction

One way to identify the text-image associations most frequently recorded by multimodal AI systems—and thus those most likely to give rise to stereotypes—is to push these AI systems to their limits. We therefore decided to ask them to produce portraits, still lifes, or landscapes in the style of painters whose work does not typically fall within these genres. By exposing generative models to unusual stylistic and genre combinations, the goal is to examine whether or not they can break free from the visual patterns most frequently present in their training data.

To conduct this experiment, three multimodal AI models were selected: Midjourney v6, ChatGPT Image (gpt-image-1.5 model), and Nano Banana (Gemini 2.5 Flash Image model).[3] For each prompt, eight images were generated in late January–early February 2026, using each model's default settings. Repeating the generations from the same prompt makes it possible to observe the potential emergence of visual patterns that allow for the identification of stylistic trends.

This experiment, which aims to test the robustness of the information embedded in the databases by analyzing the relationship between visual genres and artistic styles through historically unrecorded and deliberately incompatible combinations, relies on the development of specific prompts:

- A landscape in the style of Hilma af Klint;
- A landscape in the style of Sam Francis;
- A portrait in the style of Barnett Newman;
- A portrait in the style of Morris Louis;
- A still life in the style of Jackson Pollock;
- A still life in the style of Piet Mondrian.[4]

Once these textual descriptions are fed into a generative model, the model is required to produce images (a person, a group of objects, a natural environment) in the style of an artist renowned for abstraction. These combinations deliberately deviate from established visual conventions in order to uncover the stereotypical underpinnings of the generated images. Our approach thus invokes the concept of *trope* (cf. Dondero 2026b; Marlot 2026), for which two complementary definitions can be proposed.

One of these is provided by the Groupe μ, from the perspective of semiotics and visual rhetoric. Iconic signs are conceived here as cultural stabilizations of sensory experiences in the form of a conceptual set of typical

---

[3] *ChatGPT Image* and *Nano Banana* offer two entry points: a conversational interface and a programming interface (API). The conversational interface poses a methodological problem: it retains previous prompts in memory and may take them into account during subsequent generations. In an approach aimed at testing variables in isolation, it is therefore preferable to use the API, which guarantees the independence of each query and thus the reproducibility of the experiments.

[4] The five male painters cited are the subject of a philosophical discussion in the course *On Painting* taught by Deleuze (2023). We selected these artists because the distinctive elements of their styles are relatively well-established in Western culture. However, it seemed to us that geometric abstraction and women were not sufficiently represented. This is why Hilma af Klint was added.

characteristics. In the visual domain, tropes correspond to operations of substitution involving stabilized figurative and plastic configurations, which produce a *deviation* from a conventional usage designated as *degree zero*. Whereas the Groupe μ emphasizes rhetorical categorial operations, another relevant perspective, that of linguistic anthropology, highlights the situated nature of tropes.

Indeed, according to the linguistic anthropologist Asif Agha (2024: 8, 15–17), a *trope* is an effect of text configurations produced when forms normatively associated with a register are recontextualized within a given interactional context, thereby generating a perceptible shift in usage. A *register* constitutes a repertoire of signs—linguistic or, more broadly, semiotic—with which speakers and listeners associate a particular communicative situation to a typology of social activities and persona (Agha 2024: 1–2). This reconfiguration of the expressive resources constituting a register produces a trope-based meaning, characterized by three main features:

1. the effect of meaning cannot be grasped independently of the *co-text*, that is, the signs surrounding those to which the trope refers;
2. although common in social life, tropes do not derive their meaning from the forms themselves, but from the metapragmatic frameworks that guide their interpretation throughout the interaction;
3. correspondingly, tropic meaning is difficult to decontextualize: outside of context, stereotypical values dominate, even if certain tropic configurations can be reanalyzed and crystallized into new recorded models, endowed with their own social domain and indexicality.

We decided to transpose the conceptual framework of *register*, derived from linguistic anthropology, to the automated generation of images that mimic artistic styles, relating it to the semiotic theory of enunciative praxis (Dondero 2025a, 2025b), which explains the processes of sedimentation, typification and innovation of cultural forms from a diachronic perspective. From this standpoint, an artist's style can be conceived in relation to the concept of *register* and *enunciative praxis*. This transposition finds further support in the work of another linguistic anthropologist, Constantine V. Nakassis (2025), who coined the concept of *person-indexing registers* (as opposed to *persona-indexing registers*) to designate repertoires of signs that index a singular individual instead of a type.

Indeed, we argue that style fits easily into this category of *person-indexing registers*. On this view, artists' styles can be seen as registers whose signs appear as traces left by a person in the utterance, with certain of their properties being recognized as individuating for a given community in a

specific context. In other words, not all the constitutive features of this individual's utterances are perceived as relevant by the community, nor are they all specific to them. What matters are the properties that this community recognizes as indexical of this person as their singularity. Thus, registers raise the question of the conventions and attributions established by a community of recipients.

This line of thought brings us back to tropes when we consider that Constantine V. Nakassis (2025) links *person-indexing registers* to the concept of *rigid designators* developed by the philosopher of language Saul Kripke. According to Kripke (1980), a proper name is assigned to an entity during a baptismal ritual; this performative act serves to individualize the referent. In this way, the proper name continues to designate the same entity regardless of context. From this perspective, an artist's name should designate that artist even in counterfactual contexts—that is, possible situations in which certain properties differ while the referent of the name remains unchanged. This rigidity of reference raises precisely the question of what happens when we subvert the regularities of an artist's practice by introducing elements absent from their work (such as when we ask generative models to apply the style of an artist to a visual genre that the artist has never practiced).

2.2. Analysis of the results

Based on the results obtained, we sought to determine whether the models conceive genres and styles as we perceive them; whether they recognize the tropic incompatibilities built into the prompts; and, if so, what rhetorical strategies they deploy to resolve these tensions. In the sections that follow, we focus on a few representative cases.

2.2.1. The center-periphery relationship (landscape and abstraction)

According to the tests conducted, generative models tend to translate incompatible genre-style pairs spatially: the center accommodates the figurative elements of the visual genre, while the periphery presents the artist's stylistic characteristics. We call this topological organization the "center-periphery relationship", which separates the constraints of genre and style by distributing them across distinct areas of the image. This configuration can be observed in artificially generated landscapes and still lifes.

For example, ChatGPT Image processes the prompt "A landscape in the style of Sam Francis"[5] by placing the landscape genre at the center of the

[5] Sam Francis (1923–1994) was an American Abstract Expressionist. His work is characterized by painting on a white background that features bands of color crossing the canvas,

image in the form of bluish mountains. It thus borrows its pictorial technique from atmospheric perspective, creating an opening toward a distant space typical of the landscape. Around this figurative center, patches of varied colors spread out toward the splattered white edges of the image. This operation seems to actualize the structural role of white in Sam Francis's art, conceived as the "color of all colors" and a virtual chromatic reservoir. However, the pictorial gesture is reduced to crude splashes that only imperfectly capture the artist's technique. In short, the image does not achieve a true plastic fusion of genre and style: the tropic incompatibility is resolved through a topological juxtaposition rather than hybridization.



Figure 6: ChatGPT Image, "A landscape in the style of Sam Francis" (January–February 2026).

2.2.2. The figure-ground relationship (portrait and abstraction)

AI-generated portraits exploit the figure-ground duality characteristic of the genre as a matrix for distributing generic and stylistic constraints: figurative formants are assigned to the figure, stylistic markers to the background. In portraits in the style of Morris Louis,[6] Midjourney resolves the trope by producing a reciprocal contamination between figure and background, creating zones of indeterminacy where the two co-determine each other at the plastic and formal level.

One solution relies on the visual resonance between the subject's body position and the background. Thus, the outline of a face in profile finds an

---

emerging from this dazzling whiteness while dissolving into it. His structural use of color is organized into flat-ribbon pairs.

[6] The American painter Morris Louis (1912–1962) explored color through the infiltration of pigment into the unprepared canvas. The swaths of color diffuse into translucent veils with undefined edges. This chromatic unfolding is based on a logic of impregnation where color becomes one with the support.

echo in the curvilinear chromatic bands of the background, which follow its directional variations. This morphological analogy ensures formal coherence without merging style and genre, while simultaneously subverting their ideal-types.

In another output, the clothing of a man depicted from the waist up dissolves into the surrounding pictorial material in an oblique orientation from left to right, while his sharply defined face is turned toward the right. This configuration suggests movement. The image thus sketches out a narrative in which the blurriness of the colored lines is attributed to the subject's movement: coherence is achieved at the cost of departing from the compositional architecture specific to the portrait.

For the ideal type of the portrait to be respected (Dondero 2020), four conditions would have had to be met.

1. The tension between the figure and the background should have been more balanced. The bright colors of the background overpower the neutral tones of the figures, which struggle to assert themselves. In the second image, the background goes so far as to partially absorb the silhouette of the man depicted.
2. The figures are not centered, contrary to the compositional rules of the genre.
3. The figure in the second output is not posing; the presence of blurred traces builds a sense of movement whereas the latter is normally absent from the typical characteristics of a portrait.
4. The subjects are depicted in side view. Although this pose is well established in the tradition, frontal composition is conventionally associated with the genre of portraiture, which is defined as the highlighting of an individual and distinctive face that establishes an exchange of gazes between the depicted subject and the viewer.

2.2.3. Mereology as an iconic resolution of the trope

The examples analyzed earlier demonstrate how multimodal GenAIs reuse topological configurations virtualized by visual genres. In our tests, a still life[7] generated by Midjourney in the style of Jackson Pollock[8] illustrates

[7] The still life is characterized by a multiplicity of objects in imbalance conveying a reflection on the instability of positions, the fragility of all things, and the deceptiveness of appearances, particularly through the reflective surfaces that create mirages. The image thus serves as a *memento mori* for ordinary mortals.

[8] The American Abstract Expressionist Jackson Pollock (1912–1956) invented the contourless line—projected and poured—that encloses nothing and outlines no figure. The hand



Figure 7: Midjourney, "Portraits in the style of Morris Louis" (January–February 2026).

another way of resolving the incommensurability between style and genre, relying not so much on topology as, more precisely, on the possibilities offered by mereology—that is, the relationship between the parts composing a coherent whole.

Figure 8 features dark green and red splashes concentrated above a vase placed on a table. The observer then tends to group these splashes of paint into a bouquet: this reconstruction, linking the parts into a whole, relies on our perceptual habits and our way of organizing visual experience. The image thus presents itself as a set of perceptible qualities that, once connected, actualize a unity of meaning virtualized by the composition.

We propose to characterize this form of resolution as *iconic*, drawing on the work of Jean-François Bordron. He defines iconicity as

> un type phénoménologique, c'est-à-dire une modalité particulière selon laquelle le monde sensible se manifeste à nous, qu'il s'agisse de la perception des objets, des états de chose [*sic*], des artefacts comme les images ou la musique. La caractéristique de l'iconicité réside moins dans la présence de morphologies, de parties et de totalités, de texture et de plasticité que dans le fait que tous ces éléments prennent ensemble.
>
> [a phenomenological type, that is, a particular modality through which the sensible world manifests itself to us, whether in the perception of objects, states of affairs, or artifacts such as images or music. The characteristic of iconicity lies less in the presence of morphologies, parts and wholes, texture, and plasticity than in the fact that all these elements come together.] (Bordron 2013: 70, our translation)

---

unleashes itself upon the canvas laid on the floor in a physical dance where tactile qualities take precedence over the visual.



Figure 8: Midjourney, "A still life in the style of Jackson Pollock" (January–February 2026).

Finally, stylistically speaking, although the simulated gestures evoke a chaotic and free gesturality, the organization remains centripetal: the lines converge toward the center and remain contained within the frame. This structuring reestablishes a compositional logic that Pollock's practice seeks precisely to dissolve.[9]

2.2.4. Contextual indexicality as a resolution of the trope

In several examples from the corpus of our tests, an image imitating an artist's style is juxtaposed with figurative elements specific to the genre. The model does not directly resolve the trope incompatibility, but circumvents it: the style becomes a localized property of an object in the metavisual form of a painting or stained-glass window present in the image, while the genre structures the rest of the scene.

Sometimes the painting embedded within the generated image is re-situated in the artist's studio. In one such occurrence, the studio scene incorporates still-life elements, while Pollock's style becomes legible on

[9] This observation echoes the findings presented in an article that adapts Wölfflin's criteria regarding classical and Baroque aesthetics to evaluate images generated by Midjourney and DALL·E (Deliège, Dondero, and D'Armenio 2025: 12). This article shows that, when asked to represent situations typical of the Baroque style, the models instead predominantly adopt a classical configuration with centered, closed, and regular forms, while only superficially—in the lighting, for example—can one identify features conventionally understood as Baroque.



Figure 9: From left to right: Midjourney, "A still life in the style of Jackson Pollock"; Nano Banana, "A still life in the style of Jackson Pollock"; Midjourney, "A still life in the style of Piet Mondrian" (January–February 2026).

both the depicted canvas and the work surface, both marked by gestures of applying painterly material that evoke the artist's sensorimotor movements.

Since the image within the image is placed in a setting resembling a painter's studio, the coexistence of genre and style appears less as a formal incompatibility than as a moment in the process of producing a work of art. The generated image presents itself through media characteristics that bring it closer to a photographic rendering in which a pictorial canvas is shown on a painterly surface among brushes and paint pots. Viewers can then construct a narrative establishing a relationship of implication: it is because there was an artistic performance that the objects are scattered and the paint is present everywhere. We could name this kind of composition a "tropic resolution by contextual indexicality" in order to designate the process by which the generated image is read as evidence of an artistic performance (Leone 2026).

These examples show that the outputs of generative models actualize configurations already virtualized in visual culture, with the tropic incompatibility potentially being perceived as resolved, or at least partially so. This aligns with the hypothesis regarding *stereotopy* ("stéréotopie", a portmanteau of *stéréotype* and *topos*) formulated by Pierluigi Basso Fossali (2026). When generative models are asked to produce outputs whose configurations, at first glance, find no attestation in databases, they are compelled to find the best possible solution by relying not on the *already-seen*, but on the *already-possible*: a set of virtual configurations that visual culture renders plausible without necessarily having realized them.

Finally, the multimodal GenAIs considered here reuse the center-periphery relationship established in landscapes in the style of Sam Francis; the figure-ground duality in portraits in the manner of Morris Louis; the



Figure 10: Midjourney, "A still life in the style of Jackson Pollock" (January–February 2026).

"coming together" of certain sensory qualities in a Pollockian still life; or the integration of a scene into a plausible pre-existing scenario, as in the case of a painting re-situated in Jackson Pollock's pseudo-studio.

There is a technical explanation for this: the latent space consists of vector representations (embeddings) learned from the regularities derived from the training data, which determine the image generation process. When a prompt requests the generation of a visual configuration not already learned by generative models—since it is absent from their training dataset—the models likely employ a combination (and thus a certain deformation) of the known vectors closest to the vector corresponding to the prompt in question.

Conclusion

To conclude, it is worth briefly reviewing the methodological contributions of the two protocols described, as well as their limitations.

The main contribution of the first protocol is its attempt to formalize perceptual judgments that would otherwise remain in the realm of individual intuition and tacit judgment. The combination of the theoretical frameworks of Greimas, Floch, and Wölfflin offers, in this regard, an anchoring in two complementary traditions, which reinforces its interdisciplinary scope. However, the three experts encouraged us to include a parameter evaluating figurative stereotypy, which can play a structuring role in considering the style of certain artists. Furthermore, the evaluation methods differ across

corpora: experts do not evaluate historical works and generated images under equivalent conditions, which may undermine their comparability. This dual approach nonetheless proved thought-provoking with respect to the typification processes carried out by humans, as distinct from those of the machine.

Regarding stereotypes, the first protocol involves observing the characteristics selected by Midjourney v6 to imitate a given movement or artist, while introducing a minimum of constraints in the prompt. This approach helps identify stylistic stereotypes as they take root in common sense: trained on data linking textual descriptions to image compositions, generative models infer formal regularities encoded as vector representations. These representations then condition the generation of new images, for which the models tend to update the features inferred during the training phase in the most statistically probable way.

The second protocol introduces an additional constraint into the prompt, which no longer consists solely of stylistic guidelines but also incorporates visual genres as variables. Unlike the first protocol, where the "average solution" is more easily predictable, tropes introduce a deviation that forces the model to generate a plausible and acceptable visual configuration, whose overall coherence aligns with a form of compromise solution (or middle ground).[10] This solution does not rely on an existing prototype (there is no still life by Jackson Pollock), but nevertheless produces a generic stereotype (a vase of flowers placed on a table). These results suggest that stereotypes and tropes are closely related: the trope relies on stereotypical elements which, although repurposed, remain recognizable, attesting to their structuring role in visual generation processes.

References

Agha, A. 2024. Enregisterment. In *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press, 1–29.

Basso Fossali, P. 2026 (in press). "Les affordances entre asymétries et stéréotypies dans l'image générée par IA". *Interfaces numériques*.

Bordron, J.-F. 2013. *Image et vérité: essai sur les dimensions iconiques de la connaissance*. Liège: Presses Universitaires de Liège.

Carnap, R. 1947. *Meaning and Necessity: A Study in Semantics and Modal Logic*. Chicago: University of Chicago Press.

[10] We thank Flavia Politi for raising this point of discussion.

D'Armenio, E., M. G. Dondero, A. Deliège and A. Sarti. 2025. "For a Semiotic Approach to Generative Image AI: On Compositional Criteria". *Semiotic Review*, Vol. 9, 1–59. Available at: https://doi.org/10.71743/ee5nrx33.

Deleuze, G. 2023. *Sur la peinture. Cours mars–juin 1981*. Édition préparée par David Lapoujade. Paris: Les Éditions de Minuit.

Deliège, A., M. G. Dondero and E. D'Armenio. 2025. "Revisiting Wölfflin in the Age of AI: A Study of Classical and Baroque Composition in Generative Models". *Journal of Imaging*, Vol. 11, No. 5, Article 128. Available at: https://doi.org/10.3390/jimaging11050128.

Deliège, A., J. Marlot, M. Van Droogenbroeck and M. G. Dondero. 2025. "How Good Is the Machine at the Imitation Game? On Stylistic Characteristics of AI-Generated Images". *Journal of Imaging*, Vol. 11, No. 12, Article 429. Available at: https://doi.org/10.3390/jimaging11120429.

Dondero, M. G. 2009. *Le sacré dans l'image photographique*. Paris: Hermès.

Dondero, M. G. 2020. *The Languages of the Image: The Forms and the Forces*. Cham: Springer.

Dondero, M. G. 2025a. "Semiotics of Artificial Intelligence: Enunciative Praxis in Image Analysis and Generation". *Semiotica*, Vol. 2025, No. 262, 111–146. Available at: https://doi.org/10.1515/sem-2024-0195.

Dondero, M. G. 2025b. "Enunciative Praxis and Enregisterment in the Domain of Generative Artificial Intelligence". *Semiotic Review*, Vol. 12. Available at: https://doi.org/10.71743/bwf80g96.

Dondero, M. G. 2026a. Enunciating Pictorial Styles via Generative Artificial Intelligence: Towards a Semiotics of Stereotypes. In Peverini, P., D. Mangano (eds.). *Semiotics in the Age of Artificial Intelligence: Symbols, Discourses, Consumption, Branding*. Cham: Springer, 133–149. Available at: https://doi.org/10.1007/978-3-032-29272-8_7.

Dondero, M. G. 2026b. "The Compositional Rhetoric of Multimodal Generative AI: About Tropes and Stereotypes". *Signata*, Vol. 17. Available at: https://doi.org/10.4000/16qma.

Eco, U. 1999 [1997]. *Kant and the Platypus: Essays on Language and Cognition*. (transl. Alastair McEwen). New York: Harcourt Brace. (First edition: *Kant e l'ornitorinco*. Milan: Bompiani.)

Floch, J.-M. 1985. *Petites mythologies de l'œil et de l'esprit: Pour une sémiotique plastique*. Paris and Amsterdam: Éditions Hadès-Benjamins.

Frege, G. 1948 [1892]. "Sense and Reference". (transl. Max Black). *The Philosophical Review*, Vol. 57, No. 3, 209–230. Available at: https://doi.org/10.2307/

2181485. (First edition: "Über Sinn und Bedeutung". *Zeitschrift für Philosophie und philosophische Kritik*, Vol. 100, No. 1, 25–50.)

Greimas, A. J. 1989 [1984]. "Figurative Semiotics and Plastic Semiotics". *New Literary History*, Vol. 20, No. 3, 627–649. Available at: https://doi.org/10.2307/469358.

Groupe μ. 1992. *Traité du signe visuel: Pour une rhétorique de l'image*. Paris: Seuil.

Kripke, S. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Leone, M. 2026 (in press). "Quasi-Markovian Semiosis: Indexicality, Memory, and Large Language Models". *Semiotic Review*, Vol. 13.

Marlot, J. 2026. Des images qui ont du style ? Vers une méthodologie d'analyse des stéréotypes stylistiques (re)produits par les IAG multimodales. Paper presented at the International Semiotics Seminar in Paris, *Entre automatisme et automatisation des pratiques langagières: pour une nouvelle sémiotique du stéréotype*, 25 March 2026. Available at: https://www.youtube.com/watch?v=xczCXU1xwcA (accessed 1 April 2026).

Nakassis, C. V. 2025. "Person-Indexing Registers, Stardom, Auteurism". *Signata*, Vol. 16. Available at: https://doi.org/10.4000/14rhi.

Wölfflin, H. 1950 [1915]. *Principles of Art History: The Problem of the Development of Style in Later Art*. (transl. M. D. Hottinger). New York: Dover Publications. (First edition: *Kunstgeschichtliche Grundbegriffe: Das Problem der Stilentwicklung in der neueren Kunst*. Munich: F. Bruckmann.)

Zilberberg, C. 2012. *La Structure tensive, suivi de Note sur la structure des paradigmes et de Sur la dualité de la poétique*. Liège: Presses Universitaires de Liège.