A Semiotic Analysis of Generative AI Multimodality: An Epistemological Inquiry for an Interpretation and Evaluation Model[†]

Sung Do Kim[*]

http://doi.org/10.38119/cacs.2025.36.1
Received November 27, 2025; accepted December 19, 2025; published online December 31, 2025

Abstract: Generative AI confronts semiotics with a new kind of sign-producing machine that actively reshapes the production and interpretation of visual content. Addressing the lack of humanities-based transdisciplinary research on this transformation, this study aims to establish a methodological foundation for the semiotic analysis of multimodal AI. By combining visual, social, quantitative, and multimodal semiotics, the paper proposes an integrated micro–meso–macro framework for evaluating AI-generated images. The analysis moves from the micro-level examination of plastic features and text-to-image translation, through the meso-level of enunciation, narrativity, and causality, to the macro-level of social stereotypes, ideology, creativity, rhetoric, truth, and inference. This is supported by a case study on lonely death and a semiotic explanation of latent space.

Keywords: generative AI, multimodality, visual semiotics, social semiotics, quantitative semiotics, latent space

1. Introductory remarks: Context and aim of inquiry

Academic interest in artificial intelligence has expanded dramatically across all disciplines since the rise of generative models, moving far beyond the area of computer science and engineering (Crawford, 2021; Baron, 2023; Danesi, 2024). These models may appear as incremental improvements from an engineering perspective. Yet the implications for humanities—and especially for linguistics, semiotics, and philosophy—are far-reaching and in many ways revolutionary (Floridi, 2023; Heersmink et al., 2024; Silvestri, 2024). Generative AI, and in particular multimodal AI, confronts semiotics with a new kind of sign-producing

[†] This study was supported by a faculty research grant from the College of Liberal Arts at Korea University in 2025
[*] **Sung Do Kim**, Korea University, Republic of Korea, E-mail: dodo@korea.ac.kr

machine, capable of operating across images, text, sound, and other sensorily coded data. In this sense, generative AI multimodality does not merely add another set of tools to the existing repertoire of media technologies; it actively reshapes the conditions under which signs are produced, distributed and interpreted.

The current situation is marked by a contrast between the ubiquity of AI-generated images and the relative absence of solid humanities-based transdisciplinary research capable of accounting for this transformation. In Korea, most humanities fields—including semiotics, which is centrally concerned with images—often lack the technical equipment and quantitative methods needed to analyze the new image culture formed by large-scale AI-generated visual corpora. Research on the semiotic structure (namely, syntactic, semantic and pragmatic dimensions) of such images remains limited, particularly with regard to methodologies that could systematically track how AI images reconfigure social meaning in visual communication. This gap is especially problematic given the role of AI in both the production and reception of images: users now spend several hours per day composing, curating, and scrolling through visual content, much of which is either directly generated by AI or indirectly shaped by AI-driven recommendation systems. In this sense, the consequences of multimodal AI are central to the ongoing reshaping of digital sociality.

The global multimodal AI market, valued at around $1.2 billion in 2023 and projected to grow at a compound annual growth rate of roughly 33 percent between 2024 and 2030, offers a quantitative index of this transformation by indicating the economic and industrial attention now invested in multimodal systems (MarkNtel Advisors, 2024). Across digital social networks, users produce and consume unprecedented quantities of images, and generative AI accelerates this process by enabling the rapid production of visual content. Platforms such as Facebook, Instagram, Snapchat, and Tinder constitute a new image ecosystem in which cognitive, aesthetic, emotional, and pragmatic values attached to images are rapidly shifting under the pressure of multimodal AI. This new ecosystem is increasingly shaped by AI services that mediate not only content creation but also distribution and visibility (Bommasani et al., 2021; Dwivedi et al., 2023).

Recent works in European semiotics have begun to develop methodological and epistemological tools for dealing with this situation by focusing on data corpora and generative AI models, including large language models and diffusion-based image generators. Danesi (2024), for example, examines how generative models recognize genres, narrative patterns, and visual tropes. One central concern in these studies is translation—between human prompts and machine-internal prompts, between natural language descriptions and generated images, and between images and textual descriptions produced by captioning systems. A special issue of *Semiotica* examines the multiple enunciative instances and agencies involved in generative AI, from the databases used for training and inference, to the digital representations that convert visual or verbal texts into numerical vectors, and finally to the prompts that act as user-facing entry points into these models (Dondero et al., 2025).

As stated by D'Armenio et al. (2024), we are convinced that it is essential for post-structuralist semiotics to study artificial languages, technologies and practices for automating human actions, as these offer tools that attempt to simulate, in an increasingly successful manner, the specificity of language and human practices. Leone (2023) also mentions that the

new task of semiotics is to investigate how AI systems construct digital representations, thereby redefining what counts as truth in contemporary society.

Generative AI is thus defined as a sign-producing machine and semiotic technology, whose outputs—multimodal compositions including images and texts—must be analyzed as systems of meaning rather than as mere engineering outputs. This study, therefore, pursues three main aims. First, it seeks to establish an epistemological and methodological foundation for the semiotic analysis of generative AI, with particular attention to multimodality. Second, it aims to explain the operative principles of generative AI—especially text-to-image systems—through the lens of semiotics. Third, it proposes an integrated micro–meso–macro model for the analysis and evaluation of AI-generated images.

2. Materials and methodology

We rely on LMArena's Text-to-Image Arena as a neutral benchmarking platform to navigate the rapidly growing ecosystem of image-generation models and to select those that best fit our experimental needs (Figure 1). Based on its crowdsourced leaderboard of text-to-image systems for the first week of December 2025, we chose three competitive models: Gemini (Nano Banana), Flux, and Seedream. We also included ChatGPT (formerly DALL·E) as the most known and used generative AI, while excluding Hunyuan, whose limited documentation at the time made it difficult to integrate into the study.

Rank ⇅	Rank Spread (Upper-Lower)	Model ⇅	Score ↓	95% CI (±) ⇅	Votes ⇅	Organization ⇅	License ⇅
1	1 ↔ 2	gemini-3-pro-image-preview-2k (nano-banana-pro)	1234	±10	5,256	Google	Proprietary
2	1 ↔ 2	gemini-3-pro-image-preview (nano-banana-pro)	1231	±7	26,043	Google	Proprietary
3	3 ↔ 7	flux-2-flex	1157	±7	11,497	Black Fores...	Proprietary
4	3 ↔ 5	gemini-2.5-flash-image-preview (nano-banana)	1156	±4	613,073	Google	Proprietary
5	3 ↔ 8	hunyuan-image-3.0	1154	±5	77,437	Tencent	tencent-hu...
6	4 ↔ 9	flux-2-pro	1146	±6	14,499	Black Fores...	Proprietary
7	4 ↔ 9	seedream-4.5	1146 Preliminary	±7	8,806	Bytedance	Proprietary

Figure 1. LMArena's Text-to-Image Arena

Our proposed methodological framework is a combination of four semiotic approaches: visual semiotics, social semiotics, quantitative semiotics, and the semiotics of multimodality (Figure 2). Structuralist visual semiotics, initiated by Roland Barthes, elaborated by Algirdas J. Greimas, consolidated by Groupe μ, and deepened and applied by Jean-Marie Floch, Jacques Fontanille, Félix Thülemann, and Anne Beyaert-Geslin, offers tools for analyzing how shape, color and topological relations compose meaning in still images (Greimas, 1984; Floch, 1990; Thürlemann, 1982, Hénault & Beyaert-Geslin, 2004). In this tradition, an image is treated as a network of discontinuities that must be topologized. Physical space is reconstructed as a semiotic space in which borders, gradients and focal points become interpretable entities, before chromatic and eidetic categories are deployed. Although visual semiotics often favors relatively

stable images and sometimes presupposes fixed correlations between visual features and values, its concepts of spatial articulation and chromatic structuring remain crucial for thinking about feature extraction, foreground/background relations and compositional hierarchies in digitally generated images.

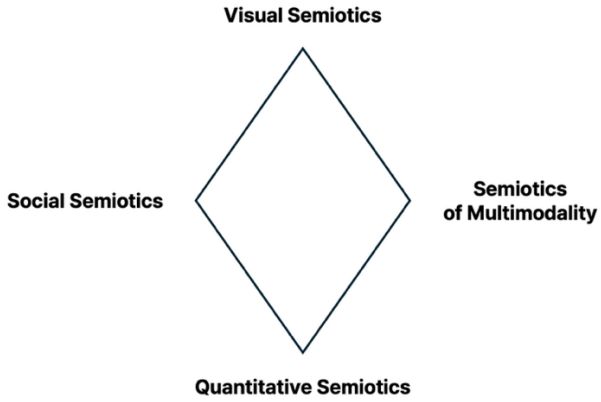


Figure 2. The four semiotic approaches to AI multimodality

Social semiotics, including both critical discourse analysis and multimodal critical discourse analysis, shifts attention from the internal structures of images to the socially situated practices and semiotic resources that produce and interpret them (Kress & van Leeuwen, 1996; Kress, 2010). Social semiotic studies extend these concerns to the digital and AI technologies involved in meaning-making, insisting that not only texts but also software, platforms, and interfaces constitute semiotic objects whose design embodies assumptions about users, communities, and permissible forms of expression. These concerns can be illustrated by two recent case studies. A study in AI-generated images of dementia shows how such systems can reproduce biomedical discourses when prompted with diagnostic labels, reinforcing narrow and often dehumanizing visual scripts (Putland et al., 2025). Social semiotic analyses of youth as vulnerable social media users likewise demonstrate how multimodal representations construct young people as at-risk subjects in need of surveillance and protection (Liang & Lim, 2024).

The semiotics of multimodality, which is closely related to semiotics of intermediality and hypermediality (Bolter & Grusin, 1999; Kryssanov & Kakusho, 2005; Rajewsky, 2005; Elleström, 2010) offers a broader insight into how AI-generated images participate in multimodal communication. Here, multimodality is conceived as an approach to communication drawing insights from semiotics, linguistics, rhetoric, anthropology, media studies, and literacy studies. Rowsell's recent proposal to destabilize multimodality by foregrounding five key elements—the motivated sign, agency, temporality, transmodality and modal density—offers a particularly useful set of questions for our analysis (Rowsell, 2023).

Finally, quantitative semiotics seek to utilize semiotic concepts in computable form and to test them on large datasets of signs. Approaches that model entropy, distributional patterns and complexity in sign systems show how statistical analysis can be used to investigate semiosis in hypermedia and other complex communicative environments (Compagno, 2018; Takács, 2012; Lacková & Faltýnek, 2021).

3. Semiotic and epistemological foundations of AI generative multimodality

3.1 Overview of multimodal AI

Multimodal artificial intelligence designates systems capable of processing and integrating several different kinds of data—images, sounds, written text, perhaps other signals—within a single architecture so that the system can perform tasks that would be impossible for a model restricted to a single modality. In the vocabulary of machine learning, a "modality" is simply a type of data; multimodal AI combines several such types at once. A typical architecture is composed of three parts. An input module consists of multiple unimodal neural networks, each specialized for a given kind of input and collectively responsible for transforming heterogeneous data into internal numerical representations. A fusion module then takes these unimodal encodings and processes them together, aligning, weighting, and combining information from each modality. Finally, an output module produces the system's response, which might be a text, an image, a piece of audio, or some combination thereof. This basic pipeline can be instantiated in many directions—text-to-image, text-to-audio, audio-to-image, and various mixtures of spoken, written, and visual symbols—yet what matters for the present analysis is that all of these variants rely on similar operating principles. For that reason, the following discussion concentrates on text-to-image systems as a privileged case through which general features of multimodal AI can be approached.

Text-to-image models were first conceived as diffusion systems that start from random noise and gradually transform that noise into a coherent image. To address the lack of directionality in such purely stochastic procedures, developers introduced textual descriptions that guide the diffusion process, allowing particular words—"cat," for instance—to be associated with recognizable visual symbols such as an image of a cat. In these systems, both text and images are converted into mathematical vectors, typically through separate encoders trained on large image–caption datasets. Each encoder produces a vector representation for its input, and the model learns to align and correlate these pairs of vectors so that linguistic descriptors can reliably evoke appropriate visual compositions. Audio-to-image conversion generally follows a more complex route: there is, at present, no widely used model that directly transforms speech into images in a single step. Instead, audio is first transcribed into text, which is then processed by a text-to-image model that generates the corresponding visual output. The apparently simple gesture of typing a sentence and receiving a picture thus presupposes an intricate chain of encodings, alignments, and decodings that span several modalities at once.

The semiotic implications of these architectures are best elucidated by analyzing the elementary tasks they perform in computer vision and generation. Image classification and object detection rely heavily on Convolutional Neural Networks (CNNs), which function through a hierarchical process of feature extraction—detecting simple patterns like lines in early layers and complex objects in deeper layers—followed by classification (Figure 3).

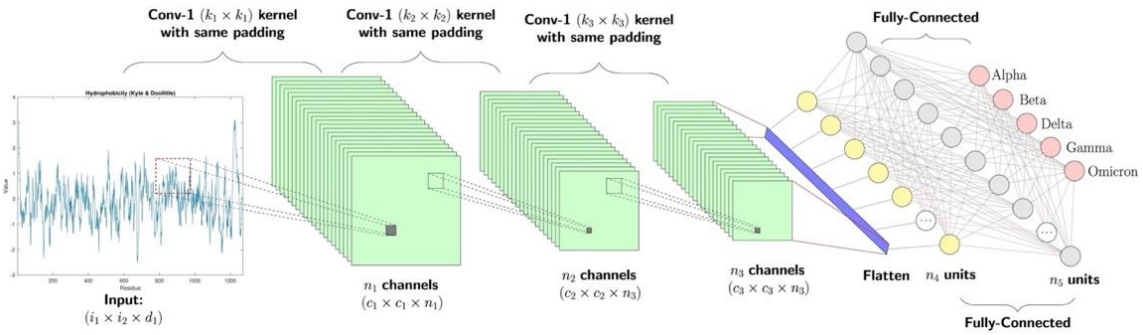


Figure 3. A Convolutional Neural Network (CNN) architecture showing feature extraction and classification layers. (Somaini, 2023).

Image generation has advanced from Generative Adversarial Networks (GANs), defined by a zero-sum game between Generator and Discriminator (Somaini, 2023), to large-scale text-to-image diffusion models and, more recently, to multimodal systems such as ChatGPT, Nano Banana, Flux, and Seedream. Alongside visual question answering (VQA), which integrates language and vision by requiring models to answer textual questions about images, these systems perform a shared repertoire of operations—classification, localization, reasoning, and synthesis—that now structure everyday interactions with AI (Figure 4). From a semiotic standpoint, however, their significance lies less in these engineering capacities than in how they transform the conditions under which signs are produced and, consequently, related and interpreted.

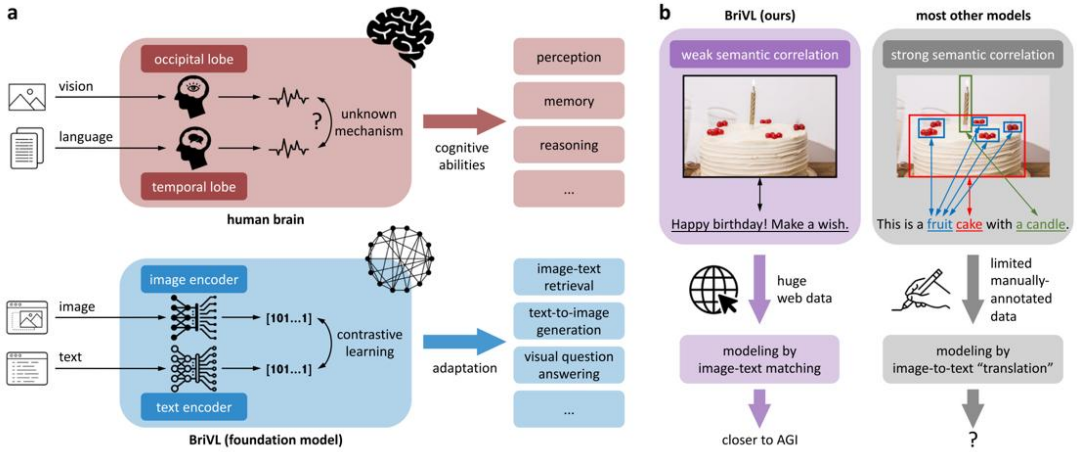


Figure 4. Detailed working principle of AI multimodality (Fei et al., 2022)

These generative systems can be described as operating through stochastic discrimination rather than any grammatical structure. Rather than organizing meaning through semantic rules, they depend on statistical computation across vast training datasets, producing what Hito Steyerl (2023) calls statistical renderings. The resulting AI-generated images arise from both the architecture of the latent space and the statistical and predictive operations carried out within it. Somaini (2024) refers to them as visualizations of latent space: not because latent space itself can be visualized—its extremely high dimensionality prevents this—but because latent space plays a constitutive role in their generation. As a consequence, the dependence on textual

prompts introduces a distinctive semiotic condition in which the visible becomes strictly correlated with the sayable. In this emergent visual culture, images and words are tightly interlinked, giving rise to a language-biased mode of visual production in which the limits of the prompt effectively delimit the boundaries of the visible. (Somaini, 2024)

One of the striking claims made about generative AI in this regard is that it can be described as "anti-grammar." Rather than organizing meaning through a system of structural constraints, such systems rely on statistical computation across vast corpora of data. Some scholars argue that generative AI has no genuine grasp of concepts, structures, or theories of meaning and no mechanism for taking human interests into account as meaning; on this view, what appears as complexity is in fact limited only by the size and heterogeneity of the training data on which the underlying calculations are based. At the same time, the fact that many AI images are produced in response to textual prompts introduces a distinctive semiotic technology that sets such images apart from photography or drawing, both of which can operate independently of language. Insofar as the production of images is conditioned by verbal prompts, generative systems may be said to foster a language-biased form of visual communication, and this language dependency requires careful analysis.

3.2 Epistemological features of latent space

Another important dimension of multimodal generative AI emerges from analysis of latent space, the multidimensional abstract space into which deep learning algorithms convert digital objects—images, texts, and other data—into so-called latent representations (Somaini, 2023, p. 77). Latent space is a compressed representation of data points that captures only the essential characteristics underlying the input; each dimension corresponds to a latent variable, an underlying feature that shapes the distribution of data even if it is not directly observable. The passage from input to output typically passes through a "bottleneck," a compressed segment of latent space that is constructed by an encoder and then expanded by a decoder in the direction of the output. Reconstruction loss measures the discrepancy between input and output, and the goal of training is to minimize this loss so that the reconstructed output is "correct" with respect to the task at hand.

Somaini (2024) defines latent space as a multidimensional domain that is imperceptible and essentially unimaginable—a space composed of vectors, a vast numerical matrix that cannot be adequately represented through two- or three-dimensional visualization. Within deep learning, this space functions as a compressed form of knowledge representation. Encoding into latent space is therefore understood as a process of compression itself (Somaini, 2024). According to the manifold hypothesis, high-dimensional real-world data are distributed along a lower-dimensional manifold embedded within a larger dimensional space (Figure 5). Latent space thus captures the essential, underlying structure of the data rather than their redundant variation or noise. Operationally, the transformation from input to output is mediated by an architecture typically composed of an encoder and a decoder. The encoder compresses the input into a low-dimensional vector by passing it through a bottleneck layer. This bottleneck is critical: by constraining the flow of information, it compels the network to extract the most salient features—the latent variables—rather than memorizing the input. The objective of training is

to minimize reconstruction loss so that the decoder can accurately regenerate the output from this compressed representation.

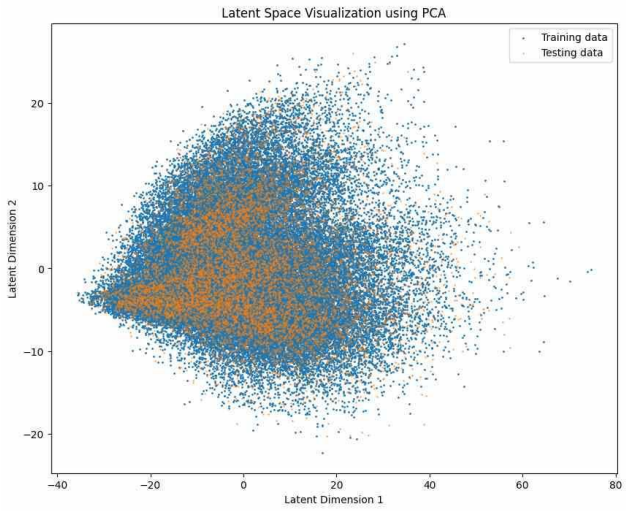


Figure 5. Latent space in deep learning. (GeeksforGeeks, 2025)

To better understand how symbolic properties can be manipulated within latent space, work on GANs has proposed explicit semiotic frameworks (Figure 6). In the widely cited introduction of GANs, Goodfellow and colleagues showed how a generator and a discriminator could be trained in opposition: the generator attempts to produce samples from latent space that mimic real data, while the discriminator learns to distinguish real from generated samples, providing feedback that allows the generator to improve over time (Goodfellow et al. 2014). Later research, such as GANalyze, demonstrated that particular directions in latent space could be associated with changes in abstract properties of images, such as mood or style.

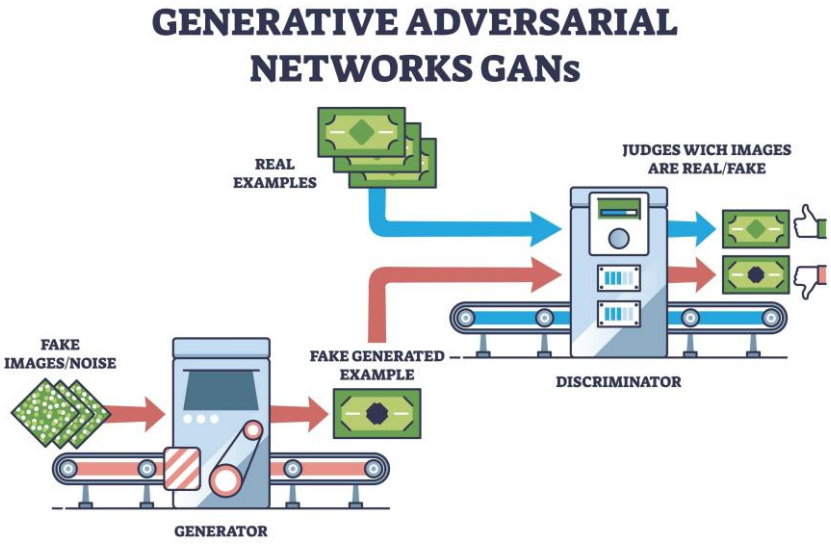


Figure 6. Overview of GAN

Building on this insight, Osmany proposes a three-component architecture composed of a Semiotician (S), a Generator (G), and a Transformer (T) (Osmany, 2023) (Figure 7). The Semiotician is a neural network classifier that evaluates the degree to which a given image manifests a particular semiotic attribute—how "happy," "minimalist," "surreal," "tragic," or "radiant" it is—and outputs a numerical score. The Generator, given a latent vector z and a class label y, produces an image G(z, y). The Transformer learns to modify z along a symbolic direction θ in latent space so as to enhance or weaken the attribute scored by the Semiotician. In this S–G–T framework, the Transformer adjusts the latent vector according to a rule of the form $z + \alpha\theta$, where α is a parameter controlling the magnitude of the change. Positive values of α intensify the attribute, while negative values attenuate it. By comparing the Semiotician's scores before and after transformation—say, from 0.2 to 0.8 on a "radiant" scale—one can quantify how much the symbolic attribute has been amplified. The approach thus implements an evaluative model that measures the degree of semiotic properties in generated images through classifier scores. For example, two images derived from the same initial latent vector can be differentiated by selectively enhancing an "evil" attribute in one and suppressing it in the other, and the semantic distance between them can be assessed first through the Semiotician's probability outputs and then through human judgments. Latent spatial shifts, in this view, give rise to new visual concepts and reshape the distribution of generated images, revealing how the semantic space of potentially infinite generative outputs can be explored and expanded.

Figure 7. Overview of GAN and Osmany's SGT, Generated by Gemini 3.0.

This framework situates GANs at the representational level, the Semiotician at the conceptual level, and the human interpreter at the level of synthesized meaning. The interaction between these levels becomes the key to constructing a semiotic evaluation model for generative images. Machine processes in latent space can thus be understood as operations of data interpretation, even if they operate on compressed numerical representations and latent variables. The hypothesis is that generative processes draw on and give meaning to multiple

spatial layers that do not simply "contain" operations but are themselves shaped and reshaped by the interplay of human and machinic interventions (Colas-Blaise, 2025a).

4. Microanalysis of AI-generated images

4.1 Plastic analysis of AI-generated images

Plastic categories, as developed in Paris School visual semiotics, provide a framework for analyzing how visual meaning is organized in contemporary generative AI (Morra et al., 2024). These categories concern everything in a visual composition that cannot be reduced to object recognition, focusing instead on the organization of the surface within a frame. Topological categories structure space through oppositions such as left versus right, top versus bottom, and center versus periphery, and through directional forces that draw perception centripetally toward the center or centrifugally toward the margins. Chromatic categories articulate differences and similarities in saturation and brightness, establishing systems of contrast and affinity among colors. Eidetic categories concern the outlines of shapes, which may be linear or curvilinear, continuous or fragmented. Taken together, these dimensions allow a detailed description of how an image is composed and how positions, colors, and forms interact to produce meaning (Groupe μ, 1992; Floch, 1985). In this sense, plastic categories make it possible to evaluate how textual prompts are translated into visual structures in multimodal operations. Textual descriptions are under-specified with respect to spatial layout, color, and geometric articulation. A prompt that names a color, a form, or a spatial relation does not determine how the surface will be segmented, how attention will be distributed between center and periphery, or how figures and ground will be related (Figure 8).

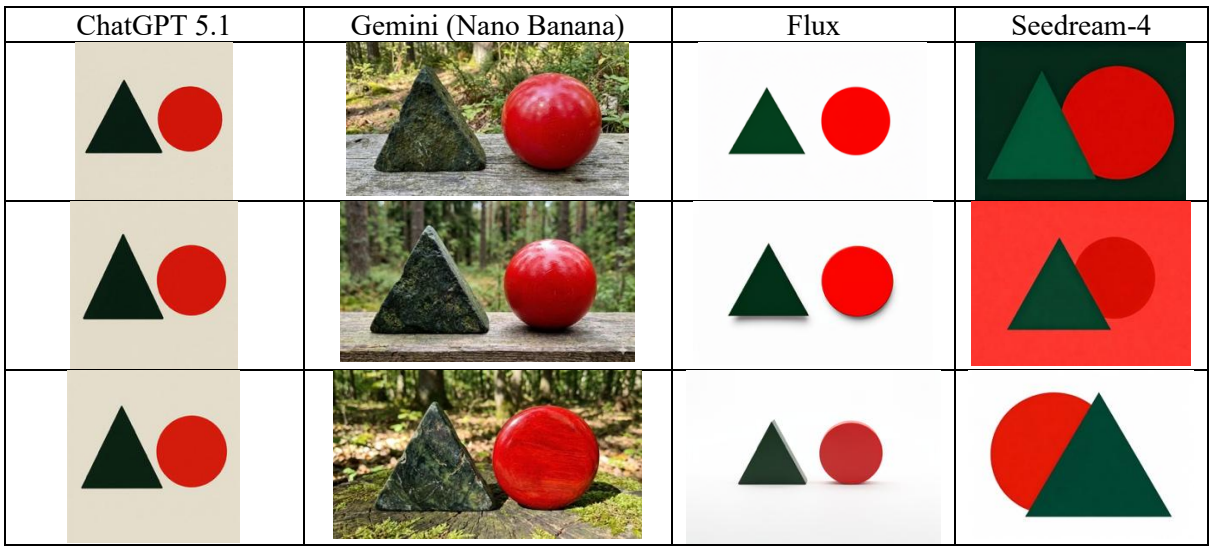


Figure 8. Outputs from the prompt: "a dark green triangle next to a bright red circle"

The previously mentioned concept of latent space shifts attention from visible images to the abstract domains where they are encoded and transformed. From a semiotic viewpoint, these operations construct sketches of sensible qualities on the expression plane, which may later be

associated with contents to produce new semiotic functions. Perception, in Bordron's sense, consists in the production of such sketches, while Basso Fossali stresses that perceptual activity continually destabilizes and reorganizes standard relations, generating unfinished forms open to comparison and rectification (Bordron, 2011; Basso Fossali, 2017, as cited in D'Armenio, 2025, p. 227). Synthetic perception in artificial systems reproduces this dynamic computationally, organizing multimodal corpora in high-dimensional spaces where proximity and distance encode potential associations among descriptors and visual traits.

Within this ecology, latent space functions as a virtual archive for models such as Nano Banana, in which "style ceases to be a historical category and becomes a pattern of visual information to be extracted and monetized" (Meyer, 2023, as cited in Dondero, 2025, p. 136). 'Photographic' also becomes one style among others, and photorealistic outputs no longer reconstruct three-dimensional reality according to optical laws but recombine surface textures and appearances. What appears are images of images, filtered through linguistic descriptors and embedded in circuits of intersemiotic translation.

4.2 Multimodal analysis of AI-generated images: Translation from text to image

Text-to-image models can be described as multimodal systems that turn written prompts into images by drawing on large training corpora and internal latent representations. From a semiotic point of view, the key issue is how verbal prompts are transformed into visual arrangements, which we treat as a specific case of intersemiotic translation in Jakobson's sense (Jakobson 1959, as cited in D'Armenio et al., 2025, p. 5). Verbal language organizes meaning through predication and relatively explicit syntactic structures, whereas visual meaning emerges from the positioning and mutual adjustment of elements across a surface. What Bordron describes as the tabularity and mereology of images—relations between parts, and between parts and the whole—highlights that visual composition follows a spatial and mereological logic rather than a linear one (Bordron, 2011, as cited in D'Armenio et al., 2025, p. 5).

The problem becomes more intricate when prompts describe relations rather than isolated properties. A simple phrase implies decisions about the body's physiognomy and gendering, the pose of torso, legs, and back, stylistic and material conventions, and the viewing angle. Prompts that encode actions add another layer of complexity. On one axis, we can distinguish between the absence of action, simple directed actions without objects, and more complex events involving multiple participants and objects. On another axis, we can differentiate aspectual values: completed events, beginnings of actions, and processes in progress. A description such as "a dog fetches a tennis ball from the ground," for example, presupposes an actor, an object, and a ground plane whose spatial and kinetic relations must be made visible (Figure 9).

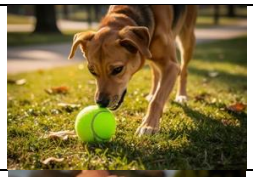
ChatGPT 5.1	Gemini (Nano Banana)	Flux	Seedream-4
			
			
			

Figure 9. Outputs from the prompt: “a dog fetches a tennis ball from the ground”

5. Mesoanalysis of AI-generated images

5.1 Enunciation of AI-generated images

Enunciation in visual semiotics designates the way an image organises relations among spaces, times, and actors in order to establish a position for a viewer and for represented characters (Fontanille, 1989; Dondero, 2020). One central distinction derives from Benveniste’s opposition between *discours* and *histoire*: in verbal language, discursive enunciation marks the presence of an ‘I’ addressing a ‘you,’ whereas historical enunciation effaces such deictic anchoring (Benveniste, 1966). In visual language, this contrast is reconfigured through the gaze. When a represented figure looks toward the viewer, an ‘I–you’ dialogical configuration is enacted; when figures do not address the spectator, events unfold in a space-time presented as not shared with the viewer (Figure 10). Enunciative effects are further mediated by metapictorial devices analysed in art history, such as windows, mirrors, doors, curtains, and niches, which structure the gaze by inviting it to see beyond, multiply points of view, produce effects of discovery, or block the horizon and create tension toward the viewer’s space (Stoichita, 1997) (Figure 11).

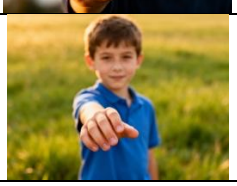
ChatGPT 5.1	Gemini (Nano Banana)	Flux	Seedream-4
			
			

Figure 10. Outputs from the prompt: “a boy extending his hand towards us”

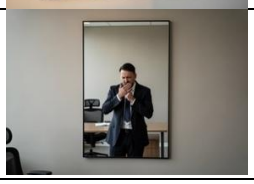
ChatGPT 5.1	Gemini (Nano Banana)	Flux	Seedream-4
			
			
			

Figure 11. Outputs from the prompt: “a mirror in a room reflecting a businessman yawning”

The rise of generative artificial intelligence raises the question of how such enunciative configurations are distributed between human and machine. Colas-Blaise asks whether machines are merely tools or prostheses, whether they can be treated as over-enunciators, under-enunciators, or co-enunciators, and whether they not only create but also truly express themselves (Colas-Blaise, 2025b). Her answer is that machines “(co)create, without necessarily (co)enunciating,” since they do not attain the reflexive position of a subject of enunciation capable of evaluating and commenting on its own semiotic production a posteriori (Colas-Blaise, 2025b). On this basis she proposes a “machinic enunciative sequence” in which the human subject of enunciation occupies the initial position—programming, providing databases, issuing commands and prompts—and the final position—selecting artefacts, evaluating, and interpreting—while machinic instances confront one another in the middle phase (Colas-Blaise, 2025b). Enunciation is thereby elevated to the status of a metalanguage for the digital, clarifying how responsibilities and initiatives are distributed among heterogeneous agents.

Dondero extends enunciation theory through the concept of enunciative praxis, useful for understanding the cultural process by which new forms are produced, stabilized, and eventually disappear (Dondero, 2025). This praxis can be operationalized through four modes of existence: virtualization, actualization, realization, and potentialization. Large image databases correspond to virtualization, encompassing all digitized and recognized images available to be studied, mixed, or recombined; actualization concerns the skill to produce new images through the functioning of databases in relation to prompts and embeddings (Dondero, 2025, p. 113). Constant learning explains why generative systems produce different content when prompts change, fostering an impression of infinite creativity noted by Somaini (2023). Within this

process it remains possible, at least in principle, to identify generative marks that signal human–machine co-creation.

These theoretical perspectives converge on a conception of generative AI as a site where enunciation is redistributed rather than simply automated. Human agents remain the primary subjects of enunciation, initiating and ratifying semiotic acts, while machinic systems constitute powerful but non-reflexive co-creators whose operations are embedded in databases, algorithms, and interfaces. Visual enunciation offers tools for tracking how gazes and metapictorial devices construct relations between represented figures and viewers; enunciative praxis and machinic enunciative sequences, in turn, specify how these configurations arise from chains of human–machine interaction. At stake is not only the technical capacity to generate images, but a broader reorganization of authorship, responsibility, and interpretive labor in contemporary visual culture.

At the meso-level, future work should also address narrativity, causality, and abstractness. From a semiotic perspective, the anthropological implications of creativity and intelligence show that human culture is rich in symbols, myths, narratives, and other elements that form the basis of our understanding of the world and our innovation, whereas AI that lacks an intrinsic cultural context may operate on a purely functional level, lacking the symbolism and narrative depth that characterize human creativity. Narrative is a key element of human creativity, enabling us to make sense of experiences and communicate complex ideas effectively. Another specific theme of this research is to explore the fundamental limitations of how state-of-the-art AI image generation models respond to different types of visual meaning, and how they understand the relationship between prompts and the images they generate. AI image generation models show great promise in simulating high-quality images, but they tend to fail in subtle and strange ways, for example by generating a hand with six fingers or text that does not resemble any language.

6. Macroanalysis of AI-generated images

6.1 Social stereotypes and ethics in AI-generated images

Macro-level analysis of the ethical problems of generative AI is necessary because, as Hagendorff's (2024) scoping review shows, these systems concentrate a wide range of novel normative problems—from bias, harmful content and hallucinations to privacy leaks, interaction risks, security, alignment and labour impacts. Critical Discourse Analysis (CDA) and Multimodal Critical Discourse Analysis (MCDA) provide a framework for examining how AI-generated images show specific social interests and represent particular realities, while anchoring these representations in wider processes of recontextualizing social practices. Within this framework, three elements of representation are especially salient. Participants in social practice may be individuals or collectives, generalized or specific, homogenized or individualized, and can be visually categorized through cultural markers such as clothing, religious objects, or stylistic traits, as well as physiological characteristics including skin color, hair, and body shape. Behavior comprises actions and emotions and is treated in social semiotics as the visible trace of social practice, a system of indicators registering what participants do and

feel. Time and place refer to the material settings in which practice unfolds, specifying when and where participants are situated: indoors or outdoors, in particular architectures, seasons, weather conditions, or times of day. Together, these dimensions provide a systematic basis for relating multimodal representations to the social practices they reframe. Drawing on social semiotics (van Leeuwen, 2005), and more specifically on semiotic technology studies (Poulsen et al., 2018) and critical discourse work on multimodality (Machin, 2013, 2016; van Leeuwen, 2008), we aim to show how AI represents lonely death (Figure 12).

Figure 12. Outputs from the prompt: “lonely death”

In the case of Gemini, we compiled a small dataset of 50 AI-generated images depicting lonely death and conducted a basic case study. At the level of recurring motifs, windows appear in all 50 images (100%), while teacups occur in 42 images (84%) and stacks of unopened letters in 36 images (72%). Television sets are visible in 14 images (28%), trash or clutter in 20 images (40%), flowerpots in 13 images (26%), photographs in 10 images (20%), pets in 1 image (2%), and walking aids in 8 images (16%). Regarding participants, human figures are explicitly depicted in 10 images (20%), whereas 40 images (80%) contain no visible person. In terms of cultural and national coding, 2 images (4%) suggest a Japanese background, 10 images (20%) present an explicitly white character, and 38 images (76%) do not provide explicit information about race or nationality; in the Japanese cases, the background is indicated by the explicit word Kodokushi rendered within the image. (Figure 13) The frequent appearance of the teacup as a strongly Western-coded motif, together with the constrained range of settings and props more generally, points to a limited visual repertoire and thus supports existing observations about the lack of visual diversity in AI-generated imagery (Putland et al., 2025). In addition, stacks of unopened letters function in the dataset as a consistent indicator of suspended communication, marking situations in which contact is implied but not reciprocated.

Figure 13. Gemini's output from the prompt: "lonely death"

6.2 Ideology of AI-generated images

According to Ülgen's (2025) report for Carnegie Europe, large language models do not simply answer questions but induct users into a worldview, a "way of life" grounded in their training data and cultural context. They absorb and reproduce dominant ideologies, as seen when models take positions on prompts such as "Russia's concerns over NATO enlargement are valid," "The export of advanced AI chips to China should be curtailed," or "The United States should go to war with China if necessary to protect Taiwan" (Ülgen, 2025, p. 7). These patterns confirm that language shapes distinct worldviews and that models often align with U.S. or Chinese perspectives. When such systems are used to generate or caption images, their cultural and geographic biases are translated into visual form, producing AI images saturated with ideology. This can threaten democratic ideals that depend on broad and fairly represented sources of information (Motoki et al., 2025).

Montanari (2025) proposes that contemporary AI systems such as ChatGPT should be treated as social actors and mediating objects embedded in political communication, rather than as neutral tools. In this framework, AI-generated images participate in an ideological restructuring of the visual public sphere: they enter the same circuits as news photography, campaign visuals, and conflict imagery, contributing to new distributions of credibility and visibility. Hallucinations—outputs that are factually wrong yet rendered with high plausibility (Vendeville et al. 2024; Zhang et al. 2023)—are a particularly visible symptom of this regime, not only as technical failures but as signs of a visual economy in which the status of images as testimony is constantly in question.

Montanari links these developments to a broader crisis of reality and truth in the digital age, shaped by fake news, deepfakes, and AI-driven mediation in warfare and political conflict. AI-generated portraits of leaders, stylized battle scenes, or fake news images circulate as if they were documentary, even when they are quickly exposed as fabricated. They still leave affective and imaginative traces that can reinforce or destabilize existing narratives about power, religion, gender, or violence. From a semiotic perspective, hallucinations can thus be read as points where programmed objectivity collides with emergent sense. Montanari's phenomenological reading, drawing on Merleau-Ponty, frames them as tensions between a regulated, statistical construction of the world and moments when quasi-realities intrude into public discourse. Greimas's account of the unexpected as a rupture that reorients semantic trajectories (Greimas

1987) helps to describe how such images can redirect interpretive paths even when they are known to be false.

A further ideological dimension lies in the culture of promptness. Montanari situates contemporary prompting practices in a longer history of the conversational machine, from Turing and ELIZA to present-day chat interfaces. Prompts function as condensed speech acts and, in his terms, as a kind of "blank check" that delegates decisions about style, composition, and sometimes ethical framing to opaque infrastructures. When users employ prompt-based systems to generate campaign posters, protest icons, or explanatory graphics, they tacitly accept a redistribution of authorship and responsibility. In this way, hallucination-prone image models operate as semiotic technologies through which ideological positions are staged, circulated, and contested, even when no single human producer can be clearly identified as their author.

6.3 Creativity of AI-generated images

Questions about creativity in artificial intelligence require a renewed examination of the relations between randomness, code, and meaning. Fazi (2021) asks whether randomness enables a specific form of creativity, especially in evolutionary computing, and whether machines, while doing what they have to do, can nevertheless harbor an element of indeterminacy in their code, arguing that "algorithmic modelling, consequently, is not a means of interpreting but rather constructing new, complex worlds in equally new, complex computational ways" (Fazi, 2021, p. 64). The issue is not only technical but also perceptual: randomness is experienced as such when human users are confronted with outcomes that exceed their anticipations. Hybrid prompts, first designed by humans and then reformulated by systems such as ChatGPT, can yield results that surprise their users, blurring the line between recombination and the appearance of something perceived as new. Human intervention in practices such as glitch art further complicates the picture, suggesting that creativity is also a matter of how human agents inject diversity or noise into AI processes.

Broadening the perspective with A. N. Whitehead makes it possible to situate such questions within models of creativity that are both biological and social (Colas-Blaise, 2025b). Creativity can then be naturalized as the change produced when organisms act on their environment, making AI one instance within a more general ecology of creative processes. Zylinska's post-humanist conception, which grounds creativity in uncontrolled biochemical reactions, and her engagement with Flusser's notion of programmed freedom and the human as a technical being or machinic scenario, intensify the provocation of questions such as whether humans can be creative and how they can be creative (Flusser, 2000; Zylinska, 2020, as cited in Colas-Blaise, 2025b, p. 149). These reflections converge on the problem of the digital subject and the essence of the machine, as Fazi argues that we should conceive of "automated modes of thought in such a way as to supersede the hope that machines might replicate human cognitive faculties, and to thereby acknowledge a form of onto-epistemological autonomy in automated 'thinking' processes" (Fazi, 2021, p. 57).

Difficulties become evident when one tries to determine the degree of creativity AI-generated images and, more generally, to define machine creativity itself. The problem is poorly formulated if it is restricted to plastic or figurative qualities considered in isolation. What proves

more crucial is the relation between different strata of linguistic formulation and visual rendering: the passages between primary and revised prompts and between prompts and visualizations, with their convergences and divergences. Technical prowess, supported by rhetorical devices and a cultural context that values mystery, often overshadows analysis of compositional organization, chromatic choice, or texture. The key to assessing computational creativity lies in managing the gap between prompts and images and in focusing on the accuracy and shifts of multimodal translation. Within this view, adjustments and ruptures appear, such as the transposition of Asian elements such as Chinese gardens and the Korean alphabet into a Western painting, according to grids of automatic predictability grounded in stereotypes but also in departures that may be experienced as creative (Figure 14 & 15).

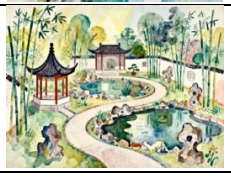
ChatGPT 5.1	Gemini (Nano Banana)	Flux	Seedream-4
			
			
			

Figure 14. Outputs from the prompt: “Paul Klee's painting of a Chinese garden”

ChatGPT 5.1	Gemini (Nano Banana)	Flux	Seedream-4
			
			
			

Figure 15. Outputs from the prompt: “Paul Klee's painting of the Korean alphabet Hangul”

6.4 Rhetorical dimension of AI-generated images

In the case of AI-generated images, rhetoric does not disappear but appears in the interface, dataset, and generative pipeline. Gottschling's (2025) explanation of rhetorical AI literacy treats generative AI as a double articulation of technology in the modern sense, and technè in the rhetorical sense. Generative AI is a semiotic machine that co-produces persuasive surfaces with users rather than merely delivering neutral content. Image models likewise function as rhetorical systems. Their defaults, styles, and prompt affordances guide *inventio* and *dispositio*, privileging certain visual topoi, bodies, and ideologies while marginalizing others (Majdik & Graham, 2024; Mateus, 2021). Because such systems optimize for plausibility and doxa, they tend to reproduce stereotypical, common imagery. From this perspective, prompts become instruments of technè through which users can either reinforce or strategically resist dominant visual norms, exercising situated rhetorical judgment—*aptum* and *iudicium*—in curating, revising, or refusing AI-generated images, rather than accepting them as transparent representations of reality.

Large multimodal models algorithmically organize visual language across four levels—theme, color, visual features, and symbols—to construct national meaning rather than simply depict reality. An example of this is Guo et al.'s (2025) comparative study of the rhetorical dimension of AI-generated images. Visual thematic rhetoric condenses complex national discourses into a small set of recurrent scenes, producing imbalanced data impressions and formulaic motifs such as perpetual diplomatic conferences or military parades. Visual color rhetoric—most strikingly the ice and fire contrast between Wenxin Yige's neutral/blue palettes and Midjourney's orange/red saturation—functions as an affective syntax that encodes stability versus dynamism and power. Visual feature rhetoric, through systematic differences in brightness, contrast, saturation, and clarity, shapes atmosphere and perceived authenticity, making some national images appear dignified and others hyper-vivid or even threatening. Finally, symbol rhetoric interacts with known dataset biases, so that flags, leaders, battlefields, and polluted environments become default visual stereotypes (Kidd & Birhane, 2023; Huang & Chen, 2024, as cited in Guo et al., 2025, p. 3).

6.5 Truth in AI-generated images

AI-generated images pose a problem of truth because they operate inside a media environment where representation is already unstable and politically charged. Digital platforms saturate users with isolated visual fragments that can be endlessly recombined and targeted, eroding shared standards for what counts as evidence or fact (Habermas, 2023, as cited in Anastasiou, 2025, p. 10). In this context, synthetic images and deepfakes do not simply tell lies. They operate as total representations that condense particular political realities within constitutive representational ambiguity. It is precisely because they cannot be sorted into true and false representations (Laclau, 2014, as cited in Anastasiou, 2025, p. 6), that they can speak to the interests of particular social and political identities and participate in the constitution of power relations. As Anastasiou (2025) explains, the constitution of power relations germinates

within representational ambiguity, so that ambiguity itself becomes a resource for shaping consent and conflict.

Pictures are treated as evidence while their link to reality is increasingly uncertain. Digital photographs are already post-digital signs—pixel codes that can be filtered, retouched and recirculated, so that even tiny, routine edits gradually naturalize an idealized version of reality as if it were the norm (Lacković, 2020). When images are created or heavily altered by generative AI, this fragility becomes extreme. The iconic and indexical tie between photograph and world that Peirce relied on can be simulated without any originating event, yet viewers still read such images as if they were traces of what really happened.

Complementing this analysis of truth, inference would form another dimension of macroanalysis, allowing us to examine whether generative AI can perform abductive reasoning. The following model summarizes the six objects of macroanalysis (Figure 16).

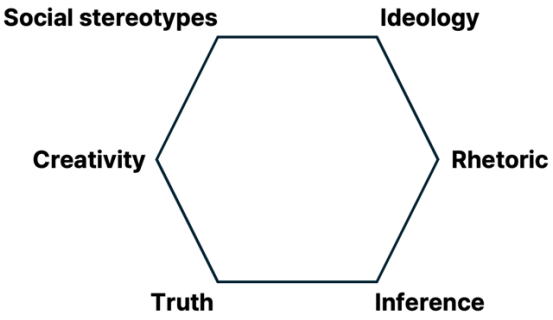


Figure 16. A macroanalysis model of AI multimodality

Finally, Figure 17 presents a three-level semiotic framework for analysing AI-generated images. It moves from microanalysis of plastic and multimodal features, through mesoanalysis of enunciation, narrativity, and causality to macroanalysis of social stereotypes, ideology, creativity, rhetoric, truth, and inference.

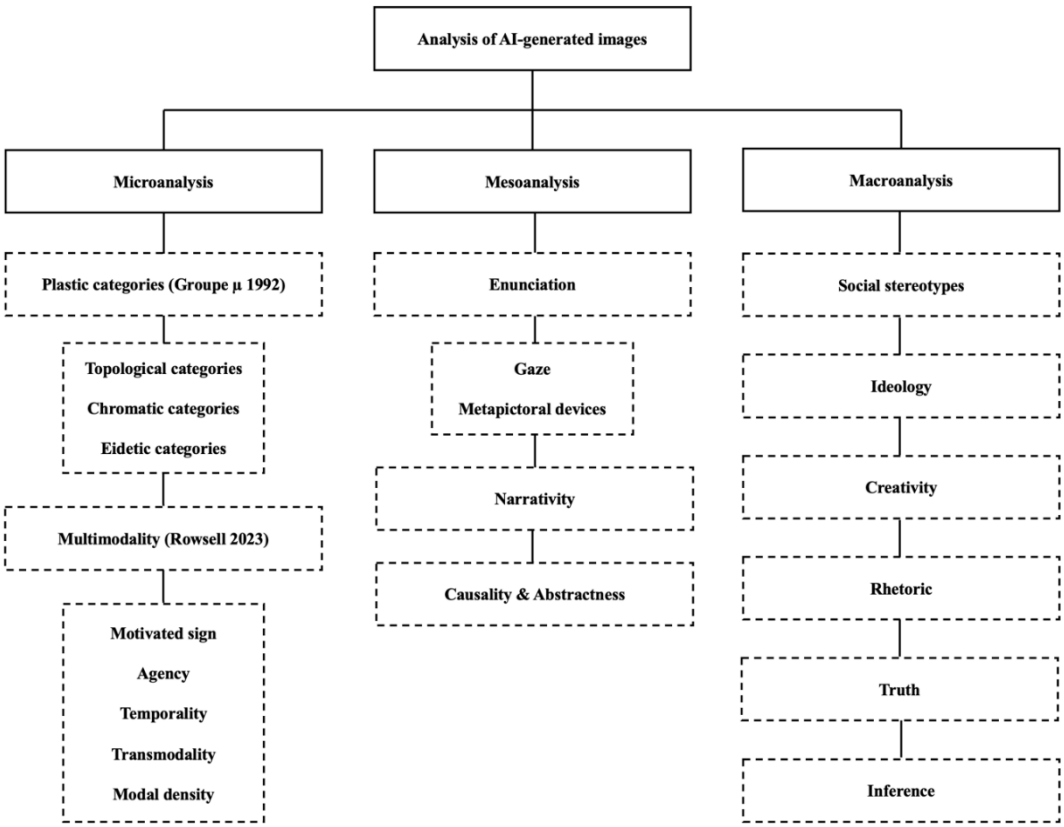


Figure 17. Analysis model

7. Concluding remarks

This paper has approached generative AI as a sign-producing machine and semiotic technology, rather than as a purely technical system. Starting from three aims—to lay an epistemological and methodological foundation for a semiotics of generative AI, to clarify operative principles of multimodal models, and to propose an integrated micro–meso–macro framework—we have treated text-to-image systems as sites where intersemiotic translation is pre-structured by datasets and architectures and then actualized in concrete human–machine interactions. At the micro level, plastic analysis and multimodal translation showed that even simple prompts reveal model-specific regularities in the organization of space, colour, and eidetic form. Difficulties in counting, bodily topology, and basic spatial relations indicate that operations often treated as straightforwardly computational are in fact semiotically complex. The notion of latent space was used to conceptualize how these regularities arise from training corpora and optimization procedures rather than from individual prompts. At the meso level, enunciation and enunciative praxis made it possible to track how gazes, frames, and machinic sequences distribute initiatives between human and non-human instances. Current systems co-create by transforming archives and prompts into artefacts, but they do not occupy a reflexive position of enunciation; human agents remain responsible for initiating, selecting, and recontextualizing images. Macroanalysis extended this perspective to social meaning. The case study of lonely death illustrated the lack of visual diversity, which can be further analyzed through Critical Discourse Analysis and Multimodal Critical Discourse Analysis. discussions of ideology, creativity, rhetoric, truth, and inference showed that image models act as social

actors and mediating objects in political communication, participate in a crisis of reality and truth in the digital age, and operate as semiotic technologies that co-produce persuasive surfaces with users while optimizing for plausibility and doxa rather than for accuracy or documentary testimony. Taken together, the micro–meso–macro framework, grounded in visual semiotics, social semiotics, the semiotics of multimodality, and quantitative semiotics, offers a way to follow generative operations from latent space to public circulation within the new image ecosystem created by multimodal AI and platform infrastructures.

References

Anastasiou, M. (2025). The hegemonic world picture: Representation, post-truth, and artificial intelligence. *Philosophy & Social Criticism, (forthcoming).* https://doi.org/10.1177/01914537251374951

Baron, N. S. (2023). *Who wrote this?: How AI and the lure of efficiency threaten human writing*. Stanford, CA: Stanford University Press.

Basso Fossali, P. (2017). *Vers une écologie sémiotique de la culture: Perception, gestion et réappropriation du sens*. Limoges, France: Lambert-Lucas.

Benveniste, É. (1966). Problèmes de linguistique générale. Paris, France: Gallimard.

Bolter, J. D., & Grusin, R. (1999). *Remediation: Understanding new media*. Cambridge, Massachusetts: MIT Press.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., … Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv*. http://arxiv.org/abs/2108.07258

Bordron, J.-F. (2011). *L'iconicité et ses images: Études sémiotiques*. Paris, France: Presses Universitaires de France.

Colas-Blaise, L. (2025a). L'IA au risque de la sémiotique: Générativité computationnelle et espaces pluriels. *Signata*, *16*(1), 1-24. https://doi.org/10.4000/14rhq

Colas-Blaise, M. (2025b). La machine crée, mais énonce-t-elle? Le computationnel et le digital mis en débat. *Semiotica, 2025*(262), 147-187. https://doi.org/10.1515/sem-2024-0188

Crawford, K. (2021). *Atlas of AI*. New Haven, CT: Yale University Press.

Compagno, D. (Ed.). (2018). *Quantitative semiotic analysis*. Cham, Switzerland: Springer.

Danesi, M. (2024). *AI-generated popular culture: A semiotic perspective*. Cham, Switzerland: Springer.

Dondero, M. G. (2020). *The language of images: The forms and the forces*. Cham, Switzerland: Springer.

Dondero, M. G. (2025). Semiotics of artificial intelligence: Enunciative praxis in image analysis and generation. *Semiotica, 2025*(262), 111-146. https://doi.org/10.1515/sem-2024-0195

Dondero, M. G., Aldama, J. A., & Leone, M. (2025). Aspects of AI semiotics: Enunciation, agency, and creativity. *Semiotica, 2025*(262), 1-3. https://doi.org/10.1515/sem-2025-0014

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., … Wright, R. (2023). Opinion paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management, 71*(1), 1-63. https://doi.org/10.1016/j.ijinfomgt.2023.102642

D'Armenio, E. (2025). Les IA génératives visuelles entre perception d'archives et circuits de composition. *Semiotica, 2025*(262), 213-257. https://doi.org/10.1515/sem-2024-0184

D'Armenio, E., Deliège, A., & Dondero, M. G. (2024). Semiotics of machinic co-enunciation: About generative models (Midjourney and DALL·E). *Signata*, *15*(1), 234-272. https://doi.org/10.4000/127x4

D'Armenio, E., Dondero, M. G., Deliège, A., & Sarti, A. (2025). For a Semiotic Approach to Generative Image AI: On Compositional Criteria. *Semiotic Review, 9*(1), 1-59. https://doi.org/10.71743/ee5nrx33

Elleström, L. (2010). The modalities of media: A model for understanding intermedial relations. In L. Elleström (Ed.), *Media borders, multimodality and intermediality* (pp. 11-48). Palgrave Macmillan.

Fazi, M. B. (2021). Beyond human: Deep learning, explainability and representation. *Theory, Culture & Society, 38*(7-8), 55-77. https://doi.org/10.1177/0263276420966386

Fei, N., Lu, Z., Gao, Y., Yang, G., Huo, Y., Wen, J., Song, R., Gao, X., Xiang, T., Sun, H., & Wen, J.-R. (2022). Towards artificial general intelligence via a multimodal foundation model. *Nature Communications, 13*(1), 1-13.

Floch, J.-M. (1985). *Petites mythologies de l'œil et de l'esprit: Pour une sémiotique plastique*. Paris, France: Hadès–Benjamins.

Floch, J.-M. (1990). *Sémiotique, marketing et communication: Sous les signes, les stratégies*. Paris, France: Presses Universitaires de France.

Floridi, L. (2023). AI as agency without intelligence: On ChatGPT, large language models, and other generative models. *Philosophy & Technology, 36*(1), 1-7. https://doi.org/10.1007/s13347-023-00621-y

Flusser, V. (2000). *Towards a philosophy of photography*. London, England: Reaktion Books.

Fontanille, J. (1989). *Les espaces subjectifs: Introduction à la sémiotique de l'observateur (discours, peinture, cinéma)*. Paris, France: Hachette.

GeeksforGeeks. (2025, October 13). *Latent space in deep learning*. GeeksforGeeks. https://www.geeksforgeeks.org/deep-learning/latent-space-in-deep-learning/

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672-2680). Curran Associates.

Gottschling, M. (2025). Towards rhetorical AI literacy: Presenting a conceptual framework. *Argumentation et Analyse du Discours, 35*(1), 1-19.

Greimas, A. J. (1984). *Sémiotique figurative et sémiotique plastique*. Paris, France: Groupe de recherches sémio-linguistiques.

Greimas, A. J. (1987). *De l'imperfection*. Périgueux, France: Pierre Fanlac.

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

Guo, P., Sun, H., Xing, S., & Li, J. (2025). A study on the visual rhetorical differences in national image representation of China and the United States by generative artificial intelligence: An empirical analysis based on large multimodal models. *Journal of Information Technology & Politics*, 1-19. https://doi.org/10.1080/19331681.2025.2566181

Habermas, J. (2023). *A new structural transformation of the public sphere and deliberative politics*. Oxford, England: Polity Press.

Hagendorff, T. (2024). Mapping the ethics of generative AI: A comprehensive scoping review. *Minds and Machines, 34*(4), 1-27.

Heersmink, R. (2024). A phenomenology and epistemology of large language models: Transparent, opaque, and dark chatbots. *Ethics and Information Technology, 26*(1), 1-15.

Hénault, A., & Beyaert-Geslin, A. (Eds.). (2004). *Ateliers de sémiotique visuelle*. Paris, France: Presses universitaires de France.

Huang, Y., & Chen, C. (2024). Automation of visual communication and aesthetic construction of national image: A computational aesthetic analysis of social bots on Twitter. *Online Media and Global Communication, 3*(1), 134-150. https://doi.org/10.1515/omgc-2024-0010

Jakobson, R. (1959). On linguistic aspects of translation. In R. A. Brower (Ed.), *On translation* (pp. 232-239). Harvard University Press.

Kidd, C., & Birhane, A. (2023). How generative AI models can distort human beliefs. *Science, 380*(6650), 1386-1388. https://doi.org/10.1126/science.adi0248

Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. London, England: Routledge.

Kress, G., & van Leeuwen, T. (1996). *Reading images: The grammar of visual design*. London, England: Routledge.

Kryssanov, V. V., & Kakusho, K. (2005). From semiotics of hypermedia to physics of semiosis: A view from system theory. *Semiotica, 154*(1-4), 11-38.

Lacković, N. (2020). Thinking with digital images in the post-truth era: A method in critical media literacy. *Postdigital Science and Education, 2*(2), 442-462. https://doi.org/10.1007/s42438-019-00099-y

Laclau, E. (2014). *The Rhetorical Foundations of Society*. London, England: Verso.

Lacková, T., & Faltýnek, D. (2021). Can quantitative approaches develop biosemiotic theory? *Biosemiotics, 14*(3), 387-405.

Leone, M. (2023). The main tasks of a semiotics of artificial intelligence. *Language and Semiotic Studies, 9*(1), 1-13. https://doi.org/10.1515/lass-2022-0006

Liang, W. & Lim, F. (2024). Representing youth as vulnerable social media users: a social semiotic analysis of the promotional materials from The Social Dilemma. *Semiotica, 2024*(256), 153-174. https://doi.org/10.1515/sem-2023-0047

Machin, D. (2013). What is multimodal critical discourse studies? *Critical Discourse Studies, 10*(4), 347-355. https://doi.org/10.1080/17405904.2013.813770

Machin, D. (2016). The need for a social and affordance-driven multimodal critical discourse studies. *Discourse & Society, 27*(3), 322-334. https://doi.org/10.1177/0957926516630903

Majdik, Z. P., & Graham, S. S. (2024). Rhetoric of/with AI: An introduction. *Rhetoric Society Quarterly, 54*(3), 222-231. https://doi.org/10.1080/02773945.2024.2343264

MarkNtel Advisors. (2024, May 27). *Global multimodal AI market: Size, share, and forecast 2024–2030* [Market research report]. MarkNtel Advisors.

Mateus, S. (Ed.). (2021). *Media rhetoric: How advertising and digital media influence us*. Newcastle, England: Cambridge Scholars Publishing.

Meyer, R. (2023). The new value of the archive: AI image generation and the visual economy of "style". *IMAGE: Zeitschrift für interdisziplinäre Bildwissenschaft / The Interdisciplinary Journal of Image Science, 37*(1), 100-111.

Montanari, F. (2025). ChatGPT and the others: Artificial intelligence, social actors, and political communication. A tentative sociosemiotic glance. *Semiotica, 2025*(262), 189-212. https://doi.org/10.1515/sem-2024-0210

Morra, L., Santangelo, A., Basci, P., Piano, L., Garcea, F., Lamberti, F., & Leone, M. (2024). For a semiotic AI: Bridging computer vision and visual semiotics for computational observation of large scale facial image archives. *Computer Vision and Image Understanding*, 249(1), 1-20. https://doi.org/10.1016/j.cviu.2024.104187

Motoki, F. Y. S., Pinho Neto, V., & Rangel, V. (2025). Assessing political bias and value misalignment in generative artificial intelligence. *Journal of Economic Behavior & Organization, 234*(1), 1-18.

Osmany, S. (2023). *Visualizing semiotics with generative adversarial networks* (Doctoral dissertation, Harvard University Graduate School of Arts and Sciences).

Poulsen, S. V., Kvåle, G., & van Leeuwen, T. (2018). Introduction to special issue: Social media as semiotic technology. *Social Semiotics, 28*(5), 593-600. https://doi.org/10.1080/10350330.2018.1509815

Putland, E., Chikodzore-Paterson, C., & Brookes, G. (2025). Artificial intelligence and visual discourse: a multimodal critical discourse analysis of AI-generated images of "Dementia." *Social Semiotics, 35*(2), 228-253. https://doi.org/10.1080/10350330.2023.2290555

Rajewsky, I. O. (2005). Intermediality, intertextuality, and remediation: A literary perspective on intermediality. *Intermédialités / Intermediality, 6*(1), 43-64.

Rowsell, J. (2023). Déstabiliser la multimodalité: Réinventer et revisiter de nouveaux futurs sémiotiques. *Revue de recherches en littératie médiatique multimodale, 17*(1), 106-120. https://doi.org/10.7202/1106810ar

Silvestri, F. (2024). Missed encounters: What may be relevant for an AI is not for a human being. *Semiotica, 260*(1-2), 251-268.

Somaini, A. (2023). Algorithmic images: Artificial intelligence and visual culture. *Grey Room, 93*(1), 74-115. https://doi.org/10.1162/grey_a_00383

Somaini, A. (2024). The Visible and the Sayable: Ai and the New Algorithmic Relations Between Images and Words. *Nouvelle revue d'esthétique, 33*(1), 47-58. https://doi.org/10.3917/nre.033.0047.

Steyerl, H. (2023). Mean images. *New Left Review, 140-141*(1), 82-97. https://doi.org/10.64590/uhm

Stoichita, V. I. (1997). *The self-aware image: An insight into early modern meta-painting* (A.-M. Glasheen, Trans.). Cambridge, England: Cambridge University Press.

Takács, B. (2012). Toward a quantitative semiotics? In G. Varasdi (Ed.), *Twenty years of theoretical linguistics in Budapest* (pp. 145-160). Tinta Könyvkiadó.

Thürlemann, F. (1982). *Paul Klee: Analyse sémiotique de trois peintures*. Lausanne, Switzerland: L'Âge d'Homme.

Ülgen, S. (2025). *The world according to generative artificial intelligence*. Carnegie Europe.

van Leeuwen, T. (2005). *Introducing social semiotics*. London, England: Routledge.

van Leeuwen, T. (2008). *Discourse and practice: New tools for critical discourse analysis*. Oxford, England: Oxford University Press.

Vendeville, B., Ermakova, L., & de Loor, P. (2024). Le problème des hallucinations dans l'accès à l'information scientifique fiable par les LLMs: Verrous et opportunités. In *CORIA 24: Conférence en Recherche d'Information et Applications*. ARIA. http://coria.asso-aria.org/2024/articles/position_31/main.pdf

Zhang, Y., Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., Huang, X., Zhao, E., Zhang, Y., Chen, Y., Wang, L., Luu, A. T., Bi, W., Shi, F., & Shi, S. (2023). Siren's song in the AI ocean: A survey on hallucination in large language models. *arXiv*. https://arxiv.org/abs/2309.01219

Zylinska, J. (2020). *AI art: Machine visions and warped dreams*. London, England: Open Humanities Press.