From the represenationalist stance to conceptual blending in AI-generated images

Francesco D'Isa

Francesco.disa@unimore.it

Lorenzo Manera

Lorenzo.manera@unimore.it

This paper examines how AI-generated images reconfigure key aesthetic and epistemological categories within contemporary visual culture. Drawing on Nelson Goodman's distinction between autographic and allographic art forms, we argue that AI image generation occupies a hybrid position, functioning as a "mixed case" where linguistic prompts serve as notational structures and generative outputs as autographic instantiations. Through the notion of semantic attractors, we describe how lexical inputs act as gravitational forces within the model's latent space, guiding the emergence of visual and affective configurations. This dynamic is interpreted through conceptual blending theory, revealing how AI systems integrate heterogeneous domains into novel perceptual and semantic syntheses. Finally, we introduce the concept of the image-ideogram to analyze the cultural impact of deepfakes and synthetic imagery, suggesting that photorealistic synthesis now operates less as an epistemic index and more as a visual and linguistic resource within the accelerated economy of contemporary visual discourse.

Keywords: AI-generated images; media Aesthetics; image-ideograms; Artificial Intelligence; semantic attractors.

From the represenationalist stance to conceptual blending in AI-generated images

Francesco D'Isa
Francesco.disa@unimore.it
Lorenzo Manera
Lorenzo.manera@unimore.it

1. Introduction

For several decades, visual studies have rejected the represenationalist stance, emphasizing how technical images are not to be intended as passive representations, but rather as agents endowed with agency, active interlocutors in human experience, capable of shaping perception, mediating knowledge, participating in the production of meaning and eliciting responses in their viewers[1]. As Alloa underlined,[2] such a perspective stems from Aby Warburg's idea of visual forms as crystallized affects and by his metaphor of images as "energy-containers".[3] The agency of images became, many years later, a recurrent topic in the field of visual studies and image theory.

In his famous 1989 essay, David Freedberg explored the idea of images intended as active instances and generative agents,[4] while Alfred Gell in *Art and Agency* analyzed the agency of visual artifact from an anthropological perspective.[5] As noted by Alloa, Gell developed a theory of performativity which sees visual artifacts as the manifestation of

[1] Cfr. Micaela Latini, Elena Tavani, Connected Images: New Paradigms for Aesthetic Experience Paradigmi, XLI, 3/2023. Itinera. 27 (2024).

[2] Emmanuel Alloa, Chiara Cappelletto, eds., *Dynamis of the Image: Moving Images in a Global World* (Walter de Gruyter, 2020).

[3] Aby Warburg, "Allgemeine Ideen," Warburg Institute WIA, III, 102.1.4.1, 1927.

[4] David Freedberg, *The Power of Images. Studies in the History and Theory of Response* (Chicago University Press, 1989).

[5] Gell, Alfred. Art and Agency. An Anthropological Theory. Clarendon Press, 1998.

the mind and uses them as means to an end.[6] Within the field known as Bildwissenschaft, Horst Bredekamp's theory of the *Bildakt* offered a distinct yet complementary perspective[7], based on the idea of images intended as prime movers, intrinsic begetters of action. As recently noted by Quartesan[8], for the Bildwissenchaft theorist one of the central tasks confronting visual-studies in the contemporary field is to engage critically with artificial intelligence generated images, heightening awareness of the ever-present danger of idolatry that accompanies our attempts to interpret these emerging visual forms. Because this risk is bound up with the very force that images exert, Bredekamp insists on bringing the resulting tension to the surface, pinpointing the pictorial mechanism that produces it, and thereby sidestep misreadings of the visual force.

In our view, the increasing spread of AI-generated images impacts, this tension bare, generating new aesthetic phenomena.

The contemporary aesthetic debate discusses the hypothesis that synthographies (AI-generated images) images imply an epistemological shift, which reconfigures the word-image relationship[9]. Secondly, the idea that AI-generated images represent a new frontier in the visual arts, as they transpose the image-creation process from the analog arts to the notational ones, providing further aspects of the distributed agency theory. Thirdly, the hypothesis that AI-generated images have an impact on visual culture at large[10].

2. Semantic attractors, conceptual blending and shared agency

With regard to the first hypothesis, our contribution focuses on the concept of the semantic attractor. We argue that certain words behave as if they were hidden magnets. When prompting, if one inserts words such as *baroque*, *noir*, or *sad*, the model's latent

---

[6] Emmanuel Alloa, "Performing an Appearance. On the Performativity of Images," *Paradigmi* 41, no. 3 (2023): 415–28.

[7] Horst Bredekamp, *Image Acts: A Systematic Approach to Visual Agency*, ed. Elizabeth Clegg (Walter de Gruyter, 2018).

[8] Ivan Quartesan, *La nuova iconologia di Horst Bredekamp. Storia dell'arte, morfologia e il potere delle immagini* (Mimesis, 2024).

[9] See Bajohr, Operative ekphrasis: The collapse of the text/image distinction in multimodal AI. Word & Image, 40(2), 2024, pp. 77-90, and J. D. Bolter, AI Generative Art as Algorithmic Remediation, in "IMAGE. Zeitschrift für interdisziplinäre Bildwissenschaft", 2023, and Arielli, E., Manovich, L. AI-aesthetics and the Anthropocentric Myth of Creativity. Nodes Journal of Art and Neuroscience, 19(1), 2022.

[10] See Antonio Somaini, Algorithmic Images: Artificial Intelligence and Visual Culturein "Grey Room", (93),2023pp. 74-115 and A. Somaini, Sortir du paradigme de la copie. *Numérique*, 2024.

space tilts, funneling the image toward specific colors, textures, and moods. A semantic attractor can be intended as any lexical item embedded within the model's latent space, but endowed with a variable attractor strength, items whose embeddings occupy denser and more coherent semantic regions exert a stronger gravitational pull on the generative trajectory, bending the output toward their characteristic visual and affective features, whereas weaker attractors influence the image only marginally[11].

As Mieke Bal argued[12], the meaning of an image is never fixed but unfolds over time through acts of recognition. Seeing, for Bal, is to participate in a process of reading that is continually renewed. Each observer, each cultural framework, and each medium modifies the image, activating its narrative and affective potentialities and diverting a sense that is always in motion. There is no essence of the image: visual forms live in the passage between interpretations and rewritings. Meaning is a provisional outcome, the transient equilibrium of a shifting field.

Semantic attractors can be interpreted as forces that orient and deform the field of meaning whenever an image enters circulation. Images activate force fields in which concepts and affects move, combine, and come into conflict.

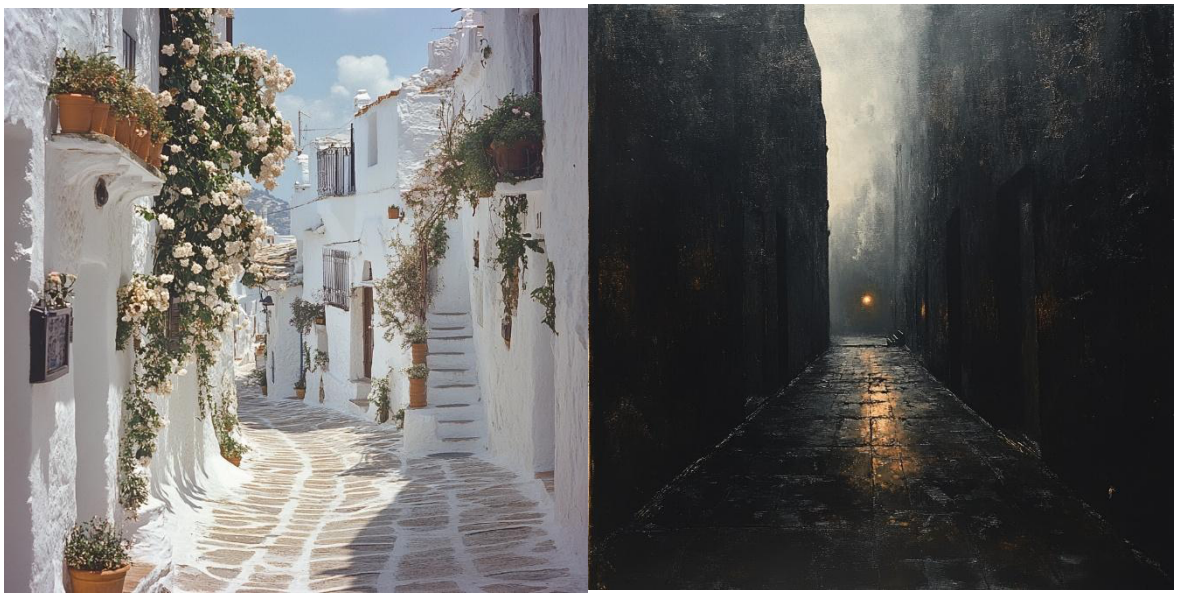
When using these generative systems, it quickly becomes evident that every word composing a prompt that acts as a gravitational pole, concentrating around itself the associations acquired during training and pushing the generative process in one direction rather than another.

[11] For a more detailed analysis of the concept of semantic attractor, see D'Isa, *Semantic Attractors and Distributed Agency: Rethinking Authorship in Generative AI*, in print.

[12] Mieke Bal, Reading "Rembrandt": Beyond the Word–Image Opposition. Cambridge: Cambridge University Press, 1991.

Anyone who has used systems such as Midjourney, Stable Diffusion, ChatGPT, or Gemini can easily observe this. For instance, when we write a prompt such as *a white street*, the model not only produces a white street but tends to introduce snow, brightness, or sunlit buildings. With *a black street*, the same scene is populated by night, winter, and melancholy. These details do not belong to the lexical content of the prompt itself; they emerge from the attractive force of associations sedimented in the model's latent space. In other words, *white* and *black* operate not only at the level of denotation but as gateways to entire semantic constellations: purity, coldness, luminosity on one side; darkness, sadness, and danger on the other.

*An example of two very simple **semantic attractors** generated with Midjourney 6.1, using the prompts "a white street" and "a black street" respectively. As can be observed, black attracts not only the chromatic value but also the concept of **night**, which expands throughout the entire image, including the cars. White, by contrast, attracts both **snow** and **brightness**, diffusing luminosity across the scene.*



Same prompt. Here, black attracts an atmosphere that is markedly sadder and, indeed, darker than white, even though such affective qualities were never explicitly requested.

From a technical point of view, this dynamic depends on how artificial intelligence models organize texts and images within the latent space. This space can be described as a mathematical map in which concepts are positioned closer or farther apart according to how frequently they co-occur in the training data. If two words or images appear in similar contexts—for instance, *sea* and *beach*—they end up in the same neighborhood of the map; whereas terms belonging to very different domains—such as *hospital* and *rainbow*—remain distant from one another[13].

We may imagine this space as a landscape of semantic hills and valleys: some areas are particularly dense, filled with recurring associations. These dense regions act as attractors, since when the model generates an image, it tends to slide toward those zones where it encounters greater statistical coherence.

It is crucial to understand that an attractor is not a natural or intrinsic property of a word. It does not exist independently. Rather, it emerges from the cultural and stylistic correlations sedimented in the data: films, critical essays, illustrations, and shared discourses. In practice, attractors reflect the cultural habits and aesthetic conventions that have accumulated within the archives on which the AI model has been trained.

[13] Antonio Somaini, 2025. "Theory of Latent Spaces." In Le Monde selon l'IA: Explorer les espaces latents, a cura di Ada Ackerman, Alexandre Gefen, Antonio Somaini, e Joanna Zylinska, 21–51. Paris: Jeu de Paume / JBE Books.

To better understand this phenomenon, it is possible to consider Fauconnier and Turner's theory of conceptual blending[14], which stems from the authors' earlier work in the fields of linguistics and cognitive science. One of the core concepts of the theory is the idea of mental space, which Fauconnier develops in his writings from the 1990s on the interpretation of counterfactuals and propositional attitude statements[15]**.** A mental space, according to the author, is an arrangement of concepts or images organized in such a way as to represent a specific type of recurring and familiar situation.

The "imaginary manipulation" of such a space allows us to anticipate the structure at a given moment and the behavior over time of the situations represented by the space itself[16]**.** This manipulation largely takes the form of blending. Blending, or conceptual integration, consists of the co-activation of two input mental spaces that are distinct instances of a generic space, which contains the general categories to which certain items from the two input spaces give form. The concepts or images of the first space correspond to those of the second if they embody the same concept or image in the generic space.

Based on such correspondences, certain constitutive elements of the input spaces are projected into the generic space, ultimately mapping onto a fourth "hybrid" space (blended space). This fourth space contains elements from the input spaces but may also include additional elements. While the Cognitive Metaphor Theory (CMT) of metaphors presupposes a binary relationship between two domains or concepts, the Blending Theory introduces and places particular emphasis on a third notion. Therefore, while CMT analyzes metaphor as directional, typically focusing on rooted and conventional relationships, the Blending Theory does not assume the directionality from "source" to "target" and often concentrates on novel conceptualizations.[17]

The pervasiveness of the conceptual metaphor supported by the CMT is rethought as the pervasiveness of blending. In this way, blending establishes a completely new structure of meaning within the fourth space, one that does not coincide with any of the original

---

[14] Fauconnier, G., Turner, M. 2002 The Way We Think: Conceptual Blending and the Mind's Hidden Complexities, Basic Books, New York.

[15] Fauconnier. 1994 Mental Spaces: Aspects of Meaning Construction in Natural Languages, Cambridge University Press, New York.

[16] Hills, D. 2017 Metaphor, in E.N. Zalta (a cura di), "The Stanford Encyclopedia of Philosophy (Fall 2017 Edition)", https://plato.stanford.edu/archives/ fall2017/entries/metaphor.

[17] Cserép, A. 2014 Conceptual Metaphor Theory: In Defence or on the Fence, in "Argu mentum", 10, pp. 261-288.

input spaces, in which the differences between the elements are not eliminated; rather, they remain active and serve in the construction of an integrated conceptual network.
Similarly, semantic attractors activate an integration space in which elements from different domains are projected and compressed into a unified configuration. This is the same process that allows metaphors to create new meaning: distant fields of significance collapse into a single space, giving rise to unexpected concepts.
Such a process brings us to the second issue emerging in the aesthetic debate, meaning the hypothesis that AI-generated contribute to transpose the image-creation process from the "autographic" to the "allographic" art forms.

3. From the "autographic" to the "allographic" art forms. Prompting as a design medium

In *Languages of Art*[18], the American philosopher Nelson Goodman proposed a distinction between the world of symbolic forms and that of representations. In Goodman's framework, works of art are to be understood as symbols operating within broader symbolic systems, which can be divided into autographic and allographic forms of art.
Autographic arts—such as painting or sculpture—lack both syntactic and semantic differentiation; that is, they do not possess a notation that can unambiguously distinguish the constitutive elements of a work from those that are merely contingent. An autographic artifact, therefore, can be regarded as a work of art precisely insofar as it resists notational codification: its identity depends on the material and historical conditions of its production rather than on a set of formal instructions or conventions.
By contrast, allographic arts—such as music, dance, or literature—are characterized by the presence of a notation system that allows works to be reproduced and correctly identified through compliance with syntactic and semantic rules. In Goodman's terms, while allographic works can be individuated through their notation and spelling, autographic ones require recourse to the history of their making as the criterion for determining their authenticity and identity. If the advent of digital photography only partially redefined the autographic nature of photographic image-making—since it still

[18] N.Goodman, Languages of arts. An approach to a theory of symbols, Indianapolis-New York, The Bobbs-Merrill Company, 1968.

presupposes the existence of a negative, whether analog or digital, and the prints derived from it—the recent advances in generative artificial intelligence (GenAI) have profoundly transformed the very process of image generation. In our view, AI-generated images can be regarded as analogous to the "mixed and transitional cases" described by Goodman. Such images may be understood as partly autographic and partly allographic, since the linguistic prompts function only as the notational component of the generated file—much as a musical score provides only the notational structure of a musical work. The creation of AI-generated images can be regarded as allographic during the design phase, insofar as its defining aspect lies in the precise formulation of a series of prompts that specify the elements constituting the image. In the subsequent generative phase, however, the process acquires an autographic dimension. What becomes most relevant here is the execution of those linguistic prompts, which determine the final characteristics of the image, turning prompting into a design medium which is impacting on visual culture at large.

4 Are deepfakes image-ideograms?

In our view, a relevant impact that the diffusion of AI-generated images and videos has had on visual culture at large is related to the phenomenon of deepfakes. To better understand this phenomenon, we propose the notion of the *image-ideogram*. It consists in a modular sign whose value lies in use, speed, and spread, conveying stance and emotion rather than truth. In our view, in fact, photorealistic synthesis has shifted from an epistemic promise (*seeing is believing*) to a linguistic resource (*seeing is signalling*). Their power, in fact, lies in shaping stance and mood, so the urgent shift is from judging *how real they look* to tracing *how and by whom they circulate*. Photorealistic synthesis enchants viewers long enough to be shared, but rarely long enough to be trusted. Its cultural power lies elsewhere: in supplying striking, endlessly re-captionable visuals that slot into the hypermemetic marketplace as stance markers and affective cues. Therefore, the deepfake's most consequential role is not to fool the eye but to unsettle the evidentiary status of video as a whole, an outcome that dovetails with the notion of the image-ideogram, one possible outcome the process of dissolution of the western image conception. If in the western tradition the image was typically treated as a secondary copy — an "image-of" something already real, assumed to possess less being and less

truth than the thing it depicted and upon which it relied, the coming of photography turned that logic. Andrè Bazin, with his interpretation of the photographic process as a semi-automatic phenomenon which entails "the eye of the machine", was among the first scholars who noted how, with the invention of photography, the old equation of image as a function of reality was changing[19].

But how is such a function reconfigured in the contemporary visual paradigm? In her 2009 essay, Steyerl describes a class of pictures that travel at accelerated speed, compressed, re-uploaded, re-edited, ripped, remixed until their resolution disintegrates but their visibility multiplies.[20] Under conditions of broadband abundance and platform competition, circulation eclipses clarity: the image becomes a unit of attention, exchanged like a low-denomination coin across social feeds. In our view, deepfakes inherit this logic while subverting its premise: they arrive in high definition yet are consumed as if they were poor. The higher the realism budget, the faster the artefact is flattened into an ideogrammatic token whose currency is speed and recognisability. In this sense, the poor image furnishes the infrastructural logic within which the deepfake can function as linguistic resource rather than evidentiary trace.

If the *poor image* explains how pictures move, the image-ideogram explains what they do once in motion. We adopt metaphorically the linguistic term *ideogram* for a sign grasped immediately and deployed combinatorially, independent of alphabetic spelling. Three intertwined attributes mark a deepfake that succeeds in this role. Firsty, the *Instant recognisability*. Viral deepfakes condense a single, highly legible motif. Recognition precedes interpretation, echoing Shifman's claim that memes are stabilised by *content* and *form* while their *stance* remains negotiable.[21]

Secondly, the *memetic portability*. Once identified, the clip is quickly shareable. Such portability performs the very function that McCulloch ascribes to emoji and GIFs: a

[19] André Bazin, and Hugh Grey, "The Ontology of the Photographic Image [1945]," *Film Quarterly* 13, no. 4 (1960): 4–9; Jean Baudrillard, *Simulacres et simulation* (Editions Galilée, 1981); Gretchen McCulloch, *Because Internet. Understanding the New Rules of Language* (Riverhead Books, 2019); Limor Shifman, *Memes in Digital Culture* (MIT Press, 2013).

[20] Hito Steyerl, "In Defense of the Poor Image," *e-flux Journal* 10, November 2009, https://www.e-flux.com/journal/10/61362/in-defense-of-the-poor-image.

[21] Shifman, *Memes in Digital Culture*.

pragmatic cue that communicates attitude faster than words can.[22] The deepfake, though technically high-definition footage, behaves linguistically.

Thirdly, the *Referential indifference***.** Because an ideogram's value is pragmatic, strict veracity becomes optional. Viewers savour the hyper-real sheen even when they know it is fabricated; what anchors meaning is not factual reference but the clip's built-in emblem already loaded with a clear, bite-size message. Context (who posts it, under which hashtag, to which crowd) then tilts that emblem toward admiration, irony, or outrage. In short, the deepfake's truth-status is secondary to the double work it performs: first, condensing a legible symbol; second, circulating that symbol through ever-shifting paratexts. These traits complete the trajectory outlined in this section: from the *index* that testified, through the *simulacrum* that masked, via the *poor image* that circulated, to the *image-ideogram* that signals. Deepfakes materialize this evolution in pure form, transforming photorealistic synthesis from an epistemic promise into the pictographic currency of twenty-first-century discourse. Deepfakes do not act as isolated files but as what Alfred Gell calls *extended objects*, whose person and artefacts form a single entity distributed in time and space.[23] Their agency unfolds through reposts, remixes and comment threads: each derivative frame becomes an index that motivates inferences and responses in the minds of recipients. the cases illuminate a broader shift in visual culture. From Bazin's indexical photograph, images have long drifted away from reality. The deepfake completes the arc: it is an image-ideogram, valued for speed, stance, and spread. In conclusion, in our view the deepfake era is less a crisis of seeing than a reconfiguration of signalling, exemplifying a broader shift from realism to symbolic currency, where images act as agents in cultural exchange rather than as proofs of reality.

[22] McCulloch, Gretchen. Because Internet. Understanding the New Rules of Language. Riverhead Books, 2019.

[23] Alfred Gell, *Art and Agency. An Anthropological Theory* (Clarendon Press, 1998).