


MULTIMODAL GENERATIVE AI AND THE METAMEDIUM CONDITION



Martin Suryajaya

Lecturer at Postgraduate School, Jakarta Institute of Arts

Abstract:

Since the decline of medium-specific approaches characteristic of early 20th-century modernism, various art theorists have introduced critical frameworks to address the fluidity of contemporary artistic practices. This article explores the implications of multimodal generative AI as a metamedium in contemporary art, building on the concepts of the post-medium condition and the postconceptual condition. As traditional notions of medium specificity dissolve, AI introduces a new paradigm where all artistic mediums converge into a single metamedium—data. This metamedium is inherently transmedial, assemblage-like, and deeply embedded in the social knowledge reflected in AI training datasets. By enabling transformations across text, image, sound, and other modalities, AI fosters a condition of metamediality, which challenges established boundaries of art forms. The article examines how this metamedium condition redefines creativity, positioning it as a collaborative process between human and non-human agents. Ultimately, it argues that the contemporary artistic landscape is now marked by a fluid interplay of mediums and temporalities, reshaping the very act of creation as a communal and boundless phenomenon.

Keywords: multimodal generative AI; metamedium condition; transmediality; creativity; contemporary art

The rise of multimodal generative AI has impacted many sectors, including the arts. There has been a significant amount of literature exploring the impacts in terms of AI replacing artists and dramatically changing job sectors (Jones 2021; Benanav 2022), whether or not AI genuinely creates artworks (Bajohr 2024; Arielli 2021), and the ethics of using generative AI in artistic practices (Flick & Worrall 2022).

I would like to focus on another issue rarely explored in the literature: the impacts of AI on the nature of contemporary artistic mediums. By focusing on shifts in the very nature of mediums, we can hopefully shed new light on the ever-elusive notions of creativity and contemporaneity, i.e., the meaning of 'contemporary' in 'contemporary art.' I argue that the rise of multimodal generative AI can be seen as a continuation of, rather than a break with, the relational and participatory aesthetics envisioned by Nicolas Bourriaud and Claire Bishop. By translating all artistic mediums into a single metamedium—data—multimodal generative AI further radicalizes the basic insight of participatory aesthetics that people themselves constitute the medium of contemporary art.

AI systems such as DALL-E and Midjourney have been trained on datasets containing every artifact of social relations, including texts, images, and sounds scraped from the internet. AI's knowledge thus reflects the general social knowledge that Marx dubbed as "the general intellect." This coming-together of the social in the form of AI creates a new understanding of contemporary art as the commoning of the medium and creativity as the commoning between different agents in different mediums.

In this presentation, we will unpack this cluster of ideas into three parts. *First*, we need to briefly review the conceptual evolution of artistic mediums, particularly the transition from the traditional version of medium essentialism espoused by modernist art theorists to the new, praxeological interpretation of medium essentialism developed over the last two decades. *Second*, we will consider the impacts of multimodal generative AI on the changing nature



of artistic mediums, leading us to examine the metamedium condition of contemporary art as an extension of the notion of people-as-medium put forth by Bourriaud and Bishop. *Third*, we will elaborate further on the consequences of the metamedium condition on the notions of contemporary art and creativity.

1. The Rise and Fall of Medium Essentialism

It is a long-held view that the properties of an art medium condition the aesthetic values of artworks and their proper evaluation. This view is called medium essentialism. According to Noel Carroll (2019: 35-36), this view can be dissected into two different of interpretations: material and praxeological.

The material interpretation of medium essentialism was proposed by late 19th and early 20th-century modernists, such as Eduard Hanslick in music, Rudolf Arnheim in film, and Clement Greenberg in visual arts. Its main tenets can be summarized in four theses: (1) each branch of the arts can and should be differentiated by its specific medium (e.g., two-dimensional image planes for painting, moving images for film, sonic events for music, bodily events in space for performing arts), (2) the medium constitutes the conceptual boundary separating the intrinsic and extrinsic aspects of an artwork, (3) art is produced by the interplay between the materiality of the medium and the artistic craftsmanship applied to it, and (4) aesthetic evaluations should be based primarily on considerations pertaining to the medium and the artistic exploration of that medium, or what Arnheim called "*Materialgerechtigkeit*" (i.e., "doing right by the material"), and only secondarily on extrinsic aspects such as socio-economic factors, broader cultural context, etc. This view had a lasting impact on the development of artistic modernism from the late 19th century up until the emergence of pop art and conceptual art in the 1950s.

The rise of pop and conceptual art posed a significant challenge to medium essentialism. Art practices based on the recontextualization of existing objects, such as Warhol's readymades, rather than the creation of new objects like modernist paintings, appear to constitute a very different notion of art and its medium. Pop and conceptual artists rely, in effect, on what could be called *second-order creation*, i.e., the creation of the context for existing objects, rather than *first-order creation*, i.e., the creation of new objects. Their medium is not the materiality of objects *per se*, but the audience's perception and expectations of these objects' roles in the world. They subvert conventional perception schemes and people's expectations by recontextualizing the system of objects appropriated from everyday life. Consequently, the conceptual framework of medium essentialism cannot cope with what Lucy Lippard in the 1970s called "the dematerialization of art" (Stephensen 2024: 133). Without a specific



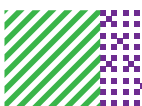
notion of materiality to rely upon, medium essentialist analysis cannot reduce pop and conceptual artworks to meditations on the medium and the artistic craftsmanship defined in relation to the medium. This marks the end of artistic modernism insofar as modernism was always conceived as the conquest of medium and the establishment of fixed boundaries between art disciplines delineated by their respective mediums.

In response to this development in artistic practices, a new approach developed in the 1960s: the institutional approach proposed by George Dickie and Arthur Danto. Dickie argues that artifactuality, i.e., the thinginess or materiality of objects, is only a necessary but not a sufficient condition of artworks. For something to be considered an artwork, it must possess an additional property beyond the material properties of the object in question. This extra-property is societal in nature. Danto (1964: 580) called it an “artworld”—an institutional context of perception consisting of an “atmosphere of artistic theory” and “knowledge of the history of art.” Similarly, Dickie proposes a definition of an artwork as “an artifact upon which some society or some sub-group of a society has conferred the status of candidate for appreciation” (Dickie 1969: 254). From this perspective, what constitutes an artwork is not merely the exploration of a specific artistic medium but the entire network of knowledge and social relations embodied in art institutions.

The praxeological interpretation of medium essentialism was developed in the last two decades by Berys Gaut, Dominic McIver Lopes, and Ted Nannicelli. This interpretation considers the view that the medium is constituted by practices rather than by the innate properties of the medium as a physical object. Gaut, for example, differentiates art photography from forensic photography by referring to the different sets of practices embodied in these two activities (Carroll 2019: 38). In doing so, the medium itself is redefined as something broader than mere materiality, encompassing what Nannicelli calls “a cluster of relatively stable, coherent practices of making something” (Carroll 2019: 41) or what Lopes called “a medium-centered practices” (Carroll 2019: 42). In all its manifestations, the praxeological interpretation of medium essentialism is essentially a retreat from the medium or an expansion of the notion of the medium into something beyond itself: praxis-as-medium.

2. From The Post-medium Condition to The Metamedium Condition

Reflecting on the works by conceptual poet-artist Marcel Broodthaers, Rosalind Kraus once proclaimed that, starting from the sixties onward, we are inhabiting “a post-medium condition” (Krauss 2000: 32). The surface flatness that Greenberg proposed as the most significant element in modernist painting now



gives way to disparate “layering of conventions” that constructs the discourse on art and its medium. What constitutes the artistic field is no longer the physicality of the medium but the social act of *détournement*, which reroutes and reassembles the order of things into new assemblages. This condition is characterized as the “leeching of the aesthetic out into the social field in general” (Krauss 2000: 56).

The implosion of the medium characterizing the post-medium condition manifests itself as the over-extension of the medium. The gatekeeping mentality inherent in the modernist attitude toward the hierarchy of genre and the demarcation of artistic fields has been dismantled almost completely over the last five decades. As a result, we now have a proliferation of mediums: weather as medium (Randerson 2018), computer code as medium (Levin & Brain 2021), the act of walking as medium (Morris 2020), even theory as medium (as reflected in the recent trend toward autotheory, e.g. Nelson 2015). If these disparate notions of the medium have a common underlying paradigm, it is that art is basically a relational phenomenon, not an individual thing. An artwork is a situation in which a social relation between people, to borrow from Marx (2010: 83), “assumes the fantastic form of relation between things.”

This social nature of artwork in post-medium condition is best exemplified in the relational and participatory aesthetics proposed by Nicolas Bourriaud and Claire Bishop. When Bourriaud (2002: 14-15) wrote that “it is no longer possible to regard the contemporary work as a space to be walked through,” or that contemporary art is a realm that takes “as its theoretical horizon the realm of human interactions and its social context,” he was highlighting the relational nature of contemporary post-medium condition. The same could be said with regard to Bishop’s proclamation that in contemporary art, “people constitute the central artistic medium and material” (Bishop 2012: 2). The relationality so pervasive in contemporary art, in the end, signals a new, all-encompassing notion of artistic medium: people-as-medium which consist of all social relations between people and between things mediated by people.

The rise of multimodal generative AI could be seen as the continuation and further realization of this broad category of people-as-medium. Nowadays, we have text-to-image AI such as DALL-E, Midjourney, and Stable Diffusion, text-to-video AI such as Runway and Sora, and text-to-music AI such as Suno and Google MusicLM. What do they have in common? They are all trained on big data, which consists of *social artifacts* scraped from all corners of the internet: images of people, portraits of animals, digitized books and articles, film footage, music and audio files, and so on. Trained in such diverse mediums, multimodal



generative AI processes all that complex information and transforms it into datasets.

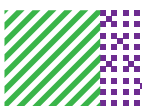
But what is a dataset? In essence, it is a collection of data points, i.e. a set of row and column containing information concerning an object. Large visual dataset such as LAION-5B, for example, contains 5.8 billion image-text pairs crawled from publicly available source in the internet.¹ It has a structure where its row contain the individual images and its column consist of image URL, captions, picture width and height, and so on. Common Crawl, one of the biggest text dataset with approximately 50 petabytes of data, contains 250 billion pages of text crawled from the internet spanning 17 years, with 3 – 5 billion new pages added each month.² Another example is YouTube-8M Dataset which consists of 6 million Youtube videos with accompanying metadata containing video title, description, tags, and so on.³

Looking at these huge datasets, one could ask: what is the nature of a dataset as an artistic medium? *First* of all, it is a *born-transmedial medium*. If we look at the metadata of those datasets, we can see that it is structured as a bundle of relations between different mediums; for example, the title column is paired with the image row, or the image column paired with the audio row, and so on. In fact, it is a medium made possible by the transmedial categorization that begets it. AI doesn't differentiate between a piece of music, painting, film, poetry, or bodily movement that some humans consider as dance. All those different mediums are the same for AI in so far as all can be transformed into a dataset, an arbitrary set of relations between arbitrary things. This transformation proceeds by two important steps: the tokenization process, which breaks down an artwork into its smallest parts such as words, pixels, or spectra of audio frequency, and the embedding process, which represents an artwork in a vector space containing strings of tokens that constitute it and its mathematical relation to another artwork, also represented as strings of tokens. The same can be said about the training processes using those datasets. An AI dataset is trained in a process that constitutes an ongoing and automatic *ekphrasis*, i.e., transmedial adaptation between different mediums (Panagiotidou 2022: 32). While in training, AI is in a state of constant, and potentially infinite, *remediation*: an ongoing representation of one medium in another that is proclaimed by Jay David Bolter and Richard Grusin (2000: 45) as “a defining characteristic of the new digital media.”

1 <https://laion.ai/blog/laion-5b/>

2 <https://commoncrawl.org/>

3 <https://research.google.com/youtube8m/index.html>



Secondly, it is an *assemblage* in the Deleuzian sense of the word. It has no innate preferences, no individuality, nor any fixed entry requirements. We could make anything into a dataset: be it things or relations, concrete or abstract, real or imaginary. It is the most arbitrary thing there is. Fully customizable, without rigid boundaries, fixed functions, or obligatory metadata structure, it can be properly called a *body without organs*, i.e. an assemblage constituted by a collection of assemblages without constituting the unity of an organism, “a body populated by multiplicities” (Deleuze & Guattari 2005: 30). AI dataset, thus, constitutes an anti-essentialist body which can be assembled and reassembled at will. This is made possible by the advances in machine learning technique, called deep learning, that enables the machine to identify patterns without step-by-step instructions, unlike symbolic AI which must be supplied with specific programming imperatives in order to execute reasoning. As a result, all data, labelled or unlabelled, categorized or uncategorized, can be used in the training.

Third important point concerning datasets as an artistic medium is that all these large datasets could be conceived as reservoirs of *general social knowledge* expressed or archived on the internet. In this context, I would like to suggest that AI is the realization of *general intellect* sketched by Marx in the seminal “fragment on machines” in *Grundrisse*. In it, Marx speculated that the development of fixed capital (such as machinery and technological implements) would arrive at a point where the “general social knowledge has become a direct force of production” and “the conditions of the process of social life itself have come under the control of the general intellect” (Marx 1993: 706). AI, in this sense, could be conceived as the accumulation of all social knowledge turned on its head: it no longer stands as a passive object of representation, an accumulated knowledge of past generations, but rather as an embodiment of the creative power of present society at large. Its medium is the people itself. With the development of super large datasets, AI can potentially be the reflection of our very own social existence.

These three properties of datasets as an artistic medium constitute what I would call the ‘metamedium condition.’ It is a condition of artistic practices where the medium is inherently transmedial, constituted arbitrarily as an assemblage, and reflects the general social knowledge. It is called a meta-medium in three senses: (1) it goes beyond the limitations of any physical mediums, much like ideas and concepts, (2) it is medium-neutral rather than medium-specific, i.e., different properties of the mediums don’t impact their structure as data, and (3) it is inherently social, or better, relational, in the sense that it expresses the general intellect or the general social knowledge. This is the condition of



artistic practices in the era of multimodal generative AI.

As such, metamediality is not too dissimilar to the Wagnerian notion of *Gesamtkunstwerk* understood as all-encompassing art form. If text could be transformed into image, image into sound, and back into text, if, in other words, there is a seamless transition between artistic mediums, then all art forms and genres are basically different instantiations of a single metamedium. The metamedium itself is an amorphous, *chora*-like receptacle, capable of forming into anything. As a consequence, it creates the condition for the *pan-aesthetics*, i.e. the aesthetics of universal transformation between art forms, be it in the form of physical mediums or social relations and ideas. An AI artwork, thus, requires an aesthetics which is not bound to a specific medium—an aesthetics of data transformations across all mediums.

3. The Meaning of Contemporary Art and Creativity in Metamedium Condition

When an AI system creates an artwork, or when an artist collaborates with an AI system to create an artwork, what is mobilized is not just people-as-medium, but also time-as-medium. This is because AI systems have an *archival logic* built into them; they are based on data archives, but not merely archives about the past in the usual sense—they can also be archives about the present, in the sense of real-time data. In this context, multimodal generative AI impacts the sense of the contemporary, which is essentially about the mode of temporal coming-togetherness.

“The contemporary” is one of the most debated concepts in contemporary art. The equation between the contemporary and postmodernity no longer holds today. Peter Osborne calls the theoretical matrix underlying the contemporary “the postconceptual condition.” It is constituted as a conjunction of the theses that art is inherently conceptual, that the medium is a necessary but not sufficient condition of artwork, that the potential mediums of art are potentially infinite, that every individual artwork has a relational nature (intertextuality with other artworks), and that the boundaries between artworks are historically malleable (Osborne 2013). This postconceptuality problematizes the conventional notion of the contemporary as “existing at the same time” because it presupposes the unity of time—the time of globalization—that abstracts away all spatial and geopolitical differences (Osborne 2018). In the postconceptual condition, the contemporary can only be experienced as fragmentary.

The postconceptual condition has some overlaps with the metamedium condition outlined here. The significant difference is that the postconceptual condition emphasizes continuity with the conceptual tradition of the sixties,



whereas the metamedium condition puts more emphasis on continuity with the relational and participatory aesthetics of the early 2000s. More importantly, the metamedium condition highlights the possibility of universal transformation between artistic mediums through deep learning, transforming every modality of art into another and, in doing so, creating new assemblages of people and time. It reveals the very process of time-making—the temporal coming-together of different people as a dataset.

The realization of this possibility could be seen in the multimodal generative AI which takes real-time data as its inputs. *Unsupervised - Machine Hallucinations - MoMA* (2022) is an example of this. Created by U.S.-based artist Refik Anadol, it is an ever-changing video installation that took 138,151 images from MoMA's collection as the training data and real-time data from MoMA's Gund Lobby, where changes in light, movement, and acoustics affects the transformation of the image. Reflecting on the question of medium, one could say that works such as this have *the contemporary* itself as the medium. It creates and re-creates the present by transforming real-time data into different modalities: from footsteps to luminosity, from air flow to color, from archival past to contemporary *Augenblick*. AI art, in other words, is the art of playing with the contemporariness of time.

What, then, is the meaning of creativity in the era of metamediality? In a romantic or modernist persuasion, creativity is often defined as an individual power to create something out of personal inspirations. Creativity is construed, in this context, as a power of individual artistic genius. We can call this *first-order creativity*. In conceptual or postconceptual art, creativity is seen as twice-removed from the objects of creation; an artwork-as-concept mediated by industrial readymades, for example. Conceptual artists, then, create the conditions of creation rather than directly creating objects. This is *second-order creativity*. An artist in the metamedium condition takes this step further by actively arranging diverse arrays of context windows to the *n*th-order in the creation of artistic objects: the context of data acquisition, corpus building, tokenization frameworks, representation technology, visualization or auralization methods, and so on. We can call this *n*th-order creativity or, borrowing from Eduardo Navas (2023: 231), a "metacreativity," which is a condition where "the creative process moves beyond human production to include non-human systems."

If a romantic poet is a first-order artist in the sense that he creates poetry as a "spontaneous overflow of powerful feelings," then a metamedial artist is an *n*th-order artist who works with *n*th-order of context windows to create artworks. Metacreativity, thus, can be understood as the commoning between different



agents in different mediums. It is an assembly of people and things, of objects and relations, of human and non-human. After all, creativity was never uniquely human. A cursory survey of the natural world displays bewildering examples, from beavers building massive networks of dams that can be seen from space to fungi creating subterranean networks spanning hundreds of kilometers that provide nutrients to the entire forest. Metacreativity is not the end of human creativity; it is the beginning of post-human creativity where humanity becomes part of a larger whole. In the end, it is creativity that is equivalent to the contemporary itself—the coming-together of beings.

As we move further into an era defined by multimodal generative AI, the boundaries between human and machine creativity continue to blur, redefining what it means to create art. The metamedium condition challenges traditional notions of medium specificity and opens up a space where artistic practices are not bound by physical constraints but are instead shaped by dynamic assemblages of people, data, and time. This new paradigm invites us to rethink the nature of artistic production, where creativity is no longer the sole domain of human genius but a collaborative process involving both human and non-human agents. In this unfolding landscape, art becomes a site of ongoing transformation and negotiation, where the interplay of diverse mediums and temporalities creates an ever-evolving reflection of our collective social and cultural knowledge. The metamedium condition, therefore, not only redefines the contemporary but also reimagines the very act of creation as a communal, fluid, and boundless phenomenon.



References

- Arielli, Emanuele. (2021). "Even an AI could do that". In L. Manovich & E. Arielli (Eds.), *Artificial Aesthetics: A Critical Guide to AI, Media and Design* (pp. 4-26): book manuscript.
- Bajohr, Hannes. (2024). *Dumb Meaning: Machine Learning and Artificial Semantics*. In E. Voigts, R. M. Auer, D. Elfein, S. Kunas, J. Röhnert, & C. Seelinger (Eds.), *Artificial Intelligence – Intelligent Art? Human-Machine Interaction and Creative Practice* (pp. 45-59). Bielefeld: transcript Verlag.
- Benanav, Aaron. (2022). *Automation and the Future of Work*. London: Verso.
- Bishop, Claire. (2012). *Artificial Hells: Participatory Art and The Politics of Spectatorship*. London: Verso.
- Bolter, Jay David, & Grusin, Richard. (2000). *Remediation: Understanding New Media*. Cambridge, MA: The MIT Press.
- Bourriaud, Nicolas. (2002). *Relational Aesthetics* (S. Pleasance & F. Woods, Trans.). Dijon: Les Presses du réel.
- Carroll, Noël. (2019). *Medium Specificity*. In N. Carroll, L. T. D. Summa, & S. Loht (Eds.), *The Palgrave Handbook of the Philosophy of Film and Motion Pictures* (pp. 29-48). Cham: Palgrave Macmillan.
- Danto, Arthur. (1964). *The Artworld*. *The Journal of Philosophy*, 61(19), 571-584.
- Deleuze, Gilles, & Guattari, Felix. (2005). *A Thousand Plateaus: Capitalism and Schizophrenia* (B. Massumi, Trans.). Minneapolis: The University of Minnesota Press.
- Dickie, George. (1969). *Defining Art*. *American Philosophical Quarterly*, 6(3), 253-256.
- Flick, Catherine, & Worrall, Kyle. (2022). *The Ethics of Creative AI*. In C. Vear & F. Poltronieri (Eds.), *The Language of Creative AI: Practices, Aesthetics and Structures* (pp. 73-91). Cham: Springer.
- Jones, Phil. (2021). *Work Without the Worker: Labour in the Age of Platform Capitalism*. London: Verso.
- Krauss, Rosalind. (2000). "A Voyage on the North Sea": *Art in the Age of the Post-medium Condition*. London: Thames & Hudson.
- Levin, Golan, & Brain, Tega. (2021). *Code as Creative Medium: A Handbook for Computational Art and Design*. Cambridge, Massachusetts: The MIT Press.
- Marx, Karl. (1993). *Grundrisse: Foundations of the Critique of Political Economy (Rough Draft)* (M. Nicolaus, Trans.). London: Penguin Books.
- Marx, Karl. (2010). *Capital Volume I*. London: Lawrence & Wishart.
- Morris, Blake. (2020). *Walking Networks: The Development of an Artistic Medium*. London: Rowman & Littlefield
- Navas, Eduardo. (2023). *The Rise of Metacreativity: AI Aesthetics After Remix*. New York: Routledge.

- Nelson, Maggie. (2015). *The Argonauts*. Minneapolis: Graywolf Press.
- Osborne, Peter. (2013). *Anywhere or Not at All: Philosophy of Contemporary Art*. London: Verso.
- Osborne, Peter. (2018). *The Postconceptual Condition*. London: Verso.
- Panagiotidou, Maria-Eirini. (2022). *The Poetics of Ekphrasis: A Stylistic Approach*. Cham: Palgrave Macmillan.
- Randerson, Janine. (2018). *Weather as Medium: Toward a Meteorological Art*. Cambridge, Massachusetts: The MIT Press.
- Stephensen, Jan Løhmann. (2024). The Marcel Duchamp Case in, against, or after Artificial Creativity. In E. Voigts, R. M. Auer, D. Elflein, S. Kunas, J. Röhnert, & C. Seelinger (Eds.), *Artificial Intelligence – Intelligent Art? Human-Machine Interaction and Creative Practice* (pp. 125-138). Bielefeld: transcript Verlag.

