THE SEMIOTIC OF AI IMAGES

Stephanie Schneider
Texas Tech University
Stephanie.Schneider80@gmail.com

Abstract

In her 1942 book *Philosophy in a New Key, A Study in the Symbolism of Reason, Rite, and Art,* American philosopher Susanne K. Langer distinguishes between two symbolic forms: discursive and presentational. The former is linear, discrete, and successive, while the latter is simultaneous and relational. She describes language as a discursive form and images as presentational. Langer argues that not all meaning can be communicated by discursive symbols; in particular, emotions cannot be expressed in discrete form. Instead, they find their symbolic form in artworks. This aligns with her aesthetic theory, which she elaborates on in her book *Feeling and Form* (1953), where she states her main proposition that art gives form to our feelings. Her distinction between the representational capabilities of language and images is well-suited for analyzing AI-generated images. Based on Langer, we can see two symbolic forms–the discursive and the

Digital Age in Semiotics & Communication, Vol. VIII, 2025, 84–102
https://doi.org/10.33919/dasc.25.8.5

presentational–collide in the process of text-to-image generation. Here, the image, as a presentational form, is created from a discursive form, i.e. a description (prompt), which, by Langer's definition, is incapable of communicating the same meaning as a presentational symbol. This article will explore the Semiotic of AI image generation based on her theory. The limitations of language as a discrete and linear form, and the resulting communicative constraints, are contrasted with images as presentational forms, such as artworks. This is not an evaluation of AI images, but rather an attempt to understand their structure and function within communication processes. The aim is to gain a new perspective on this medium and assess whether these images can help us "make our ideas clear" (Charles Peirce 1878).

Keywords: AI generated images, Art, Symbolic Forms, Susanne Langer, Presentational & Discursive Symbols

Introduction

The introduction of the first powerful image-generating artificial intelligence to the public has provoked intense debate. In 2022[1] and 2023[2], respected prizes were awarded to artworks created using artificial intelligence (AI), sparking intense discussions across the art world about the legitimacy of creating art with AI (see, for example, Parshall 2023; Roose 2022). With the growing dissemination of AI and its continuous use in image production across disciplines, the public debate has broadened, and the focus has shifted towards questions of evaluation and interpretation of AI generated images. Furthermore the ethical implications, copyright violations, and the effect on the job market, particularly for professions involving image generation are being discussed. Finally, the status of AI-generated images *as images* continues to be undefined.

The sustained critical discourse is of importance, since new technological developments are published at an ever-increasing rate: For example, in March 2025 Open AI integrated its image generation model DALL-E, into ChatGPT to improve image generation (Open AI 2025). At the same time, MIT reported on a new tool combining an autoregressive model and a small diffusion model to generate high-quality images faster (Zewe 2025).

[1] Jason Allen's A.I.-generated work, "Théâtre D'opéra Spatial".

[2] Boris Eldagsen's entry to the Sony World Photography Award, "Pseudomnesia: The Electrician".

Given the integral role of AI-generated images in the current technological transformation, there is a need for robust theoretical frameworks that can lead the study of this new generation of images. In order to facilitate an informed discussion regarding the effects of AI images on humans and society, it is necessary to first determine their specific characteristics. The present paper starts from the observation that AI image generation models produce images by using language. In order to generate the desired image, it is necessary to enter a prompt that instructs the AI model. Even image-to-image models that transform an existing image require a written prompt to provide instructions, and for a model to be trained, all images used for training the model need to be paired with a description. The technology has the capacity to generate images from natural language, thereby inaugurating a new form of image creation–one characterized by the absence of vision. The application of semiotic analysis to the domain of AI-generated images necessitates a consideration of the distinctive characteristics inherent to such images, with a particular emphasis on the relationship between text and image. The paper argues that there is a fundamental difference between the expressive qualities of a text compared to those of an image, which is exemplified by Joseph Kosuth in his artwork *One and Three Chairs* (1965).

The paper proposes the semiotic theory of Susanne K. Langer to explain this difference by distinguishing between discursive and presentational symbolic forms and provide a suitable theoretical framework for the study of AI images. This article will introduce the main argument she presented in her book *Philosophy in a New Key. A Study in the Symbolism of Reason, Rite, and Art* (1942), which is the differentiation between discursive and presentational symbols. This will be complemented by her aesthetic theory, introduced in *Feeling and Form* (1953) and her article *The Cultural Importance of the Arts* (1966), where she argues that art gives form to our feelings. This will guide the ensuing discussion on the (im)possibility of AI creating art. Subsequent to the introduction of the theoretical framework, the article will study the architecture of image-generating AI models, including historical references (without aiming for a comprehensive coverage). To test the applicability of the theory, the article will focus on examples drawn from art, a field in which a long and productive tradition of critical reflection of images and imaging technologies has been established. Even a cursory study of the art history of the late 19th and 20th century reveals that every significant technological invention has been extensively studied by artists. Wolf Vostell, Nam June Paik, and Frieder Nake, for example, have explored contemporary technologies such as photography, video, television, and computers in their artistic practice. The final paragraph offers some concluding remarks on the nature of AI images and their role in art and visual culture.

The Semiotic Theory of Susanne K. Langer

The American philosopher Susanne K. Langer, who was born in 1895 in New York is renowned for her book *Philosophy in a New Key. A Study in the Symbolism of Reason, Rite, and Art* from 1942. Based on the philosophy of Alfred N. Whitehead and Ernst Cassirer, she develops her unique approach to the theory of symbol, meaning, and the human mind. The key to Langer's symbol theory is the differentiation between discursive and presentational symbols. The first of these is characterized by linearity, discrete nature, and succession (cf. Langer [1942] 1953: 76), whereas the second one is marked by simultaneity and relational nature (cf. [1942] 1953: 86). Language is an example of discursive symbols; images and music are examples of presentational symbols. Langer emphasizes that not all meaning can be communicated by discursive symbols, especially emotions cannot be expressed in language. They find their symbolic form in presentational symbols.

Langer's theory is of particular interest in the study of images. Her writings offer a distinguished analysis of images as a non-textual symbolic form. In *Philosophy in a New Key*, she decidedly argues against the limitation of meaning or intellectual activity being attributed to language only. She states, "For there is an unexplored possibility of genuine semantic beyond the limits of discursive language" ([1942] 1953: 81). For Langer, presentational symbols adhere to the laws of logic, yet considerably expand the general notion of rationality. As she did not restrict thinking unilaterally to discursivity, she was able to take a holistic view of human mental activity and to include imagination, emotions, myth, and rituals in her theory: "The recognition of presentational symbolism as a normal and prevalent vehicle of meaning widens our conception of rationality far beyond the traditional boundaries, yet never breaks faith with logic in the strictest sense" ([1942] 1953: 89–90).

With reference to J. E. Creighton's article *Reason and Feeling*, she saw feeling as part of the process of understanding and cognition. At the same time, she noted that language is an inadequate medium for expressing emotions–those are better represented by presentational forms. According to Langer, the special characteristic of presentational symbols is that a multitude of concepts are brought together and conveyed in a single form. Images, for example, present their content not in a linear way, as text does, but in a relational form. We interpret an image by looking at it as a whole, then, as we enter the process of interpretation, work our way through the details. This is fundamentally different from reading, where we connect words, one after the other, into a meaningful context or a "Sinnzusammenhang", as it is called in German. Langer explains, "Wherever a symbol operates, there is meaning; and conversely, different classes of experience – say, reason, intuition, appreciation – correspond to different types of symbolic mediation" ([1942] 1953: 90).

Images are a familiar type of non-discursive symbols and, at the same time, a highly complex form of sign. They are coded in multiple layers, culturally, perceptually, and biologically. Unlike language, images are rich in detail; a photograph, for example, can convey more information than a description.[3] Langer argues that images communicate meaning in a unique way. Contrary to language, there is no vocabulary that would allow for definition (as in a dictionary) or translation, and the complex relationships cannot be broken down into single units in the same way a complex sentence could be subdivided into single words.

The elements of an image have no meaning independent from the individual context; their meaning only exists within the relationship of the composition in this very image (cf. ([1942] 1953: 87–88). However, images do not necessarily have to take on the form of the depicted object; they can differ in terms of color, surface, size, etc., since the image is a symbol and not a duplicate of the depicted object. Our extensive interpretive ability enables us to interpret even abstract pictorial representations. An image only needs to possess certain distinctive visual elements analogous to the properties of the object to be intelligible as a symbol. More details, however, will ensure the image's reference to a particular object (e.g., a portrait of a person).

As meaning only emerges from the concrete image, it cannot be separated from the perceptible material in which the image is realized. Each form, when placed in a new context, takes on a different meaning. There is no dictionary of lines, forms, or colors that explains a line (form, or color) using another line (form, or color). This means that presentational symbols lack generality. We can only arrive at the general concept through a process

[3] Langer's findings are supported by contemporary research. French computer scientist Yann LeCun, for example, compared the amount of information a human can retrieve through the visual system with the information a Large Language Model is trained on. According to his calculation, the information a four-year-old human has *seen* equals the data the biggest LLMs are trained on (Father of AI: AI Needs PHYSICS to EVOLVE…: 2025 n.p.).

of abstraction from an individual object or idea. An abstract concept, however, only becomes accessible through the perception or imagination of a particular presentation. According to Langer, the ability to see abstractly is the basis of human rationality, and since animals cannot interpret symbols, they are also unable to see images. The same can be said about computers and Large Language Models.

Langer's theory can be illustrated using Joseph Konsuth's artwork *One and Three Chairs* (1965). The artwork consists of three elements: a wooden folding chair, a photograph of the chair, and a definition of a chair from a dictionary. Only the definition was determined by the artist; the chair is interchangeable. Therefore, the work consists of the definition and the instructions for assembly. The chair is photographed in its respective exhibition context, so these two elements are subject to change. The photograph and definition are displayed as prints on the wall, and the chair stands next to them.[4] According to Langer's theory the dictionary definition of a chair is a discursive symbol, while the photograph is a presentational symbol of the actual object, the chair itself. The definition is general, therefore, it can be represented by different chairs, but can never be exhaustively represented by a single chair. The photograph and the chair, on the other hand, are in a direct, unique relationship to one another based on symbolization. While there is a visually perceptible similarity, the tactile properties have changed: the chair is no longer a three-dimensional wooden object, but rather paint on paper. Since the definition describes chairs in general, and the word chair in an extended sense–when it no longer refers to a piece of furniture, but to a position that a person occupies–it is clear that a general definition cannot convey any information about a specific object. As explained above, the general concept, such as a definition in a dictionary, is attainable only through abstraction.

To understand the general concept, an individual object is needed, for example, an image in the mind. When reading the definition without seeing the photograph or the chair, the reader will form an idea of the described object in their head that may differ from the one used in the artwork. In order to deliver a specific idea linguistically, a lengthy description of all the individual features of the chair would be needed, and due to the discrete and successive form of discursive symbols, the reader can only form their mental image as they read. The photograph, however, will deliver the idea

[4] The significance of the work for conceptual art and its art-historical references, e.g., the relationship with the works of Marcel Duchamp, are omitted here as they are discussed extensively in the relevant literature (see, for example, Foster et al.). The work of art is used here exclusively as an example to illustrate the theory.

of the specific chair in one act of perception. All essential information, such as color, material, and form, is delivered instantly. The presentational symbol, however, is limited to representing this specific chair. No information about other chairs is conveyed; only through abstraction are we able to conclude that there are other forms of chairs. This then extends beyond the current chair; we expect that all future chairs will have common characteristics and can be used in a certain way.

Langer's distinction between discursive and presentational symbols provides a suitable theoretical approach for analyzing AI-generated images. I argue with Langer that the two symbolic forms she describes collide here. The image, as a presentational form, is created from a discursive description which, by Langer's definition, is incapable of communicating the same meaning as a presentational symbol. The question of how images can be created from text may be translated into the question of whether it is possible to translate between linear and discrete discursive symbols and relational presentational symbols. In addition, there is the question of whether the resulting images can be considered art. To approach these questions, the next paragraph takes a look at image generation models (software) such as DALL-E 2.

The General Architecture of AI Image Generation

The concept of computer-generated images has been a subject of research for several decades. Frieder Nake, the German mathematician and computer scientist, is a pioneering figure in the field of computer art. In 1965, he designed an experiment using a drawing machine ("Graphomat Z64") with a computer connected to it and a program he had written for the computer (using a punched card) to study how the computer would draw. The drawing machine operated a writing head that could hold up to three pens for drawing. This resulted in the picture *Hommage à Paul Klee (13/9/65 No. 2)* by Frieder Nake (Nake & Grabowski 2005).

Lev Manovich has traced the use of AI in computer graphics back to Ivan Sutherland's *Sketchpad* (1961–1962), which "had a feature that would automatically finish any rectangles or circles you started drawing" (Manovich & Arielli 2024: 72). He argues that the history of digital media contains many such "AI Moments" (Manovich & Arielli 2024: 73).

When studying the history of digital media, and especially AI, one has to look past the 1960s to find the first steps in digital computing taken by Charles Babbage, Ada Lovelace and Charles S. Peirce (Ketner 1984) in the 19th century. The introduction of the concept of AI, however, is attributed to Alan Turing who published a paper discussing the question whether machines can think in 1950 (Turing 1950). The subsequent decades have been characterized by a dichotomy of progress and stagnation, with the 1970s being designated as the "AI Winter" due to the failure to meet expectations. However, the research continued at universities such as MIT and companies like Siemens and IBM until, eventually in the 2000s, it reached a new level of development (see, for example, Siemens: Tracing the AI family tree […] 2025 n.p.; Mucci 2024).

The arts have critically explored the technological inventions and discovered new forms of art alongside the technologies: today, the term "computer art" is rarely used and has been mostly replaced by the term media art or digital art. The underlying rationale for this phenomenon can be traced back to the evolution of technological media. The advent of the Internet, for example, led to the integration of personal computers into a networked infrastructure, thereby transforming them into components of a broader digital landscape. As Taylor argues in his book *When the machine made art : the troubled history of computer art*: "Art employing the latest digital technologies no longer relies on stand-alone computers, but is embedded in multiple devices, interacting globally with mobile and Web-based technologies" (Taylor 2014: 1). The artistic practice has evolved to employ a global infrastructure that is no longer restricted to a single computer and associated devices within the artist's studio. This is especially true for today's AI-generated images. AI models such as DALL-E are capable of image generation from a text input (prompt). They can also modify images, such as incorporating new elements into an existing image. To develop a more profound comprehension of the architecture of these models, the following paragraph takes a closer look at the essential elements.

An image generation model, such as DALL-E2 and 3, is composed of four basic elements: a text embedding, a prior, an image embedding, and a decoder (c.f. O'Connor 2023: n.p.; Ramesh 2022:3). The first element is the text embedding. IBM, for example, defines an embedding as "[…] a means of representing objects like text, images and audio as points in a continuous vector space where the locations of those points in space are semantically meaningful to machine learning (ML) algorithms" (Barnard 2023: n.p.). In general terms, an embedding converts different input formats such as text, images and audio into a numerical format. The embedding is essential for enabling ML models to find similar objects, meaning to learn patterns and relations in the data. Embeddings are created through a learning process using training data that, afterwards, can be integrated into other applications such as image generation. In the process of AI image generation, an embedding is needed to convert the natural language prompt from the user into a numerical format. To be able to link this prompt with a possible image, DALL-E2 employed a model called CLIP (**C**ontrastive **L**anguage-**I**mage **P**re-training), whose function was slightly modified after the integration of DALL-E into Chat-GPT (DALL-E 3) (O'Connor 2023: n.p). In its training phase, CLIP learns to relate images and descriptive text from a vast amount of images with descriptive labels: "[…] CLIP jointly trains an image encoder and a text encoder to predict the correct pairings of a batch of (image, text) training examples" (Radford 2021:2).

The 2025 Special Edition on AI by Scientific American illustrates the process with a simple example. To train a model to generate an image of a man on the beach with a dog, it is trained on many images showing a man walking on a beach with a dog, which are labelled with descriptive captions including the key words (c.f. Bushwick 2025: 24). However, it is important to remember that the model is not able to *see* those images. They are converted into pixel values to be utilized in the subsequent image generation process. Meaning is represented by the relation (similarity) between the numerical representation of the image and the numerical representation of the text description. After being trained on hundreds of millions of images from the internet, and often with no respect to copyrights, the model will then be able to correctly connect a keyword with an image. In the example of the man walking on a beach with a dog, it learns to relate single elements (man, dog, beach) of the image with their respective descriptions. While CLIP is capable of relating a description to an image, it is unable to generate images from a given description.

The next step in the process is to generate image embeddings using another model called a prior. This is a diffusion model, which is a neural network, meaning a network of computer neurons, defined as a mathematical function with a number of inputs and one output (Bergmann & Stryker 2024: n.p.). To process more complex tasks, those single neurons are arranged in layers. The input data for image generation is a grey-scale value of pixels generated in the embedding. If we have two input resources, they are mapped as one pixel onto a two-dimensional space, and as the number of inputs increases, the number of dimensions increases. Given enough input information, a meaningful image can be mapped. IBM offers the following summary,

> Diffusion models are generative models used primarily for image generation and other computer vision tasks. Diffusion-based neural networks are trained through deep learning to progressively "diffuse"

samples with random noise, then reverse that diffusion process to generate high-quality images (Bergmann & Stryker 2024: n.p.).

For image generation models such as DALL-E, diffusion models are trained in a process in which images are first deconstructed and then reconstructed. In the first step, clean data from a training set (e.g., an image sample) is transformed gradually into pure noise (imagine an old TV screen with faulty reception). This ensures that the essential qualities of the image are slowly fading, giving the model the chance to learn the patterns and structure of the original. The next step is to reverse the noise steps and generate a clean image. Since the noise was added gradually, the structure of the noise was derived from the original image: " Therefore, by learning to accurately predict the noise through reverse diffusion, the model not only learns to denoise the image, but also implicitly learns the structure of [the image]" (Bergmann & Stryker 2024: n.p.). It is then able to generate a new image from a random noisy image. Scientific American summarizes the process as follows, "Once trained, the AI can read any given text prompt, start with an image of pure noise, and reduce the noise until it has a new image that matches the written description" (Bushwick 2025: 25). The result of this step is the image embedding.

The final step uses yet another model called GLIDE, which acts as the decoder. GLIDE is another diffusion model, but with a modification that allows it to include textual information: "GLIDE builds on the generative success of Diffusion Models by augmenting the training process with additional textual embeddings. This results in text-conditional image generation. It's this modified GLIDE model that enables DALL·E 2 to edit images using text prompts" (Singh 2022: n.p.). The image generated with the GLIDE model undergoes an up-sampling process to create a high-resolution image from the preliminary small image, and with that final step, the image generation is completed. The latest development, as mentioned in the introduction, is the integration of an image-generating diffusion model like DALL-E3 into a Large Language Model (LLM) such as ChatGPT to optimize prompts for image generation (c.f. OpenAI).

Now that a general idea of the process of AI image generation has been established I will discuss this in the light of Susanne Langer's semiotic theory. It has been shown that AI image generation is based on text and numbers, both, in Langer's terms discursive symbols. The training of the model consists of the AI learning to relate written labels with a pattern, or a distribution of pixels represented by points in a multidimensional space. Those patterns can then be retrieved and assembled into new structures to generate new images that match the keywords in the user prompt.

AI Images in the Light of a New Key

The previous paragraph has demonstrated how an AI model operates to generate images. Although the model is trained on images, which have been characterized as presentational symbols, the model does not *see* them as images but as numbers and mathematical functions. The relationship between the image and its object–what it is showing–is established by descriptive labels. In terms of Langer's theory, we see the image treated as a discursive symbol. This establishes a form of dictionary to be able to define one symbol in terms of another. Moreover, the complex relational form, typical of presentational symbols, is broken down into its smallest parts (pixel), thereby divesting the image of its meaning. Contrary to visual perception, where the image is seen as a single form that conveys a multitude of concepts. To achieve the goal of AI image generation, images have to be treated as if they were discursive symbols to conform to the binary and linear architecture of computers and digital media. This contradicts the relational form of the presentational symbol: The meaning of the image is inferred in visual perception in a non-linear way. Moreover, as Langer explains, the significance of presentational symbols is derived solely from the individual image and the medium in which it is materialized. Every color, line, or shape takes on a different meaning upon its placement within the broader visual narrative. The same is true for music, where the context of a note's performance significantly affects its interpretation. It can be concluded that the AI model does not process presentational symbols; however, it does generate an image as a result of its processing. This raises more questions than I can discuss here, but I hope the article will serve as an inspiration for the reader. The question previously posed was how images could be created from a written text input and whether or not it is possible to translate between a linear and discrete discursive symbol and a relational presentational symbol. Studying the structure of current AI image-generating models, it can be shown that by translating the image into a mathematical function, it becomes possible to visualize a short text (prompt). However, the question remains whether this constitutes a translation of a discursive symbol into a presentational symbol. This question also relates to the earlier one asking whether the resulting image could be art.

To address this question, I suggest incorporating Langer's aesthetic theory into the discussion. In her book *Feeling and Form* (1953), she states her main proposition that art gives form to our feelings, on which she further elaborated in a later article titled *The Cultural Importance of the Arts* (1966). There she states,

> Art objectifies the sentience and desire, self-consciousness and world-consciousness, emotions, and moods that are generally regarded as irrational because words cannot give us clear ideas of them. (...) I believe the life of feeling is not irrational; its logical forms are merely very different from the structures of discourse. But they are so much like the dynamic forms of art that art is their natural symbol. Through plastic works, music, fiction, dance, or dramatic forms we can conceive what vitality and emotion feel like (Langer 1966: 10).

Art, she explains, objectifies feeling and makes it intelligible–a task discursive symbols such as language cannot render. Her argument here is a logical one, she states that the logical form of language does not reflect that of feeling. The incompatibility Langer sees is based on the complexity of feeling because, although we use words to denote a certain feeling, such as joy for example, the single word is incapable of communicating the nuances of a feeling we experience.

She states,

> But human feeling is a fabric, not a vague mass. It has an intricate dynamic pattern, possible combinations and new emergent phenomena. It is a pattern of organically interdependent and interdetermined tensions and resolutions; a pattern of almost infinitely complex activation and cadence (Langer 1966: 9).

The spatial character of human feeling she describes, resonates with her notion of the virtual space, the real but non-existing space that artworks as presentational symbols create (c.f. Langer 1953). Contrary to the idea of attributing the notion of virtual space to the realm of digital media, she describes it as the "primary illusion" (72) that is created by art. It is created, for example, by deploying colors on a canvas (cf. 1953: 71), but it is not the canvas or the colors themselves. The pictorial space is like the space behind the mirror, only visual, without continuity with the space in which we live. It is, as Langer puts it,

> […] limited by the frame, or by surrounding blanks, or incongruous other things that cut it off. Yet its limits cannot even be said to divide it from practical space; for a boundary that divides things always connects them as well, and between the picture space and any other space there is no connection. The created virtual space is entirely self-contained and independent (72).

Langer's concept of a virtual space can be illustrated with numerous examples from art history and contemporary art. One may imagine the landscape paintings from Georgia O' Keeffe or the late French-Lebanese artist Etel Adnan. Known as a writer and poet, she developed an intimate connection between language and visual art in her oeuvre. A recurrent motif in her oil paintings is Mount Tamalpais, located in California. In countless variations of color combinations, she painted the peak as seen from her Californian home in a style "continually dancing between figuration and abstraction" (Etel Adnan: Lights… 2021: n.p.). What does one see when looking at one of these paintings? According to the American founder of Semeiotic, Charles S. Peirce, we may see colors, shapes, and relations. But certainly not a landscape, a mountain, the sun, or the sky. This is what we *think* we see. Virtual space emerges from interpretation. We do not *see* the Californian Landscape, but we may be able to imagine it from what we see. To do so, humans do not need an image to be an exact copy of the depicted object or have a caption to label it. What enables humans to interpret an image is their experience; their ability to see, remember, and imagine, thus connecting memory, perception, and imagination. The virtual space is not the artwork but the intangible space of relations created by the artist and the beholder. The reality of the actual space and its objects, or physical movements as in a dance performance, are the *media* to access virtual space.

The virtual space created by artworks, as described by Langer, is not accessible for digital media, since the technology cannot see, remember or imagine. In short, digital media have no experience. A computer algorithm works in a discrete and linear way, and no AI model has its own memory. We may call a computer's storage capacity its memory, but stored data is not the same as past experience. The lack of experience results in a lack of contextual knowledge, which is crucial for understanding meaning, especially in presentational symbols that have no lexical translation. As a result, there is also no capacity for imagination:[5] AI models such as DALL-E2 and Midjourney generate new images based on probability instead of imagination.

A further point in this discussion is that since the model can only process a natural language prompt, the input the model gets from the user is already limited to the discursive form.[6] When overlooking the whole process, it becomes evident that there is no presentational knowledge or

[5] As recent research shows, memory and imagination are interrelated in the human brain. See for example Wickelgren 2023.

[6] This paper does not analyze image-to-image models, however they also do not process images as presentational symbols.

meaning involved. Thus, according to Langer's theory, AI neither meets the criteria for being an artist nor for producing artworks for it is not operating with presentational symbols, creating and interpreting virtual space as defined by Langer. Nevertheless, it is part of contemporary art production and theory. To define its role, not only in artistic but also in cultural production, the characteristics of this new symbolic form need further exploration. Similar to the 1960s when artists explored computers (but also TV and video) as new artistic media, AI-image generation is being tested today. Art created using AI models is crucial to understanding the characteristic potential and shortcomings of the technology and, as a result, its societal consequences. As Epstein and Hertzmann in their article *Art and the science of generative AI* summarize,

"Every artistic medium mirrors and comments on the issues of its time, and the debates surrounding contemporary AI-generated art reflect present issues surrounding automation, corporate control, and the attention economy. Ultimately, we express our humanity through art, so understanding and shaping the impact of AI on creative expression is at the center of broader questions about its impact on society" (Epstein & Hertzmann 2023: 1111).

AI Image Generation in Contemporary Art

Boris Eldagsen's entry to the Sony World Photography Award, "Pseudomnesia: The Electrician," was one of the images that sparked intense debate about AI in art in 2023. Taking a closer look at his artistic practice, it becomes obvious that the image generation is a multi-step process that involves text and image prompts, as well as "inpainting" and "outpainting" techniques (Eldagsen 2023: n.p.). The artist coined the term "promptography" for his artistic practice, an artificial term he defines as welding photography and prompt together to denote the use of language prompts and AI image-generating programs such as *DALL-E* and *Midjourney*. His practice, however, has–strictly speaking–nothing to do with photography. The suffix *-graphy* does not denote what is specific about photography, only the prefix *photo*. Upon closer inspection, we see what the process is about: drawing with prompts, meaning instructions given to an AI. Thus, the process of AI-image generation is fundamentally different from photography; the commonality lies in the appearance. AI images imitate or simulate a photographic image. The transition from photography to promptography is currently explored in exhibitions contrasting both imaging technologies.[7]

[7] See for example: RIVALS Photography vs. Promptography at Photo Edition Berlin c/o Galerie Guelman und Unbekannt.

The goal, Eldagsen proclaimed when refusing the Sony award–to spark a public debate of AI-image generation–was accomplished. And with that, a societal responsibility of artists, namely, to critically reflect on the media we use in our cultural communication, was fulfilled. However, it remains only a first step, and the term "promptography" shows that a detachment from photography and, as a consequence, a definition of this new form of image has not happened yet.

Another example is Refik Anadol's show at the Museum of Modern Art (MoMa), titled *Unsupervised* (2022–23). He trained an AI model using the publicly available data of MoMA's collection. The aim was not to imitate a man-made image, but to create something new by allowing the AI model to find new relationships. The process is not fully automated; in fact, the artist describes it as a "collaboration between machine and human" (Anadol et al. 2021: n.p.). It appears like a mutual blindness; the AI model does not *see* the artwork, whereas humans do not see the data points. This results in a new perspective on the collection, art history, and the technology itself. Using an AI model to search, sort, and classify the data, the artist created a "complex spatial map of the archive in 1024 dimensions" (Anadol 2022: n.p.). The questions, however, remain: In navigating this map and ultimately creating new forms does the AI enter the virtual space Langer describes? Or does it generate a virtual space from which it is excluded because it can only process a computational space? Or, is it obsolete to try to discriminate those spaces? In their exploration and experimentation with technology, artists are not relinquishing control of their artistry to algorithms and corporations that develop them. Instead, they are asserting their authority and agency in this domain. Many of today's questions have been discussed by artists before, including human agency in early computer art, abstract painting, and ready-made. Every inquiry into image technology invariably entails an inquiry into the nature of the image itself and its function as a sign. The 20th-century artists' proclamation to disengage from the image has, in fact, signified a re-entry to the discussion of images and their characteristics (cf. Bisanz 2010: 65).

With Susanne Langer, we could also speak of an exploration of the image as a virtual space that, by definition, cannot come to an end. The image as a virtual space is entirely independent of actual space because it is intangible and most of all relational. The dynamic nature of the virtual space as Langer describes it, may again best be illustrated by artworks. The delicate sculptures made by German artist Günter Haese of brass wire, coil springs, and other materials used in watchmaking are constantly reacting to their environment; every movement in their surroundings sets them in motion.

Therefore, they are a beautiful example of the vitality of virtual space. A more widely known example are the mobiles by Alexander Calder. Both exemplify another mode of virtual space, one that is not exclusively visual but also tactile. They make, as Langer puts it, "[…] tactual space visible" (1953: 90). The translation of tactile experience into a visual experience is another very important topic to discuss in the context of today's disembodied media, but that is to be left to another time.

Conclusion

The objective of this article was to explore the applicability of Susanne Langer's semiotic and aesthetic theory to the study of AI generated images. Her discrimination between discursive and presentational symbols facilitates a more precise terminology and consequently a lucid analysis. The discrete and linear nature of discursive symbols is in contrast to the relational character of presentational symbols. By identifying images as presentational and language as discursive symbols, it becomes apparent that AI models are incapable of processing images. Current systems translate images into numerical forms mapped in a vector space and generate images by calculating the most probable conformity with keywords, which have also been translated into numerical form and mapped onto that vector space. Given the framework of Langer's aesthetic theory and her concept of virtual space it was concluded that there is no evidence to support the idea of AI models producing art. At this point, the theory agrees with observations FRIEZE magazine made at museums worldwide, stating that exhibitions are "[…] moving the conversation away from the dead-end question of whether AI can make art to the question of what meaning artists can make with AI" (Droitcour 2024: n.p.). From a semiotic perspective, the question of how new technologies are altering the image as a medium remains unanswered. Since it was concluded here that AI-image-generating models cannot process presentational symbols–hence, images–the question of whether a generated image is, in fact, an image remains unresolved. From a strictly technical perspective, it could be considered a visualization of a text, which would make a meaningful difference. Images, and to a greater extent, artworks, play a pivotal role in our culture, having a profound influence on our collective visual environment. In light of this, it becomes imperative to foster a renewed emphasis on interdisciplinary research, integrating the insights of both the arts and the sciences to determine once again what images are, while recognizing the perpetual evolution of our symbols in the age of emerging technologies. It is our responsibility to ensure that symbols grow, as Charles S. Peirce once stated (Peirce 1878), and that they are not diminished by technology in their capacity to communicate meaning.

References

Adnan, E. 2021. Etel Adnan: Light's New Measure. *Guggenheim*. Available at: https://www.guggenheim.org/exhibition/etel-adnan-lights-new-measure (accessed 4 June 2025).

Anadol, R., C. Reas, M. Kuo and P. Antonelli. 2021. "Modern Dream: How Refik Anadol Is Using Machine Learning and NFTs to Interpret MoMA's Collection". *MoMa Magazine,* 15 November 2021. Available at: https://www.moma.org/magazine/articles/658 (accessed 4 June 2025).

Anadol, R. 2022. Refik Anadol on AI, Algorithms, and the Machine as Witness. *MoMa Magazine,* 20 December 2022. Available at: https://www.moma.org/magazine/articles/821 (accessed 4 June 2025).

Barnard, J. 2023. What is Embedding? *IBM*. Available at: https://www.ibm.com/think/topics/embedding (accessed 4 June 2025).

Bergmann, D., C. Stryker. 2024. What are diffusion models? *IBM*. Available at: https://www.ibm.com/think/topics/diffusion-models (accessed 4 June 2025).

Bisanz, E. 2010. *Die Überwindung des Ikonischen. Kulturwissenschaftliche Perspektiven der Bildwissenschaft*. Bielefeld: transcript Verlag.

Bushwick, S. M. Twombly, A. Hobbs. 2025. "How AI Generates Images From Text". *Scientific American Magazine,* Special Edition, Vol. 34, No. 1, Winter/Spring 2025, 24–25.

Droitcour, B. 2024. The Year in Review: Museums Are Leaving AI Hype Behind. *FRIEZE,* 10 December 2024. Available at: https://www.frieze.com/article/year-review-ai-art-2024?utm_source=FRIEZE&utm_campaign=ed5c (accessed 4 June 2025).

Eldagsen, B. 2023. "PSEUDOMNESIA III. Fake Memories of Art History". Available at: https://www.eldagsen.com/pseudomnesia3/ (accessed 4 June 2025).

Epstein, Z., A. Hertzmann. 2023. "Art and the Science of Generative AI". *Science,* Vol. 380, Issue 6650, 1110–1111, Available at: https://doi.org/10.1126/science.adh4451 (accessed 4 June 2025).

Foster, H., R. Krauss, Y.-A. Bois, B. Buchloh, D. Joselit. 2004. *Art Since 1900: Modernism, Antimodernism, Postmodernism*. London: Thames & Hudson.

Ketner, K. L. 1984. "The Early History of Computer Design: Charles Sanders Peirce and Marquand's Logical Machines." With the assistance of A. Stewart, *The Princeton Library Chronicle*, Vol. XLV, No. 3, 186–224.

Langer, S. K. 1942 [1951]. *Philosophy In A New Key. A Study In The Symbolism Of Reason, Rite, And Art*. 2nd edition. New York: Mentor Books.

Langer, S. K. 1953. *Feeling and Form. A Theory of Art Developed From Philosophy in a New Key*. New York: Charles Scribner's Sons.

Langer, S. K. 1966. "The Cultural Importance of the Arts". *Journal of Aesthetic Education,* Vol. 1, No. 1, 5–12. Available at: https://doi.org/10.2307/3331349 (accessed 3 June 2025).

Manovich, L., E. Arielli. 2024. *Artificial Aesthetics*. Available at: https://manovich.net/index.php/projects/artificial-aesthetics (accessed 4 June 2025).

Mucci, T. 2024. The history of AI. *IBM*. Available at: https://www.ibm.com/think/topics/history-of-artificial-intelligence (accessed 4 June 2025).

Nake, F., S. Grabowski. 2005. Zwei Weisen, das Computerbild zu betrachten. In Warnke, M., W. Coy, G. C. Tholen (eds.). 2005. *HyperKult. Zur Ortsbestimmung analoger und digitaler Medien*. Bielefeld: transcript Verlag.

O'Connor, R. 2023. How DALL-E 2 Actually Works. *AssemblyAI,* 29 September 2023. Available at: https://www.assemblyai.com/blog/how-dall-e-2-actually-works (accessed 4 June 2025).

OpenAI. Introducing 4o Image Generation (2025). Available at: https://openai.com/index/introducing-4o-image-generation/ (accessed 3 June 2025).

OpenAI. (n.d.). DALL·E 3. Available at: https://openai.com/index/dall-e-3/ (accessed 4 June 2025).

Parshall, A. 2023. How This AI Image Won a Major Photography Competition. *Scientific American*, 21 April 2023. Available at: https://www.scientificamerican.com/article/how-my-ai-image-won-a-major-photography-competition/ (accessed 3 June 2025).

Peirce, C. S. 1878. "Illustrations Of The Logic Of Science. Second Paper. How To Make Our Ideas Clear". *Popular Science Monthly*, Vol. XII, 286–302.

Radford, A., J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger and I. Sutskever. 2021. "Learning Transferable Visual Models From Natural Language Supervision". *ArXiv*, arXiv:2103.00020v1. Available at: https://arxiv.org/pdf/2103.00020 (accessed 4 June 2025).

Ramesh, A., P. Dhariwal, A. Nichol, C. Chu, M. Chen. 2022. "Hierarchical Text-Conditional Image Generation with CLIP Latents". *ArXiv,* arXiv:2204.06125v1. Available at: https://cdn.openai.com/papers/dall-e-2.pdf (accessed 4 June 2025).

Roose, K. 2022. An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy. *The New York Times*. Available at: https://www.nytimes.com/2022/09/02/technology/ai-artificial-intelligence-artists.html (accessed 3 June 2025). A version of this article appears in print on Sept. 3, 2022, Section B, Page 1 of the New York edition with the headline: "Creating Art By Typing In Some Words".

Siemens 2025 Tracing the AI family tree: say hello to ChatGPT's grandmother. Available at: https://www.siemens.com/global/en/company/insights/tracing-the-ai-family-tree-say-hello-to-chatgpts-grandmother.html (accessed 4 June 2025).

Singh, A. 2022. How Does DALL·E 2 Work?. *Medium*, 27 April 2022. Available at: https://medium.com/augmented-startups/how-does-dall-e-2-work-e6d492a2667f (accessed 4 June 2025).

Taylor, G. D. 2014. *When the Machine Made Art : The Troubled History of Computer Art. International Texts in Critical Media Aesthetics*. Vol. 7. New York, London: Bloomsbury.

This Is World. 2025. Father of AI: AI Needs PHYSICS to EVOLVE. prof. Yann LeCun. *YouTube*, 7 March 2025. Available at: https://www.youtube.com/watch?v=RUnFgu8kH-4&list=PL80I41oVxglJOokgsaIKw4ZJORQC0Qc_5&index=4 (accessed 3 June 2025).

Turing, A. M. 1950. "Computing Machinery and Intelligence". *Mind* 49, 433–460. Available at: https://courses.cs.umbc.edu/471/papers/turing.pdf (accessed 4 June 2025).

Wickelgren, I. 2023. Where Imagination Lives in Your Brain. *Scientific American*, 8 June 2023. Available at: https://www.scientificamerican.com/article/where-imagination-lives-in-your-brain/?utm_source=Klaviyo&utm_medium=campaign&utm_campaign=TIS_051925&utm_term=the%20purpose%20of%20memories%20is%20actually%20for%20future%20planning&_kx=G7ZVzZ4uR4H437KxN8yj8ye4BK7GqvSaRqWTVZPpFmZSK7F5u8_VzERn6kKRYBxo.WEer5A (accessed 4 June 2025).

Zewe, A. 2025. AI tool generates high-quality images faster than state-of-the-art approaches. *MIT News*, 21 March 2025. Available at: https://news.mit.edu/2025/ai-tool-generates-high-quality-images-faster-0321 (accessed 3 June 2025).