By now, being the summer of 2023, the development of generative AI has become so technically advanced and culturally accepted, that both researchers and media theorists agree that media generation with AI algorithms is not a passing trend but here to stay (Offert, 2022; Wilde, 2023).

Text-to-"insertmedium" models have become extremely powerful in generating media and are now popular beyond academia, research organizations or early adopters in creative AI. In the field of (generative) AI art, "the summer of 2022 introduced a moment of radical shift in the public awareness, mainly due to the fact that generative imagery since left the confinement and control of companies, research labs, or specialized artistic experiments, becoming available to the general public (Wilde, 2023, p. 7)" (Manovich, 2023; Wilde, 2023). In terms of artistic engagement with generative imagery, this period also marks the arrival of the so-called *Photoshop era* or even a new beginning of (generative) AI art (Offert, 2022, 2023; Wilde, 2023). Today's multimodal AI systems not only generate images, they also have built-in editing capabilities and the option to upgrade them with extensions that allow the user to control the outcome of these models in ways that were previously impossible. For example, in *Stable Diffusion*, which is one of the most advanced systems to date, it is now possible to generate images based on hand-drawn sketches or to work with personalized, pre-trained AI models. Users can even customize their own models by adding just a few sample images to the software (Wilde, 2023; Xiao, 2023).

The pace of development in this field is incredibly fast and surprises both observers and experienced researchers alike (Wilde, 2023). Therefore, this work aims to contextualize important developments in (generative) AI art, starting from its early days up to the introduction of multimodal AI systems, which have been identified as a "fundamentally new method" (Manovich, 2023, p. 35) in media production and can be seen as the most popular way of generating images to date (Manovich, 2023).

A Short History of (Generative) AI Art

General note: *The included figures have been heavily color edited by the author and often only slightly resemble the original references.*

A Short History of (Generative) AI Art

by Matthias Grund

By now, being the summer of 2023, the development of generative AI has become so technically advanced and culturally accepted, that both researchers and media theorists agree that media generation with AI algorithms is not a passing trend but here to stay (Offert, 2022; Wilde, 2023).

Text-to-"insertmedium" models have become extremely powerful in generating media and are now popular beyond academia, research organizations or early adopters in creative AI. In the field of (generative) AI art, "the summer of 2022 introduced a moment of radical shift in the public awareness, mainly due to the fact that generative imagery since left the confinement and control of companies, research labs, or specialized artistic experiments, becoming available to the general public (Wilde, 2023, p. 7)" (Manovich, 2023; Wilde, 2023). In terms of artistic engagement with generative imagery, this period also marks the arrival of the so-called *Photoshop era* or even a new beginning of (generative) AI art (Offert, 2022, 2023; Wilde, 2023). Today's multimodal AI systems not only generate images, they also have built-in editing capabilities and the option to upgrade them with extensions that allow the user to control the outcome of these models in ways that were previously impossible. For example, in *Stable Diffusion,* which is one of the most advanced systems to date, it is now possible to generate images based on hand-drawn sketches or to work with personalized, pre-trained AI models. Users can even customize their own models by adding just a few sample images to the software (Wilde, 2023; Xiao, 2023).

The pace of development in this field is incredibly fast and surprises both observers and experienced researchers alike (Wilde, 2023). Therefore, this work aims to contextualize important developments in (generative) AI art, starting from its early days up to the introduction of multimodal AI systems, which have been identified as a "fundamentally new method" (Manovich, 2023, p. 35) in media production and can be seen as the most popular way of generating images to date (Manovich, 2023).

Early Steps: Inceptionism & GANism

Earlier technologies used to create generative AI art, such as *Google Deep Dream (2015)* or *Style Transfer Algorithms (2016),* used neural networks to identify patterns within images and either enhance them or apply them to other images. Deep Dream, for example, was originally developed for scientists and engineers to better understand the vision of deep neural networks, before being repurposed as a creative tool (Morris, 2022; Zylinska, 2020).

The Deep Dream algorithm received notable attention for the distinctive psychedelic style it generated, which was described as *inceptionism.* However, the general interest in the technology quickly faded with the discovery of its unilateral visual feature, which often generated eyes or dogs within a regular image by enhancing its visual pattern (Morris, 2022; Zylinska, 2020).

Figure 1. Google Deep Dream example based on The Starry Night by Vincent Van Gogh (Cascone, 2016, n.p.) (edited by author)

Figure 2. Neural Style Transfer Examples based on the Mona Lisa (Finlay, 2021, n.p.) (edited by author)

The technological development of *GAN*, which stands for *generative adversarial network,* has revived and popularized interest in the artistic exploration of artificial intelligence, helping to characterize this art form as a more generative one (Scorzin, 2021).

GANs are generative machine learning models that can generate photorealistic images by learning to mimic the visual features of the input data. Technically, GANs consist of two neural networks, which are set in an adversarial relation. One of these neural networks, which is called the *generator model,* tries to generate a convincing output based on the input data, while the other neural network, called the *discriminator model,* evaluates the output against the real training data and classifies whether the output appears to be real or not. During the training process, both neural networks learn from their interactions and improve as they try to outperform the other. Over time, the output of the generator model should become indistinguishable from the real training data (Kana, 2020; Scorzin, 2021).

Starting with the invention of the first GAN by the computer scientist Ian Goodfellow in 2014, there are now a variety of versions of this technology, with their own specifications and different adaptations, which are constantly being further developed. One of the most prominent examples of its use and capabilities is the creation of photorealistic human faces with the so-called *styleGAN* architecture (Cohen & Giryes, 2022).

In recent years, GANs have become a fairly standard method of image synthesis in art, especially since the artist collective *Obvious* sold their work *Edmond de Belamy* for $432,500 in 2018. The GAN-generated and unfinished looking portrait of a man in antique clothing was the result of an algorithm trained on 15,000 paintings from different historical periods (Bajohr, 2022).

Figure 3. (top) Examples of GAN-Generated Human Faces (This Person Does Not Exist, n.d., n.p.) (edited by author)

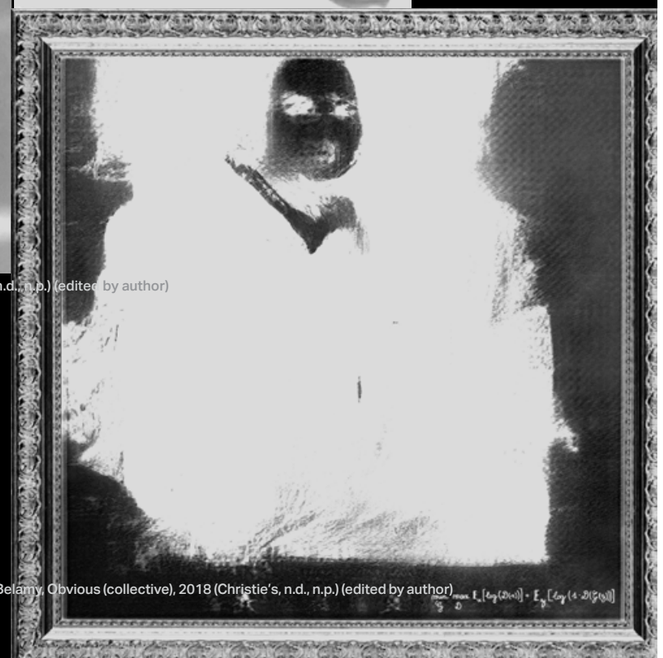

Figure 4. (bottom) Edmond de Belamy, Obvious (collective), 2018 (Christie's, n.d., n.p.) (edited by author)

The Latent Space (and its Limitations)

A crucial part of machine learning algorithms is called the *latent space.* To learn the similarities and distinguishable features of the input data, machine learning algorithms compress and simplify the representation of the data to find patterns within the dataset . The data is compressed into multidimensional vectors, where the most important information is stored as coordinates in a multidimensional space – the latent space. Similar data points are closer together within this space. The latent space representation of the data makes it manageable for the machine learning model to analyze and reconstruct the data as well as generate new, similar data (Tiu, 2020).

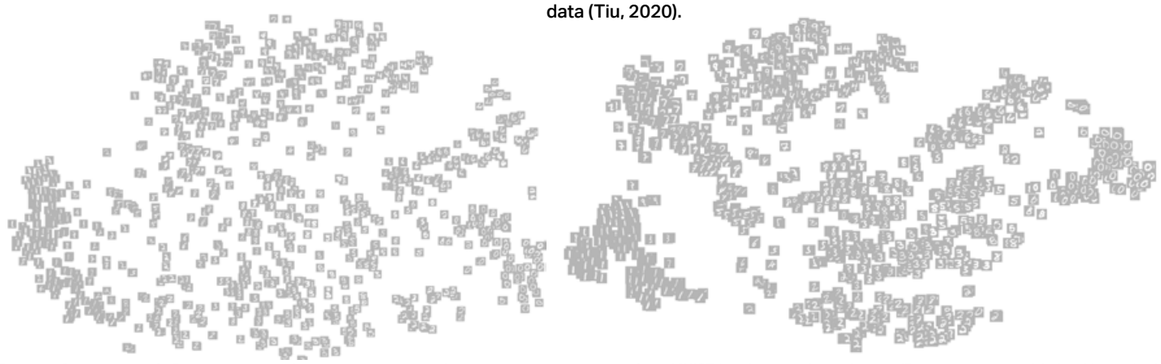

Figure 5. Comparison of 2D mappings visualizing image space vs. latent space representation from MNIST handwritten digits dataset (Despois, 2017, n.p.) (edited by author)

Literature and mathematics researcher Peli Grietzer explains that in the process of simplifying data into data points, some information is inevitably lost, "since the variety of possible media files is much wider than the variety of possible short lists of short numbers" (Grietzer, 2017, n.p.). He further points out, that machine learning algorithms therefore do not learn to create perfect reconstructions of their training set, but rather approximate reconstructions of it (Grietzer, 2017).

Fabian Offert, Assistant Professor of History and Theory of Digital Humanities at the University of California, Santa Barbara, describes this aspect more concretely when he explains that the "latent space is essentially a lossy compression of an image space, some features inevitably get lost in the training process and thus cannot be reconstructed" (Offert, 2021, p. 10) and further "we cannot know which features are lost in the training process" (Offert, 2021, p. 10).

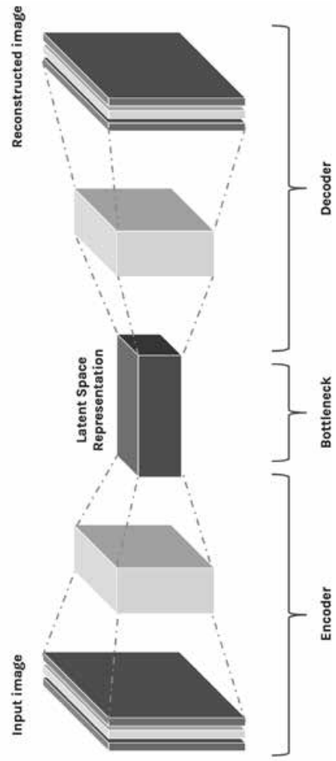

Figure 6. Visualization of the Convolutional Encoder-Decoder architecture (Despois, 2017, n.p.) (edited by author)

This technical process can be retrieved in the aesthetics of (GAN) generated imagery, which includes distinctive visual features such as seriality & variability, amalgamation, virtuality, mutations, and morphing (Scorzin, 2021). Also referred to as "GANism" (Chollet, 2017, n.p.), (generative) AI art shows striking similarities to contemporary mash-up and popular remix culture (Scorzin, 2021). Furthermore, the way machine learning algorithms learn and behave has been compared to the way vibes capture cultural patterns to make sense of the world (Grietzer, 2017), described as compost heaps (Salvaggio, 2021), or associated with dreaming (dark-taxa-project, 2021).

Generative machine learning models rely purely on their training dataset and their developed latent space, which as well describes their limitations. They simply "cannot reproduce what they have not seen" (Offert, 2021, p. 11). In terms of visual discovery, therefore, it can only take place within this image space, which is technically limited. Intelligent generative models "operationalize the epistemological distinction between invention and discovery by rendering the space of discovery a technically determined space. This determination is a historical determination: where [they] serve as a medium, what there is to know is what is already known" (Offert, 2021, p. 11).

Multimodal AI Systems

The possible generative space has expanded significantly with the invention of *multimodal neural networks.* These AI systems have the ability to learn and combine concepts from different modalities. Typically text and image based, they are able to receive information from one modality and use that knowledge to operate in another (Singer, 2022).

Also known as *text-to-image models,* they are trained on massive amounts of text-image pairs and can generate images of seemingly any

imaginable subject in any imaginable visual style from text input. The latent space of such generative AI systems seems to include the ability to render images of almost any conceivable phenomenon (McAteer, 2021; Singer, 2022).

It should be noted that generating media from text input goes beyond generating images. Other generative systems can apply text input, also referred to as *prompts,* to other forms of media such as video & animation, 3D models, music, or other text itself. However, text-to-image models are currently one of the most developed technologies (Manovich, 2023; Wilde, 2023). It is worth mentioning that there are numerous other AI methods for generating media. In terms of (generative) AI art, for example, "prompt-to-image generation is only one aspect of generative imagery, and there is also *image-to-image* generation or techniques like 'outpainting' that do not necessarily require linguistic input" (Wilde, 2023, p. 18).

"Text-to-another media methods are currently the most popular" (Manovich, 2023, p. 37), which has been argued to be due to the fact that most people are literate in one or more languages and that communication through language is generally considered natural (Manovich, 2023; Wilde, 2023).

Historically, the introduction of the two neural networks *DALL·E* and *CLIP,* in early 2021 by the research laboratory *OpenAI,* marked a major milestone in the field of image synthesis (Sutskever, 2021).

DALL·E (a combination of the artist Salvador Dalí and Pixar's WALL·E) is a 12 billion parameter transformer language model trained on a dataset of text-image pairs based on *GPT-3,* a large language model with approximately 175 billion parameters, that has been shown to produce realistic and human-like text. With the primary goal of generating images from given text prompts, and the ability to combine unrelated concepts and generate plausible images from text input (Ramesh et al., 2021), DALL·E "[replaced] model training and tuning as the artistic approach to image synthesis" (Offert, 2022, n.p.). Thus, replacing the previous GAN era with *prompt engineering* (Offert, 2022), a term that will be shortly discussed in the following.

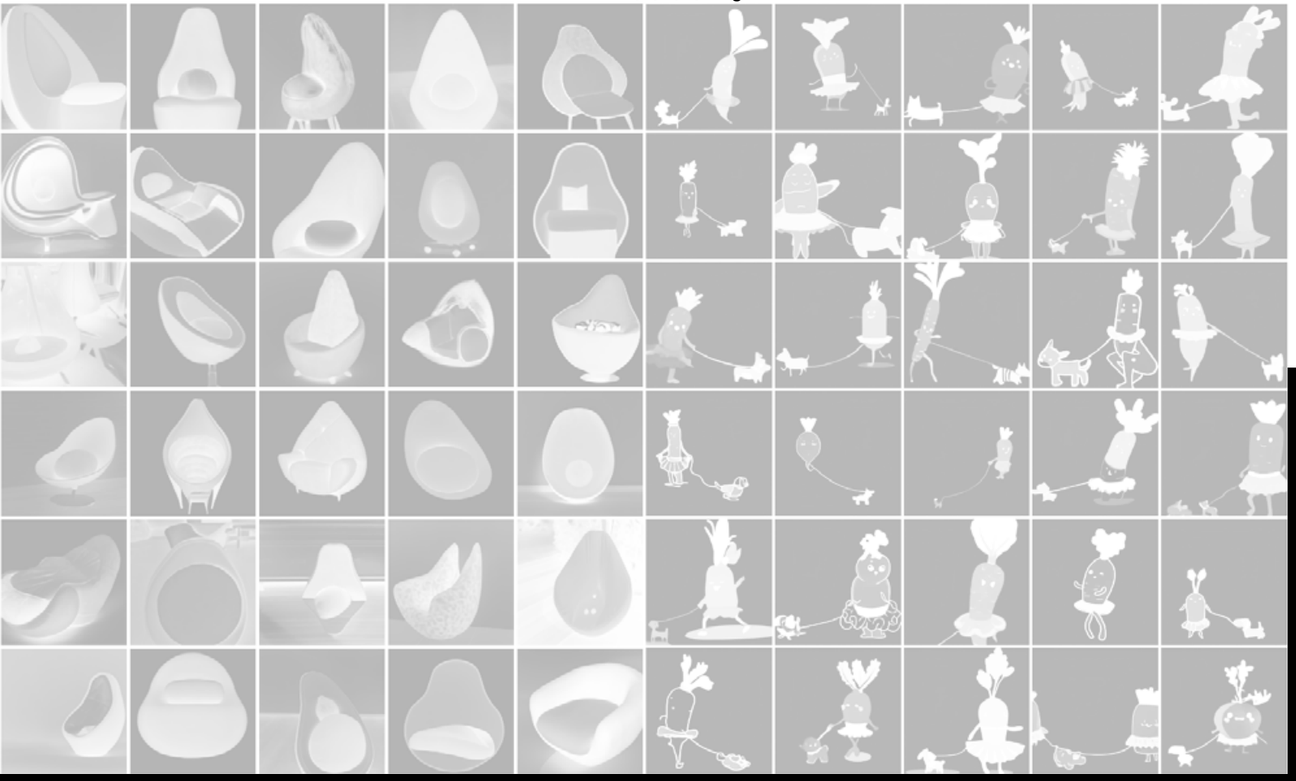

Figure 7. DALL·E generated images based on the text prompt: an armchair in the shape of an avocado. an armchair imitating an avocado (Ramesh et al., 2021, n.p.) (edited by author)

Figure 8. DALL·E generated images based on the text prompt: an illustration of a baby daikon radish in a tutu walking a dog (Ramesh et al., 2021, n.p.) (edited by author)

CLIP (Constrastive Language-Image Pre-training) is said to be an all-purpose classifier that connects text and images by classifying an image into a certain category and predicting which text caption of the dataset it is paired with. Essentially, it compares images and text and provides a similarity score. CLIP is able to complete this process in a *zero-shot* manner, meaning that is has the ability to perform classification tasks on unseen data by making predictions through the combined word vector embedding space of its training set (Radford et al., 2021). Pre-trained on 400 million publicly available web text-image pairs, CLIP encodes the given classes and is then able to link text input to image classifications of new data without having to be retrained on that specific data (Solawetz, 2022).

CLIP can be seen as the enabling technology for the so-called *artist-critic paradigm*, in which multimodal AI systems are able to classify whether the generated images match the entered text. This iterative process allows the generative model (artist) to refine its output based on the classifier's (critic's) evaluation, resulting in a closer correlation with the input text. As mentioned above, this practice was not possible with pre-CLIP generative models, as they merely represented the given training data with limited ability to guide what they produced, and could only be adjusted by manipulating and changing the training dataset (Morris, 2022).

Community-Driven Technology

A crucial aspect for the further development of multimodal AI models was the open source release of the CLIP code by Open AI, which, however, did not include the trained model with the training data. This led a group of AI enthusiasts to build a similarly large and then even larger dataset themselves, known as the *LAION-400M* and *LAION-5B,* and to train their own similar models using CLIP (Beaumont, 2022; Morris, 2022). In addition, various other researchers, developers, and artists created their own adaptations and combinations of various generative models connected to CLIP and published them openly online, in so-called *Google Colab*[1] notebooks, fostering a thriving AI art community online (Morris, 2022).

These community-created tools and methods not only enabled other researchers, developers, and artists to enter and develop the field of (generative) AI art, but also became very useful and inspiring for academic research, as shown by the example of *GLIDE,* another generative model developed by OpenAI that combines CLIP with diffusion models (Morris, 2022).

Diffusion is another generative machine learning model that uses a noise process to generate images based on the input data. The key concept of diffusion models is to "learn the systematic decay of information" (Siddiqui, 2022, n.p.) by gradually adding *Gaussian noise* to the input data and then reversing this process to "recover the information back from the noise" (Siddiqui, 2022, n.p.). This iterative forward/reverse process appears to be very successful in high-quality image synthesis and has taken over the field of image generation by providing superior quality and more diverse output compared to generative adversarial networks (Demochkin, 2021; Dhariwal & Nichol, 2021; Siddiqui, 2022).

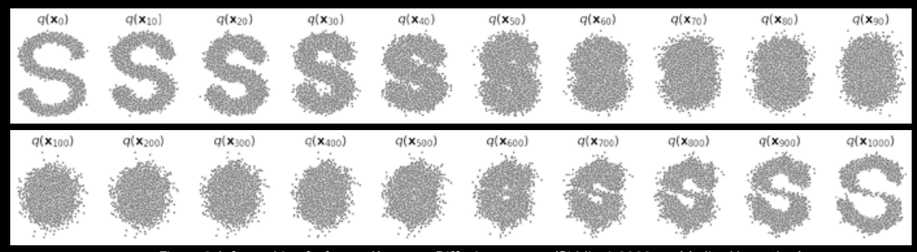

Figure 9. Infographic of a forward/ reverse Diffusion process (Siddiqui, 2022, n.p.) (edited by author)

1 *Colaboratory* is a Google-owned product that allows Python code to be written, shared, and executed in the browser without any installation or setup on the computer, and is often used for machine learning applications (Google Colab, n.d.).

Social Images

The method of guiding image synthesis with diffusion models using CLIP was originally discovered, developed, and published by artist Katherine Crowson in August 2021. To create GLIDE, OpenAI adapted Crowson's method and trained the system on its own unpublished training dataset. OpenAI often refers to Crowson's tweets in its research paper on GLIDE, emphasizing the importance of community contributions in this otherwise very monopolistic and scientific research-driven field, dominated by large tech companies (Morris, 2022).

The community's involvement in the development of tools and modifications has fostered a new and active Internet subculture in the field of (generative) AI art. Social platforms such as *Twitter, YouTube, Discord,* and *Reddit* are used for public discourse, with the community openly sharing their experiments, works, and resulting knowledge for everyone to access (Bajohr, 2022).

Platforms generally serve as facilitators for (generative) AI art and shape the process of creating generative imagery, which researcher and artist Andreas Ervik calls *AI imagenesis*. Services such as the latest version of DALL·E, namely *DALL·E 2*, for example, prohibit certain words or phrases such as nudity, violence, politics, or public figures in the text input to generate images, and regulate access to these tools by offering only a limited number of free trial images before users are required to sign up for a subscription to continue using the tool (Ervik, 2023).

However, alternative AI systems with similar capabilities, such as Stable Diffusion, are open source, can be installed on a personal computer, and can be customized, fine-tuned, and extended through additional community-built extensions (Ervik, 2023; Wilde, 2023).

With these systems, users are able to circumvent the regulations of commercial platforms, leading to a large amount of pornographic, violent and deepfake content, which is an ongoing problem that accompanies the progress of this technology. The generative models fine-tuned to explicit content, as well as their outputs, are openly shared and form a majority on general generative AI sharing platforms such as *Civitai,* which can be clearly observed by simply visiting its landing page (Civitai, n.d.). Although this phenomenon is not the focus of this piece of work, it is important to mention that AI porn must be seen as a separate subset of generative imagery, with various services offering subscription models to specifically generate this type of imagery. This content is increasingly populating social media platforms, especially those with more flexible rules regarding explicit content, such as Reddit (Shrivastava, 2023). The same is true for image manipulation in general, as in the age of generative imagery, the technical expertise required to create fake images that are difficult to distinguish from real photos is greatly reduced. Although the manipulation and retouching of photographs is nothing new and has been practiced for over 150 years, the notions of authenticity and trustworthiness of images are once again up for debate and may need to be rethought based on the massive amount of fake images that will continue to populate our media channels (Meyer, 2023b).

(Generative) AI art is essentially an interaction between the AI system, the user's input, and the platform owner, with shared agency. This can be best observed through *Midjourney,* another service for generative imagery. Midjourney operates through *Discord,* a social platform focused on instant messaging, and creates a social environment where users can interact with each other, generate images, and instantly share and discuss their results (Ervik, 2023). All three involved parties have a clear role within the social activity of image generation and influence the outcome in their own way. For example, within Midjourney, the generative model has a default aesthetic, a so-called *house style* (Manovich, 2023), which is so distinctive that other systems adapt it to such an extent that there is talk of a *midjourneyfication* (Meyer, 2023c). The active community members, who share their knowledge and the public manner in which images are generated, gradually shape the use of this system, while the platform provides specific features and (content) regulations (Ervik, 2023).

Ervik captures this phenomenon well, when he highlights that "the experience [of image synthesis through Midjourney] thus becomes undeniably social, but this applies to AI imagenesis in general. AI imagenesis is made possible by training data consisting of an enormous number of images, and the generated images are often shared in social networks, entering into ecosystems of likes, re-sharing, influencers, followers, trends, and algorithmic influence. AI creates a uniquely social form of images" (Ervik, 2023, pp. 49–50).

Promptism & Prompt Engineering

Figure 10. Selected examples of CLIP + VQGAN keyword comparisons by Twitter User @kingdomakrillic (Imgur, 2021, n.p.) (edited by author)

As quickly highlighted earlier in this text, the rise of text-to-image models has given birth to the notion of *promptism,* a term coined by artist Johannez, an early adopter of this technology, to describe the art movement resulting from the use of natural language to communicate with neural networks to generate artworks based on specific descriptions (Herndon Dryhurst Studio, 2022).

These textual inputs are commonly called *prompts* and can be defined as "written statements, acting as requests for the program to run its diffusion, detailing what the field of noise is supposed to coalesce into displaying" (Ervik, 2023, p. 6).

Early experiments with text-to-image models resulted in unexpected results and generally poor-quality images, leading to a new subfield of generative imagery called *prompt engineering,* also known as *prompt design.* Wanting to improve the output of generative models, the community experimented extensively with these new systems to understand what ideas, specific words, and structures would have what effect on the generated results. It became clear that "prompts can include descriptions of motifs of varying specificity, as well as stylistic registers and media technologies to be simulated" (Ervik, 2023, p. 6), and that there is an underlying syntax and semantics that lead to specific visual results. Consequently, over time, numerous tricks have been discovered to develop prompts and control the outcome of multimodal AI systems (McAteer, 2021; Merzmensch, 2022b).

According to machine learning engineer Matthew McAteer, the key to effectively controlling the results of generative models is to understand the language behind the text-image embeddings on which CLIP was trained. Most text-image models use CLIP as their classifier, which is trained on web data and makes it possible to identify structural patterns and specific parameters to develop successful prompts. However, due to the variety of AI systems that connect different generative models to CLIP, it is somewhat impossible to develop a universal prompt engineering guideline, as each system will generate different results when given a specific prompt (McAteer, 2021).

However, some services, such as Midjourney, do provide rough instructions on how to better structure your prompts to generate desired images. They also emphasize the importance of taking the time to create specific prompts, noting that "a well-crafted prompt can help make unique and exciting images" (Midjourney, n.d.-e, n.p.). While prompts can be as short as a single character, it is advised to be fairly specific and detail oriented. The shorter the prompt, the more random the result (Midjourney, n.d.-e). Community-created guides such as the *DALL·E 2 Prompt Book* and others, combine the shared knowledge of the AI art community and provide more detailed instructions, tips and hints on what is possible, and advice on how to *successfully* build a prompt (Diab et al., 2022; Parsons, 2022).

Before such prompt books existed, several members of the generative AI art Internet community developed and published extensive cheat sheets, tutorials, and sample catalogs comparing various styles and themes, sharing the knowledge gained from the many hours of experimentation and prompt discovery ((CLIP+VQGAN keywords), 2021; Harmeet G, 2022; Unlimited Dream Co., 2022a, 2022b, 2022c). Prior to DALL·E, one of the most extensive lists called Artist Studies, curated by artist and programmer *Remi Durant,* featured a collection of several hundred examples of artist names and the visual appearances they evoke, based on the then state-of-the-art VQ GAN + CLIP model architecture, developed simultaneously by the previously mentioned Katherine Crowson and artist Ryan Murdock (Crowson et al., 2022; Durant, 2022).

Figure 11. Selected examples of Remi Durant's Artist Studies list (Durant, n.d., n.p.) (edited by author)

Modifiers

Adding a distinctive visual style to one's prompt is probably the most important parameter in building a prompt, and the ability to generate images in essentially any style imaginable is one of the most prominent selling points of text-to-image models (McAteer, 2021; OpenAI, n.d.).

Multimodal AI systems do not simply apply a style to an image, but are able to differentiate the essential visual characteristics of a particular style connected to the prompt and apply those visual features to a different situation. For instance, by adding the phrase *in style by/of* to the text prompt, generative systems not only identify the stylistic elements, but also interpret and apply them in new contexts (Merzmensch, 2022a; Unlimited Dream Co, 2022b).

In generative AI art, a style is not limited to specific artists or art genres, but the visual features of any medium, environment, emotion, even historical period or more, can be added to almost any subject (Midjourney, n.d.-e; Midjourney, n.d.-a). In fact, as visual culture and media scholar Roland Meyer repeatedly argues, in generative imagery "everything becomes a ›style‹" (Meyer, 2023a, n.p.).

"Everything becomes a 'style', and while, in name, all these different 'styles' are still associated with people, media, genres, techniques, formats, places, or historical periods, in the production logic of the AI model they are nothing more than typical visual patterns extracted from a latent space of possible images accessed through generative (and often iterative) search queries" (Meyer, 2023d, p. 107).

By adding various other parameters, one can better adjust the output of generative systems and combine different concepts. Emphasizing or de-emphasizing certain elements, adjusting image quality or the accuracy of a prompt, setting aspect ratios, using specific model versions, or making use of other built-in editing features, is now possible and considered as common practice when using multimodal AI systems (McAteer, 2021; Midjourney, n.d.-d, n.d.-c; OpenArt, 2023; Wilde, 2023).

For example, by inserting a vertical bar glyph (|) or a double colon (::) between phrases, one can modify each phrase and submit the entire construction as a combined prompt. Adding numeric values such as -2; 1; 0.2; 0.0; -0.2; -1; -2, etc. to specific prompt parts will increase or decrease the effect of the associated phrase relative to the value of the other phrases. By slightly changing the seed value, it is therefore possible to generate a variety of similar images based on the initial output image (McAteer, 2021; Midjourney, n.d.-c).

Since the release of the second version of Stable Diffusion, it is especially recommended to add a second, *negative prompt* to the text input for undesired elements in the generated outcome (OpenArt, 2023). Text-to-image systems not only function with correctly spelled words or

Figure 12. DALLE·2 generated image based on the text prompt: Good morning in style of Arcimboldo. (OpenAI Labs, n.d., n.p.) (edited by author)

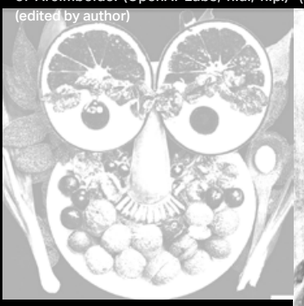

Figure 13. Vertumnus, Giuseppe Arcimboldo, 1591 (Wikimedia Commons, 2022, n.p.) (edited by author)

grammatically correct sentences and phrases; in fact, adding unusual word combinations, misspelled words, or even emojis, can lead to unexpected results or a change in the expected visual features of the correctly spelled term (McAteer, 2021; Russel, 2021). Finally, prompts do not have to be exclusively textual, as images themselves, alone or in combination with text, can also function as prompt input (Midjourney, n.d.-b).

It should be noted, that while different text-to-image models can be operated in a very similar manner, some functions, their modifiers, and the specific prompt syntax differ from system to system (McAteer, 2021; Midjourney, n.d.-d; OpenArt, 2023).

Prompting as a Process

In general, as Meyer (2023d) observes, multimodal AI systems for image generation operate somewhat like a black box. Their outputs seem unpredictable, as there is no clear prompt structure and no instructions to learn; a phenomenon that admittedly plays well into the fascination surrounding this new mode of image production.

"[Prompts] do not follow a standardized syntax, nor are they interpreted according to transparent protocols. Most importantly, they do not produce predictable and repeatable results. Rather, and this seems to be true of all diffusion models to date, one can never predict what specific image a particular prompt will produce, since minimal changes in the prompt will lead to visually completely different results, and even the exact repetition of a formula will conjure up ever novel, though in some respects similar images" (Meyer, 2023d, pp. 102–103).

Prompting can be described as the activity of searching the model's latent space; it is an iterative process of navigating the vast possibilities of image generation until images emerge that meet one's expectations or, better, one's preferences, since the result is more likely to surprise than to visualize exactly what one expected (Meyer, 2023d).

"The relationship between description and image seems to be less one of instruction and interpretation than one of navigation and matching: Verbal description does not determine what is to be produced, but functions as a means of narrowing down selections in a space of possibilities not yet realized" (Meyer, 2023d, pp. 103–104).

Hidden Language

The surprising nature of multimodal AI systems is also why it is possible, that in early June 2022, two researchers from the University of Texas reportedly discovered that DALL·E 2 uses "a hidden vocabulary that can be used to generate images with absurd prompts" (Daras & Dimakis, 2022, p. 1). Daras and Dimakis (2022) observed, that by asking DALL·E 2 to generate text output within the image, it would usually "lead to generated images that depict gibberish text" (Daras & Dimakis, 2022, p. 2). While rendering text is a common limitation of text-to-image models, the researchers found that these text fragments contain specific patterns and do not appear to be randomly constructed. They achieved this by generating images with specific textual content, such as written words or subtitles, and then copying and feeding these incomprehensible outcomes back into the model. For example, when prompted with "Two farmers talking about vegetables, with subtitles" (Daras & Dimakis, 2022, p. 2), the generated images included words such as *Vicootes* and *Apoploe vesrreaitais*. Prompting the term Vicootes led to images of vegetables, while Apoploe vesrreaitais led to images of birds. According to the researcher, this result could then be interpreted as "that the farmers [were] talking about birds that interfere with their vegetables" (Daras & Dimakis, 2022, p. 2).

Further examples revealed that within this vocabulary it is feasible to merge two separate concepts into one text prompt. However, the results of these experiments did not work consistently over the course of the research and often resulted in random or inconsistent outcomes (Daras & Dimakis, 2022), which further supports Roland Meyers' hypothesis.

With reference to Meyers' observations, one could explain that the somewhat consistent gibberish texts found in the research, formed by the combinations of letter shapes from the rendered words in the output images, were constructed from overrepresented visual features of images in the dataset that contained textual content in images of, for example, vegetables or birds. However, due to the opaque and complex nature of multimodal AI systems, such conclusions should be viewed as interpretations rather than explanations.

Figure 14. (right) DALLE·2 generated image based on the text prompt: Two farmers talking about vegetables, with subtitles. (Daras & Dimakis, 2022, p. 3) (edited by author)

Figure 15. (left) DALLE·2 generated images based on the text prompt: Apoploe vesrreaitais. (Daras & Dimakis, 2022, p. 3) (edited by author)

Attributing meaning to AI systems is generally controversial because there is no consciousness behind such systems. However, due to the complexity of multimodal AI systems, researchers are recognizing their computational capabilities beyond symbolic understanding and their ability to generate what researcher Hannes Bajohr (2023) calls *dumb meaning*[2].

However, as Meyer notes, regarding text-to-image models, they do not "show us … images of the world, but images of images – indeed, ultimately images about images, filtered through language" (Meyer, 2023d, p. 108).

Complex Systems

The level of complexity of so-called *foundation models*[3] can be exemplified by looking at the technical structure and image generation process of DALL·E 2, one of the current state of the art multimodal AI systems.

As mentioned above, DALL·E 2 is an improved version of the previously developed text-to-image generation systems DALL·E & GLIDE and can generate more caption-accurate and photorealistic high-resolution images (OpenAI, n.d.). Compared to GLIDE, DALL·E 2 produces a wider variety of images while maintaining a similar image quality (Ramesh et al., 2022). Furthermore, DALL-E 2 is equipped with additional features that allow to edit or transform an image by text input (inpainting & text diff), as well as to generate different versions that differ in composition but share distinctive visual features of a given reference image (Ramesh, n.d.).

Figure 16. Example of DALLE·2 inpainting feature: Original image vs. Adding a Corgi in different locations (OpenAI, n.d, n.p.) (edited by author)

Essentially, the key system on which DALL·E 2 is based, is called unCLIP, which uses CLIP and diffusion combined with a separate model called *Prior,* to generate images in a two-stage process. Aditya Ramesh, who created DALL·E and co-created DALL·E 2, describes the process on his personal website as follows: "In the first stage, a model which we call the prior generates the CLIP image embedding (intended to describe the 'gist' of the image) from the given caption. In the second stage, a diffusion model which we call unCLIP generates the image itself from this embedding" (Ramesh, n.d., n.p.).

The Prior is important due to of the infinite number of possible images that could be generated from the given text prompt. Therefore, the translation of the prior increases the likelihood of matching text and image embeddings (Ramesh, n.d.).

"During each step of training, unCLIP receives both a corrupted version of the image it is trained to reconstruct, as well as the CLIP image embedding of the clean image. This model is called unCLIP because it effectively reverses the mapping learned by the CLIP image encoder. Since unCLIP trained to 'fill in the details' necessary to produce a realistic image from the embedding, it will learn to model all of the information that CLIP deems irrelevant for its training objective and hence discards" (Ramesh, n.d., n.p.).

The process of learning the described irrelevant or non-essential details, enables the diffusion decoder to generate multiple image versions, within its aesthetic realm, according to the given image embedding. Together, with the CLIP embedding space, this allows for text-controlled image editing (Ramesh et al., 2022).

In conclusion, to explain the DALL·E 2 process in a simpler manner, one could say: To generate an image, CLIP first encodes the given text prompt, then the CLIP encoded text is fed to the Prior, which pre-selects a fitting image encoding that is then sent to the diffusion decoder. In a second step, the diffusion decoder generates the image from the given image encoding. To generate multiple versions of images and to perform tasks beyond text-to-image translation, the decoder learns more than the essential details of its given data.

2 In short, the term dumb meaning describes a nuanced level of meaning below human capabilities.

3 Foundation models are large machine learning models that are used as a basis and can perform a variety of tasks.

Other text-to-image models with similar generative abilities such as Google's *Imagen,* Stability AI's *Stable Diffusion,* and others, use slightly different approaches, however since all belong to the same domain, they are similarly complex (Ervik, 2023; Rajput, 2022; Wilde, 2023).

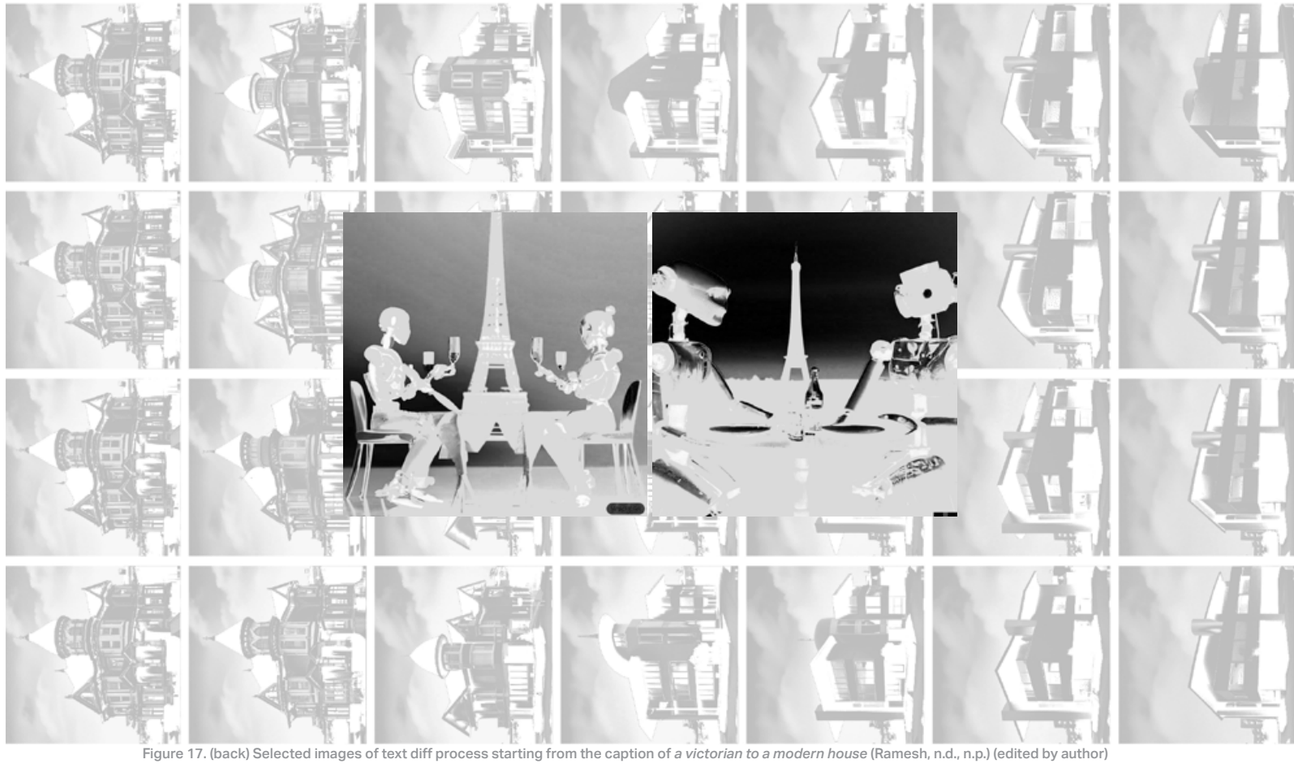

Figure 17. (back) Selected images of text diff process starting from the caption of *a victorian to a modern house* (Ramesh, n.d., n.p.) (edited by author)
Figure 18. (front) Side-by-side comparison of generated images by Imagen (left) and DALLE·2 (right) based of the prompt: A robot couple fine dining with Eiffel Tower in the background. (Hilton, 2022, n.p.) (edited by author)

Figure 19. Cosmopolitan Magazine June 2022 cover (Liu, 2022, n.p.) (edited by author)

Due to the complexity of this technology, and the general topic in all areas of artificial intelligence being the automation of human abilities (Manovich, 2023), discussions of *intellectual property,* and specifically in generative imagery, *aesthetics,* and *creativity* arise almost daily (Bolter, 2023).

The immense interest in generative imagery, as well as generative AI in general, has been greatly enhanced by the availability of AI tools (Bolter, 2023) and will most likely continue to grow, as generative AI methods are integrated into conventional creative software. First major implementations can be observed in *Adobe Firefly,* the "creative generative AI engine" (Adobe, n.d., n.p.) for *Adobe Photoshop,* one of the most popular photo and graphics editing software on the market (Adobe, n.d.).

While (generative) AI art is often presented as an independent, automated process, it is still entirely dependent on human involvement and effort, as Ervik (2023) points out with the example of the June 2022 issue of *Cosmopolitan Magazine.* The magazine's cover claims to be the "world's first artificially intelligent magazine cover" (Liu, 2022, n.p.), which "only took 20 seconds to make" (Liu, 2022, n.p.), while the final artwork was actually the result of a much more extensive and time-consuming creative process of building the right prompt and editing the final image (Ervik, 2023).

Musician and sound artist Holly Hernon and researcher and artist Mat Dryhurst, of Herndon Dryhurst Studio (2022), early adopters of generative AI, have coined the novel term *spawning* to describe the creative process of working with generative models, also known as prompting. Introduced as a contemporary update to the activity of sampling, in spawning, artists use text-to-image models to create something new by generating an output in the style of others; or in short, spawning describes the activity of "creating infinite new works from training data" (Dryhurst, 2023, n.p.). Andreas Ervik mentions that "the term spawning opens for an understanding of image generation as a co-creative process between the human and the generator. It is thus a form of computational symbiogenesis in which the genesis of the images is characterized by the symbiotic relationship between technology and humans" (Ervik, 2023, p. 49). Returning to Bajohr (2023) and the notion of dumb meaning, it is important to note that this attributed creative agency is not evenly distributed between the human and the generative model.

The artist duo shares Ervik's point of view, arguing that the developments in text-to-image models have finally made collaborative practices with machines possible. For Hernon and Dryhurst, the "act of conjuring artworks from language feels very very new" (Herndon Dryhurst Studio, 2022, n.p.). It is important to note that the duo puts a special emphasis on *feelings* being a crucial aspect of making art: "It feels like jamming, giving and receiving feedback while refining an idea with an inhuman collaborator, seamlessly art-ing. It intuitively feels like an art making tool" (Herndon Dryhurst Studio, 2022, n.p.). Before the introduction of multimodal AI systems, producing art with AI was a much slower and a more laborious activity, including the creation of data sets and a long training process (Herndon Dryhurst Studio, 2022).

The Issue of Consent

However, the sheer size of the data sets that are used to train multimodal AI systems, and the manner how that data is collected, highlights a fundamental problem with generative AI. Trained on a massive amount of data from the internet, generative imagery do not only have the tendency to reinforce existing stereotypes and biases -which is another topic worthy of its own discussion - but the data is also scraped without consent and therefore often contains copyrighted material (Meyer, 2023b).

A significant number of creative professionals from various disciplines consider this practice to be theft of their intellectual property and are worried about a further devaluation of cultural production. In fact, creative practitioners are not consulted before their work is included in large data sets, such as the LAION-5B and subsequently used to train models that automate creative and human labor (Chayka, 2023; Meyer, 2023b).

This phenomenon therefore has led to various lawsuits currently being filed against the companies behind generative AI systems. However, as Roland Meyer (2023b) recently pointed out, even if AI companies

become legally required to pay royalties for the work of others, the only ones likely to benefit will be large image right holders, such as stock photo companies. One can already observe such a phenomenon when looking at influential music streaming services such as *Spotify,* where major music labels and popular artists receive most of the revenue, while smaller artists are left with close to nothing (Ross, 2022).

In general, the internet and social media have enabled the emergence of generative AI, and the resulting medium of user-generated content has transformed cultural production into data assets produced as free labor. With AI-powered media, we are now experiencing the consequences of this media model, as this data is now being analyzed. This has evolved into an algorithmically driven media landscape (Dryhurst, 2023).

However, steps are already being taken to at least improve the situation for cultural producers, as Herndon Dryhurst Studio has launched the website *HaveIBeenTrained,* which allows artists to scan the LAION-5B dataset, that Stable Diffusion used to train their model, for their own work. Recently, they also launched an *opt out* service in collaboration with Stability AI, where cultural producers can remove their work from the training data used to train their updated system *Stable Diffusion 3* (Heikkilä, 2022).

Unlike others, Herndon Dryhurst Studio is not afraid of the possible replacement of the artist by artificial intelligence, seeing the artistic process as a social activity and much more complex than what multimodal AI systems are capable of (Dryhurst, 2023). They see generative AI as a collaborative tool that enables human creativity, a view shared by others, as media studies scholar Jay David Bolter notes: "Researchers in machine learning today, … seem to welcome the idea that AI systems would be used in collaborative relationships with human agents, rather than replacing humans altogether" (Bolter, 2023, p. 199).

Returning to the point of view of generative imagery as an art form per se, rather than as a tool or agent, the medium is formed by the process that begins with the creation of the model and ends with the activity of generating images. Bolter argues, based on its technological architecture, that this "medium is rooted in the principle of remix or remediation" (Bolter, 2023, p. 200) as it is both, based on other forms of media and generally "intermedial, [as it is] a blend of text and images that is both at the same time" (Bolter, 2023, p. 200). Unlike traditional remix practices, however, (generative) AI art involves less human intervention. Referring to the work of Andreas Ervik (2023), he adds that generative images are based on our *collective imagination* and at the same time shape it, as generative AI is not possible without prior data (Bolter, 2023).

In this sense, "images can no longer be understood as distinct (material or digital) artifacts, but instead appear as networked interfaces between human and non-human actors (including platforms, databases, and corporations) within [socio-technical systems]" (Wilde, 2023, p. 21), as media studies scholar Lukas R. A. Wilde argues.

Lev Manovich adds *predictive* attributes to generative AI, as multimodal AI systems attempt to generate prompt-accurate output. Furthermore, generative AI is not only *intermedial* but also *crossmedial,* since generative models can translate from one medium to another (Manovich, 2023).

A final essential aspect of generative images is that they are multiple and arbitrary in nature:

Text-to-image models are characterized by the fact that they usually generate more than one image for each text input. They generate a multitude of image versions in a matter of seconds, always with the possibility of generating an (almost) infinite number of others at will (Ervik, 2023; Wilde, 2023).

Therefore, one could agree with Roland Meyer, who states, that "for DALL·E, Midjourney, and Stable Diffusion the single image doesn't matter much" (Meyer, 2023d, p. 109).

To support the above point, AI-generated images are essentially arbitrary, since the outcome of text-to-image models is not precisely foreseeable. "The algorithmic 'blackbox' is part of their mediality" (Wilde, 2023, p. 14). The ability to choose from multiple options is due to the inability of precisely controlling the outcome of generative models (Wilde, 2023).

“Generative imagery is … remarkable perhaps not in quality but in quantity, speed, and availability as platforms like DALL·E, Midjourney, or Stable Diffusion can generate, through rapid feedback loops, an infinite number of pictures in all possible stylistic variations at incredible speed …. All the resulting individual pictures then seem so arbitrary and ephemeral that they hardly seem to deserve deepened individual attention or analysis“ (Wilde, 2023, p. 10).

Matthias Grund
Hidden Spheres, 2023
00:22:15, 1080 x 1080 resolution

The work *Hidden Spheres* provides a deep dive into the inner workings of a generative machine learning model by revealing the otherwise invisible structures of a deep neural network. For this work, a *Generative Adversarial Network (GAN)* was trained on a dataset of planetary imagery using the *StyleGan3* architecture, and then inspected using its interactive model visualization tool.

To learn about the visual features of the given dataset, machine learning models transform the input data through an iterative and layered process, where information is updated and passed onto the next layer.

By “zooming in“ on the trained model, this work brings the transformational qualities to the forefront, revealing the process of how such models learn simple features at the beginning of the training and learn to represent more complex and abstract features as the training progresses.

After an opening sequence that uses a *3D Ken Burns* effect that leads the viewer into the following microscopic view of this complex technology, Hidden Spheres gradually reveals the visual characteristics of the different layers involved in the neural network, starting with the output layer and gradually working backwards towards the initial input layer.

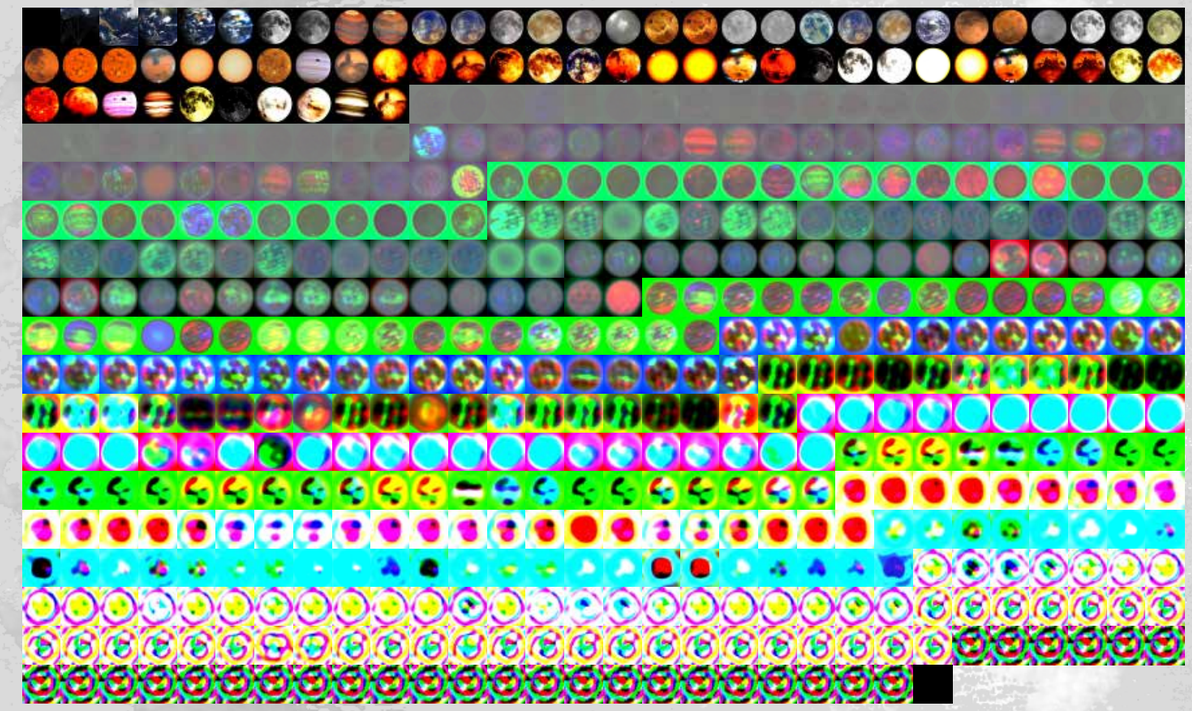

Figure 20. *Hidden Spheres*, exported frame selection

List of References:

Adobe. (n.d.). *Meet Adobe Firefly. Now in Photoshop.* Adobe. Retrieved June 4, 2023, from https://www.adobe.com/sensei/generative-ai/firefly.html

Bajohr, H. (2022, May 7). *Malen nach 0 und 1.* Republik. https://www.republik.ch/2022/05/07/malen-nach-0-und-1

Bajohr, H. (2023). Dumb Meaning: Machine Learning and Artificial Semantics. *IMAGE: The Interdisciplinary Journal of Image Sciences, 37*(1), 58–70. https://doi.org/10.1453/1614-0885-1-2023-15452

Beaumont, R. (2022, March 31). LAION-5B: A new era of open large-scale multi-modal datasets. LAION. https://laion.ai/laion-5b-a-new-era-of-open-large-scale-multi-modal-datasets/

Bolter, J. D. (2023). AI Generative Art as Algorithmic Remediation. *IMAGE: The Interdisciplinary Journal of Image Sciences, 37*(1), 195–207. https://doi.org/10.1453/1614-0885-1-2023-15472

Chayka, K. (2023, February 10). Is A.I. Art Stealing from Artists? *The New Yorker.* https://www.newyorker.com/culture/infinite-scroll/is-ai-art-stealing-from-artists

Chollet, F. (2017, July 13). *GANism (the specific look and feel of seemingly GAN-generated images) may yet become a significant modern art trend* [Tweet]. Twitter. https://twitter.com/fchollet/status/885378870848901120

Civitai. (n.d.). *Civitai.* Retrieved May 22, 2023, from https://civitai.com/

(CLIP+VQGAN keywords). (2021, July 23). *a comparison of roughly 200 CLIP+VQGAN keywords across 4 subjects full images here: https://imgur.com/a/SnSlQRu https://t.co/CjickItdlr* [Tweet]. Twitter. https://twitter.com/98_0634741763/status/1418440857578098691

Cohen, G., & Giryes, R. (2022). *Generative Adversarial Networks* (No. arXiv:2203.00667). arXiv. https://doi.org/10.48550/arXiv.2203.00667

Crowson, K., Biderman, S., Kornis, D., Stander, D., Hallahan, E., Castricato, L., & Raff, E. (2022). *VQGAN-CLIP: Open Domain Image Generation and Editing with Natural Language Guidance.* https://doi.org/10.48550/ARXIV.2204.08583

Daras, G., & Dimakis, A. G. (2022). *Discovering the Hidden Vocabulary of DALLE-2* (No. arXiv:2206.00169). arXiv. https://doi.org/10.48550/arXiv.2206.00169

darktaxa-project. (2021). *DARKTAXA-PROJECT: noPUBLICATION. [The whole digital insanity].* Verlag der Buchhandlung Walther und Franz König.

Demochkin, K. (2021, December 26). *75: Guided Diffusion.* Casual GAN Papers. https://www.casualganpapers.com/guided_diffusion_langevin_dynamics_classifier_guidance/Guided-Diffusion-explained.html

Dhariwal, P., & Nichol, A. (2021). *Diffusion Models Beat GANs on Image Synthesis* (No. arXiv:2105.05233; Version 4). arXiv. https://doi.org/10.48550/arXiv.2105.05233

Diab, M., Herrera, J., Sleep, M., Chernow, B., & Mao, C. (2022, November 13). *Stable Diffusion Prompt Book.* OpenArt. Retrieved June 2, 2023, from https://openart.ai/promptbook

Dryhurst, M. (2023, January 10). AI Art and the Problem of Consent. *ArtReview.* https://artreview.com/ai-art-and-the-problem-of-consent/

Durant, R. (2022, January 4). *I made this tool to make it easier to find good artist names to include in your prompts. I've just added a few more, and am now up to about 850 artist names! 😅 https://remidurant.com/artists/* [Tweet]. Twitter. https://twitter.com/remi_durant/status/1478216135514505219

Ervik, A. (2023). Generative AI and the Collective Imaginary. *IMAGE: The Interdisciplinary Journal of Image Sciences, 37*(1), 42–57. https://doi.org/10.1453/1614-0885-1-2023-15450

Google Colab. (n.d.). *Frequently Asked Questions.* Google Research. Retrieved May 5, 2023, from https://research.google.com/colaboratory/faq.html

Grietzer, P. (2017). *A Theory of Vibe.* Glass Bead. https://www.glass-bead.org/article/a-theory-of-vibe/?lang=enview

Harmeet G. (2022, March 25). Disco Diffusion Modifiers. *Weird Wonderful AI Art.* https://weirdwonderfulai.art/resources/disco-diffusion-modifiers/

Heikkilä, M. (2022, December 16). Artists can now opt out of the next version of Stable Diffusion. *MIT Technology Review.* https://www.technologyreview.com/2022/12/16/1065247/artists-can-now-opt-out-of-the-next-version-of-stable-diffusion/

Herndon Dryhurst Studio. (2022, May 6). *Infinite Images and the latent camera.* Mirror. https://mirror.xyz/herndondryhurst.eth/eZG6mucI9fqU897XvJs0vUUMnm5OITpSWN8S-6KWamY

Kana, M. (2020, April 20). Generative Adversarial Network (GAN) for Dummies — A Step By Step Tutorial [Medium article]. *Towards Data Science.* https://towardsdatascience.com/generative-adversarial-network-gan-for-dummies-a-step-by-step-tutorial-fdefff170391

Liu, G. (2022, June 21). The World's Smartest Artificial Intelligence Just Made Its First Magazine Cover. *Cosmopolitan.* https://www.cosmopolitan.com/lifestyle/a40314356/dall-e-2-artificial-intelligence-cover/

Manovich, L. (2023). AI Image Media through the Lens of Art and Media History. *IMAGE: The Interdisciplinary Journal of Image Sciences, 37*(1), 34–41. https://doi.org/10.1453/1614-0885-1-2023-15448

McAteer, M. (2021, December 20). *CLIP Prompt Engineering for Generative Art.* Matthew McAteer. https://matthewmcateer.me/blog/clip-prompt-engineering/

Merzmensch. (2022a, April 27). DALL·E: an AI Treasure Chest in Action: Creative and comprehensible capacities of Artificial Intelligence [Medium article]. *Towards Data Science.* https://towardsdatascience.com/dall-e-an-ai-treasure-chest-in-action-894c3a9cca92

Merzmensch. (2022b, May 28). A perfect Illustration with DALL·E: Art of Prompt Design: it's all about YOU. [Medium article]. *Merzazine.* https://merzmensch.medium.com/a-perfect-illustration-with-dall-e-550b9d941148

Meyer, R. (2023a, February 22). *Playing around with AI-generated, non-existent movies and the idea of , I started thinking about the notion of ›vibe,‹ how it might describe what I'm looking for in these images and what distinguishes it from concepts like ›style‹ or ›mood‹ 1/16* [Tweet]. Twitter. https://twitter.com/bildoperationen/status/1628438141420859394

Meyer, R. (2023b, June 2). Digitale Kunst: Bildermaschinen. *nd Aktuell.* https://www.nd-aktuell.de/artikel/1173690.digitalisierung-digitale-kunst-bildermaschinen.html

Meyer, R. (2023c, March 20). Es schimmert, es glüht, es funkelt - Zur Ästhetik der KI-Bilder. *54 Books.* https://www.54books.de/es-schimmert-es-glueht-es-funkelt-zur-aesthetik-der-ki-bilder/

Meyer, R. (2023d). The New Value of the Archive: AI Image Generation and the Visual Economy of "Style." *IMAGE: The Interdisciplinary Journal of Image Sciences, 37*(1), 100–111. https://doi.org/10.1453/1614-0885-1-2023-15458

Midjourney. (n.d.-a). *Explore Prompting.* Midjourney Documentation. Retrieved June 2, 2023, from https://docs.midjourney.com/docs/explore-prompting

Midjourney. (n.d.-b). *Image Prompts.* Midjourney Documentation. Retrieved June 2, 2023, from https://docs.midjourney.com/docs/image-prompts

Midjourney. (n.d.-c). *Multi Prompts.* Midjourney Documentation. Retrieved June 2, 2023, from https://docs.midjourney.com/docs/multi-prompts

Midjourney. (n.d.-d). *Parameter List.* Midjourney Documentation. Retrieved June 2, 2023, from https://docs.midjourney.com/docs/parameter-list

Midjourney. (n.d.-e). *Prompts.* Midjourney Documentation. Retrieved June 2, 2023, from https://docs.midjourney.com/docs/prompts

Morris, J. (2022, January 28). The Weird and Wonderful World of AI Art. *Notion.* https://jxmo.notion.site/The-Weird-and-Wonderful-World-of-AI-Art-b9615a2e7278435b98380ff81ae1cf09

Offert, F. (2021). Latent Deep Space: Generative Adversarial Networks (GANs) in the Sciences. *Media+Environment, 3*(2), 1–20. https://doi.org/10.1525/001c.29905

Offert, F. (2022, November 10). *Ten Years of Image Synthesis.* Zentralwerkstatt. https://zentralwerkstatt.org/blog/ten-years-of-image-synthesis

Offert, F. (2023). Five Theses on the End of AI Art. In *forthcoming* (p. forthcoming). forthcoming.

OpenAI. (n.d.). *DALL·E 2: DALL·E 2 is a new AI system that can create realistic images and art from a description in natural language.* OpenAI. Retrieved May 9, 2023, from https://openai.com/dall-e-2/

OpenArt. (2023, February 13). *The Most Complete Guide to Stable Diffusion Parameters.* OpenArt Blog. https://blog.openart.ai/2023/02/13/the-most-complete-guide-to-stable-diffusion-parameters/

Parsons, G. (2022, July 13). *The DALL·E 2 Prompt Book.* DALL·Ery GALL·Ery. Retrieved June 2, 2023, from https://dallery.gallery/the-dalle-2-prompt-book/

Radford, A., Sutskever, I., Kim, J. W., Krueger, G., & Agarwal, S. (2021, January 5). *CLIP: Connecting text and images.* OpenAI. https://openai.com/research/clip

Rajput, V. (2022, May 31). Google's Imagen vs OpenAI's DALLE-2 [Medium article]. *AIGuys.* https://medium.com/aiguys/googles-imagen-vs-openai-s-dalle-2-f760b60de800

Ramesh, A. (n.d.). *How DALL·E 2 Works.* Adityaramesh. Retrieved May 9, 2023, from http://adityaramesh.com/posts/dalle2/dalle2.html

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). *Hierarchical Text-Conditional Image Generation with CLIP Latents* (No. arXiv:2204.06125; Version 1). arXiv. https://doi.org/10.48550/arXiv.2204.06125

Ramesh, A., Pavlov, M., Goh, G., & Gray, S. (2021, January 5). *DALL·E: Creating images from text.* OpenAI. https://openai.com/research/dall-e

Ross, A. (2022, February 2). Reasons to Abandon Spotify That Have Nothing to Do with Joe Rogan. *The New Yorker.* https://www.newyorker.com/culture/cultural-comment/imagine-a-world-without-spotify

Russel, D. (2021, September 28). *CLIP invented an artist because I mispelled their name.... and that made-up artist seems super interesting Quoted tweet's prompt was `art by rene magirrite` Here's the corrected `art by rene magritte` https://t.co/J1kTzm7WDh* [Tweet]. Twitter. https://twitter.com/danielrussruss/status/1442786335127846913

Salvaggio, E. (2021, December 19). *Artificial Intelligence is a Compost Heap (Ideally): Looking beyond the human mind for new metaphors of intelligence* [Substack newsletter]. Cybernetic Forests. https://cyberneticforests.substack.com/p/artificial-intelligence-is-a-compost?s=r

Scorzin, P. C. (2021). ARTificiality: Künstliche Intelligenz, Kreativität und Kunst. In P. C. Scorzin (Ed.), *Kann KI Kunst? AI Art: Neue Positionen und technisierte Ästhetiken* (Vol. 278, pp. 50–75). KUNSTFORUM International.

Shrivastava, R. (2023, May 11). *How Real People Are Caught Up In Reddit's AI Porn Explosion.* Forbes. https://www.forbes.com/sites/rashishrivastava/2023/05/11/reddit-ai-generated-porn/

Siddiqui, J. R. (2022, May 2). *Diffusion Models Made Easy: Understanding the Basics of Denoising Diffusion Probabilistic Models* [Substack newsletter]. Towards Data Science. https://towardsdatascience.com/diffusion-models-made-easy-8414298ce4da

Singer, G. (2022, February 2). *Multimodality: A New Frontier in Cognitive AI: Enabling smarter, adaptive AI with innovative multimodal systems* [Medium article]. Towards Data Science. https://towardsdatascience.com/multimodality-a-new-frontier-in-cognitive-ai-8279d00e3baf

Solawetz, J. (2022, January 8). *How to Try CLIP: OpenAI's Zero-Shot Image Classifier: Earlier this week, OpenAI dropped a bomb on the computer vision world — you can now make image classifications with no training required.* [Medium article]. Towards Data Science. https://towardsdatascience.com/how-to-try-clip-openais-zero-shot-image-classifier-439d75a34d6b

Sutskever, I. (2021, January). *BLOG: Multimodal.* OpenAI. https://web.archive.org/web/20230218064737/https://openai.com/blog/tags/multimodal/

Tiu, E. (2020, February 4). *Understanding Latent Space in Machine Learning: Learn a fundamental, yet often "hidden," concept of deep learning* [Medium article]. Towards Data Science. https://towardsdatascience.com/understanding-latent-space-in-machine-learning-de5a7c687d8d

Unlimited Dream Co. (2022a, March 16). *Writing good VQGAN+CLIP prompts part one – basic prompts and style modifiers: This is the first in a series of guides where I'll go through some techniques on how to write better prompts, improve the quality of your generated images and get the best out of VQGAN+CLIP.* Unlimited Dream Co. https://www.unlimiteddreamco.xyz/articles/writing-good-prompts-part-1/

Unlimited Dream Co. (2022b, March 31). *Writing good VQGAN+CLIP prompts part two – artist and genre modifiers: Techniques and ideas to use artist and genre prompts with VQGAN + CLIP to shape your images and create better images.* Unlimited Dream Co. https://www.unlimiteddreamco.xyz/articles/writing-good-prompts-part-2/

Unlimited Dream Co. (2022c, May 5). *Writing good VQGAN+CLIP prompts part three – environmental modifiers: How to use environmental modifiers such as materials, lighting and surroundings to further refine your AI generated images.* Unlimited Dream Co. https://www.unlimiteddreamco.xyz/articles/writing-good-prompts-part-3/

Wilde, L. R. A. (2023). Generative Imagery as Media Form and Research Field: Introduction to a New Paradigm. *IMAGE: The Interdisciplinary Journal of Image Sciences, 37*(1), 6–33. https://doi.org/10.1453/1614-0885-1-2023-15446

Xiao, Z. (2023). *Dreambooth on Stable Diffusion* [GitHub Repository]. https://github.com/XavierXiao/Dreambooth-Stable-Diffusion (Original work published 2022)

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

List of References (Figures):

Cascone, S. (2016, March 2). Auctions: Google's 'Inceptionism' Art Sells Big at San Francisco Auction: You've never seen 'Starry Night' like this before. artnet news. https://news.artnet.com/market/google-inception-ism-art-sells-big-439352

Christie's. (n.d.). 2018 | LIVE AUCTION 16388: PRINTS & MULTIPLES: Edmond de Belamy, from La Famille de Belamy. Christie's. Retrieved June 11, 2023, from https://www.christies.com/lot/lot-6166184/?intObjectID=6166184

Daras, G., & Dimakis, A. G. (2022). Discovering the Hidden Vocabulary of DALLE-2. arXiv preprint arXiv:2206.00169, 1-6. https://doi.org/10.48550/ARXIV.2206.00169

Despois, J. (2017, February 23). Latent space visualization — Deep Learning bits #2. HackerNoon. https://hackernoon.com/latent-space-visualization-deep-learning-bits-2-bd09a46920df

Durant, R. (n.d.). Artist Studies by @remi_durant. Remi Durant. Retrieved June 11, 2023, from https://remidurant.com/artists/#

Finlay, P. J. (2021, February 1). Neural Style Transfer Images. [GitHub Repository]. https://github.com/PJ-Finlay/Neural-Style-Transfer-Images

Hilton, B. [@benjamin_hilton]. (2022, May 25). 6. "A robot couple fine dining with Eiffel Tower in the background."(L: Imagen, R: DALL-E) [Tweet]. Twitter. https://twitter.com/benjamin_hilton/status/1529510726997622784

Imgur. (2021, July 23). CLIP + VQGAN keyword comparison by @kingdomakrillic. Imgur. https://imgur.com/a/SnSlQRu

Liu, G. (2022, June 21). The World's Smartest Artificial Intelligence Just Made Its First Magazine Cover. *Cosmopolitan.* https://www.cosmopolitan.com/lifestyle/a40314356/dall-e-2-artificial-intelligence-cover/

OpenAI. (n.d.). DALL·E 2: DALL·E 2 is a new AI system that can create realistic images and art from a description in natural language. OpenAI. Retrieved May 9, 2022 from https://openai.com/dall-e-2/

OpenAI Labs (n.d.). DALL·E 2: My collection: "Good morning in style of Arcimboldo.". OpenAI Labs. Retrieved June 11, 2022 from https://labs.openai.com/s/x6w3jiwlXPDcyK9ZpWchdccT

Ramesh, A. (n.d.). How DALL·E 2 Works. Adityaramesh. Retrieved May 9, 2022 from http://adityaramesh.com/posts/dalle2/dalle2.html

Ramesh, A., Pavlov, M., Goh, G., & Gray, S. (2021, January 5). *DALL·E: Creating images from text.* OpenAI. https://openai.com/research/dall-e

Siddiqui, J. R. (2022, May 2). Diffusion Models Made Easy: Understanding the Basics of Denoising Diffusion Probabilistic Models. Towards Data Science. https://towardsdatascience.com/diffusion-models-made-easy-8414298ce4da

This Person Does Not Exist. (n.d.) Image. This Person Does Not Exist. Retrieved June 11, 2023, from https://thispersondoesnotexist.xyz/

Wikimedia Commons (2022, May 4). File:Vertumnus årstidernas gud målad av Giuseppe Arcimboldo 1591 - Skoklosters slott - 91503.tiff. Wikimedia Commons. Retrieved June 11, 2023, from https://commons.wikimedia.org/wiki/File:Vertumnus_%C3%A5rstidernas_gud_m%C3%A5lad_av_Giuseppe_Arcimboldo_1591_-_Skoklosters_slott_-_91503.tiff

Bundesministerium für Bildung und Forschung

Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen

KISD Köln International School of Design

Technology Arts Sciences TH Köln

This publication was created at *Köln International School of Design* as part of the joint project *KITeGG - Making AI tangible and comprehensible: Connecting technology and society through design.* Funded by the *German Federal Ministry of Education and Research (BMBF)* and the State of *North Rhine-Westphalia.*

Diese Publikation entstand an der *Köln International School of Design* im Rahmen des Verbundprojektes *KITeGG - KI greifbar und greifbar machen: Technologie und Gesellschaft verbinden durch Gestaltung.* Gefördert durch das *Bundesministerium für Bildung und Forschung (BMBF)* und das *Land Nordrhein-Westfalen.*