

Jay David Bolter

AI Generative Art as Algorithmic Remediation

Abstract: As the essays in this collection demonstrate, AI generative imagery raises compelling theoretical and historical questions for media studies. One fruitful approach is to regard these AI systems as a medium rooted in the principle of remediation, because the AI models depend on vast numbers of samples of other media (painting, drawing, photography, and textual captions) scraped from the web. This algorithmic remediation is related to, but distinct from earlier forms of remix, such as hip-hop. To generate new images from the AI models, the user types in a textual prompt. The resulting text-image pairs constitute a kind of metapicture, as defined by William J.T. Mitchell in *Picture Theory* (1994).

Introduction

The quality of the essays in this collection attests to the rich potential of generative AI for media studies. Even if AI imagery threatens the practices and perhaps livelihoods of designers and artists – a question on which opinions differ – it certainly does not threaten media studies researchers, who are already responding creatively to the theoretical and historical questions posed by this relatively new practice. These essays provide evidence that we cannot understand AI simply as a threat; instead, they attest to the complexity of our visual culture before these generative programs became available and address the ways in which AI imagery may participate in that culture.

Since the essays engage with DALL·E 2 and the other generative image systems on a variety of levels, we might begin by giving DALL·E 2 itself a chance to engage: that is, by submitting the titles of a few of the essays or the earlier presentation titles as prompts and seeing what kind of images emerge. The titles are not the kind of phrases typically submitted to DALL·E 2, and the results do not seem typical either. Pamela Scorzin’s presentation title “Meta-Images and Meta-Humans” (cf. SCORZIN 2023) produces a visually coherent result (cf. fig. 1). The depiction of human figures taking pictures of pictures does suggest the self-referential

quality of the prompt. In this case, at least, DALL-E 2 seems to be functioning as its makers intended: we could imagine using this image on a book cover for a monograph on meta-images and meta-humans. Two further examples are harder to interpret as responses to their texts. Andreas N. Ervik’s presentation title “Towards an Ontology of AI Generated Images” (cf. ERVIK 2023) produces the image we see in figure 1 in the middle, and Hannes Bajohr’s title *Dumb Meaning: Machine Learning and Artificial Semantics* (2023a) gives us perhaps the most humorous image (cf. fig. 3 on the right). What is notable about the last two results is that what appear to be alphabetic symbols have found their way into the images. How text functions here is suggestive of the ontology of these generative programs, as we will discuss below.

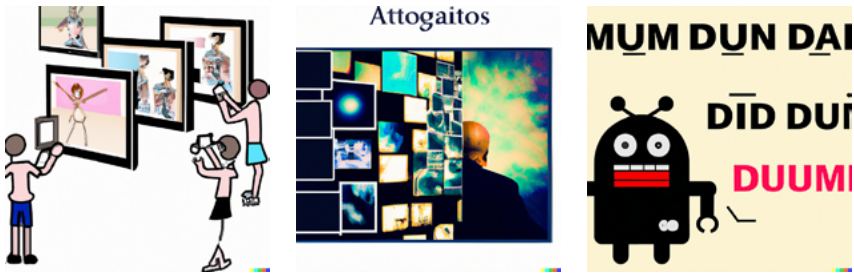


Figure 1: DALL-E 2 creations with prompts based on titles of presentation of the Tübingen-workshop, generated in February 2023

Debating the Status of AI Imagery

It is not only media studies scholars who are fascinated with generative AI; there is also enormous popular interest in the technology, particularly since the private company OpenAI has released DALL-E 2 and then more recently ChatGPT. Although the technologies and even some of the manifestations of AI image generation and art stretch back years, it is only since last fall (2022) that everyone seems to be talking about them in blog posts, podcasts, and mainstream newspapers and magazines (such as the *NYTimes*, *The Economist*, and *Der Spiegel*). Two interrelated issues are of greatest interest:

1. The question of intellectual property: What is the legal status of these generative images? Are they original or derivative works? The systems do not generate images *ex nihilo*. They draw on millions or even billions of text-image pairs scraped from the web for their underlying databases, such as, for example, LAION used by Stable Diffusion and some other systems. Do

they therefore infringe on the rights of the human artists and producers whose works were scraped from the web and fed into the model?

2. The aesthetic issue: are these generated images creative artifacts at all? And where does the creativity reside? Are they art? And if so, who is the artist?

Although the first of these, the legal question, is not a central focus of this *IMAGE* special collection, some of the essays do address it. In particular, Nicolle Lamerich's (2023) contribution on fan art shows that a significant portion of the fan community objects to their works being used without permission and possible remuneration. A class-action lawsuit has already been filed in the United States against Stability AI, Midjourney, and Deviant Art by three artists claiming that their work has been used to train the model which in turn can generate images similar to their art style (cf. WIGGERS 2023). This amounts, they claim, to a "21-century collage" technique and is not fair use. In response, a website titled "Stable Diffusion Frivolous" was created by "tech enthusiasts uninvolved in the case, and not lawyers, for the purpose of fighting misinformation" (STABLE DIFFUSION FRIVOLOUS n.d.: n.pag.). The site sets out to refute several points made by lawyers for the plaintiff, essentially claiming that what Stable Diffusion and the other models do is indeed fair use.

The legal case will depend on technical issues concerning the storage, modification, and use of the image-text pairs and on how well intellectual property lawyers and judges understand these technicalities. Other lawsuits concerning generative AI have already been filed, and more are sure to follow (cf. WIGGERS 2023). The legal issues may well be decided differently in different countries with different intellectual property regimes. While we do not know how these issues will be settled, it seems likely that these disputes will continue for some time (years?), and that there eventually will be two related but distinct resolutions: one legal and the other cultural. By that I mean that the law on AI image generation will be more or less settled and there may well be some system of remuneration for the human artists. That settlement, however, will not necessarily be the same as what our media culture in general comes to accept as 'fair use' in image generation. Such a divergence between the law and cultural practice has happened before. The situation over the last three decades of the practices of remix provides a good example. Based on various legal rulings, there is an elaborate set of rules about how older samples can be used in new works. This is especially the case for music sampling. However, most amateur remixers do not pay much attention to the rules. The web is full of remixes that may technically be illegal, but unless they reach a certain level of economic significance, they are largely ignored. Our media culture has come to a shared understanding of the legitimacy of the creative appropriation of earlier works, which is what Lawrence Lessig (2008) was arguing for in his book on remix.

The second question is: are these images creative artifacts in their own right? And if so, who is the creator? Several essays in the present collection, especially

those by Erwin Feyersinger, Lukas Kohmann, and Michael Pelzer (2023) and by Scorzin (2023), have important insights to contribute. On the other hand, there seems to be no consensus among the larger public on these questions. Two observations here:

1. A wide group outside of the traditional arts and media studies communities (artists, critics, scholars) feel compelled to express their views on this question. The technical community certainly does; computer scientists have become theorists of art.
2. In many cases, these views often begin from the assumption that the definitions of art and creativity, prior to AI, are themselves relatively unproblematic. The question is simply how these AI systems fit into those definitions.

The larger popular debate is connected to a question, then, of the cultural status of art in an era of digital media. The notion that art is the province of cultural and academic elites has been eroding for decades. The belief in a cultural hierarchy in which the fine arts and literature are superior to film, popular music, romance novels, and comics is largely, if not entirely, gone. In its place is the sense that all forms of creative expressions have equal status: the network is replacing the hierarchy as a cultural form (cf. BOLTER 2019). With its blogs, fan pages, streaming services, and social media sites, the Internet is perfectly suited to foster this networking of culture. At the same time, the vocabulary and the implicit values of that earlier hierarchical era have not disappeared. Our current media culture has adopted them. The term ‘artist’ has vastly expanded to include all sorts of creative practices that would not have been called art before the 1960s. And this trend makes it easier to imagine a further expansion to include these generative programs and their neural nets, which are themselves among the most impressive products of network thinking.

AI as a Tool

As several of the essays in the present collection suggest, there are other ways to regard these AI programs than as threats to or replacements for human artists or creators. One is to regard AI as a new tool in the hands of human agents. The authors of the Stable Diffusion Frivolous website pages take this position, referring to these programs as “AI Art tools”. They argue that there are earlier instances of new technological tools reconfiguring artistic practice,

anti-AI artists [fear] being replaced by artists who use AI tools in their workflows. Just like the fear was of manual artists being replaced by digital artists when tools like Photoshop emerged, and the fear of painters being replaced by photographers when the camera was developed (STABLE DIFFUSION FRIVOLOUS n.d.: n.pag.).

Writing in the *American Scientist* an article entitled *AI Is Blurring the Definition of Artist*, a top researcher in this field, Ahmed Elgammal (2018), supports this view: “[J]ust because machines can almost autonomously produce art, it doesn’t mean they will replace artists. It simply means that artists will have an additional creative tool at their disposal, one they could even collaborate with” (ELGAMMAL 2018: n.pag.).

Already we can see generative AI integrated into popular software tools such as Photoshop with plugins and experimental add-ons – a trend that will likely continue to create a much tighter connection between traditional digital and AI-supported art creation. The idea of AI as a tool for improving a human artist’s work is reminiscent of one line in the original AI debate from the 1950s to the 1980s, a period when artificial intelligence was still controversial even in the computer science community. Some computer scientists thought that instead of aiming for artificial intelligence the goal should be to develop interfaces and systems that would serve to amplify human intelligence. That was the implicit, and sometimes explicit, assumption behind the development of personal computing in this period: for example, Douglas Engelbart, one of the pioneers of the desktop interface, whose 1968 demonstration of his NLS system introduced a number of key elements of desktop computing, spoke of “augmenting human intellect” (ENGELBART n.d. [1962]: n.pag.).

Many advocates for AI in that original debate, such as John McCarthy and Marvin Minsky, believed that so-called ‘symbolic AI techniques’ would lead to intelligent systems that could function without human collaborators. However, researchers in machine learning today, whose neural nets are far more powerful than anything that the era of symbolic AI produced, seem to welcome the idea that AI systems would be used in collaborative relationships with human agents, rather than replacing humans altogether. Some of the essays in our collection explore this line, too. Feyersinger, Kohman, and Pelzer (2023: 134) argue that the “fuzzy” image generation of DALL-E 2 can serve as a form of externalized visual thinking for human creators. Scorzin observes that we may think of these systems as co-creative agents. Human-machine co-creativity has become a topic of interest in the computer science community, a common theme of papers in the ACM conference “Creativity and Cognition” and elsewhere. A recent anthology entitled *The Language of Creative AI* (VEAR/POLTRONIERI 2022: xi) “builds on [...] and extends the notion of embedded and cooperative creativity with intelligent software. It does so through a human-centered approach in which the AI is empowered to make the human experience more creative or join in/cooperate with the creative enterprise in real time”.

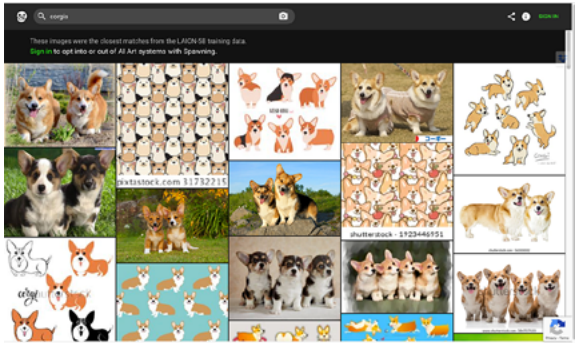
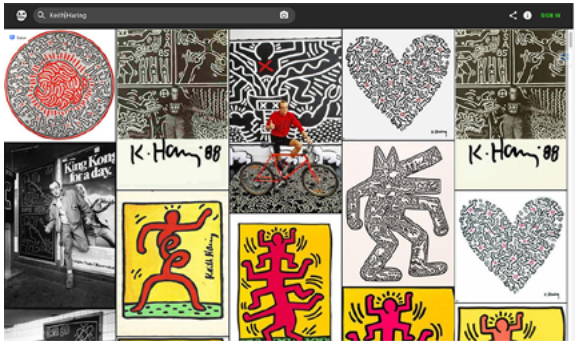
AI as a Medium

There is yet another perspective to consider. Instead of regarding the systems as agents or tools, they can be thought of as a new medium. In the case of text-to-image generators like DALL·E 2, not just the prompt itself but the whole process of creating the model and producing images would constitute the medium. There would be nothing particularly novel in imagining that the characteristics of the medium would themselves impose constraints on or facilitate the making of the art. But the degree to which the medium of AI would participate in the fashioning of the images is new and perhaps without parallel. We could say that the system of artist and AI constitute both maker and medium. I suggest that if the database, model, and algorithms behind systems like DALL·E 2 are constituents of a new medium, then that medium is rooted in the principle of remix or remediation for two reasons. First, the existing generative models depend on other media – above all painting, drawing, and photography. These are the media that constitute the imaginary of the current web that is scraped to generate the models. Second, these systems are constituted from text-image pairs, and the generated images are therefore the product of two heterogeneous media. The model itself is intermedial, a blend of text and images that is both at the same time.

AI generative imagery is remix, but there is an important difference from earlier forms. We can trace one strand of remix to hip-hop practices dating back decades. Then there is the somewhat younger video remix, which involves the editing and often complex layering of a series of video clips together with an underlying musical soundtrack. This practice became popular among amateurs in the 2000s because of the availability of affordable editing tools and inexpensive computers powerful enough to handle the tasks. Both audio and video remix require the step-by-step intervention of the human remixer, and this is obviously different from the process of image generation in DALL·E 2, Midjourney, and the like. True, the creation of the data model itself is a step-by-step process by a team of programmers and an anonymous crowd of image taggers. But for the human user providing a text prompt, the rhythm of interaction is redefined, and human intervention is reduced. Even at this stage, however, it is possible to use these systems for interactive refinement and skilled manipulation, as several of the present essays are remarking. Eventually, as noted above, we will likely see a workflow similar to that of a skilled user applying filters in Photoshop. In any case, this kind of AI image generation will always be remix because the systems begin with the visual samples and captions scraped from the Internet.

One of the largest such text-image databases is LAION (Large-scale Artificial Intelligence Open Network). This publicly available database was used to generate the models for Stable Diffusion, Imagen, and others. The March 2022 release contained more than 5 billion text image pairs. You can query this database from

the site haveibeenentrained.com, whose explicit purpose is to allow artists to see if their work is present in the database and request its removal. Typing in the name of the graffiti artist Keith Haring, for example, produces the eclectic results seen in figure 2 upper part. As we would expect, the database has scraped not only images of Haring's work, but other related images such as mimickings of his style. Searches on the site are not limited to names; one can search for other terms as well. We can use haveibeenentrained.com to view the kinds of images that underlay results in DALL·E 2. (Although OpenAI used its own database for the model behind DALL·E 2, the data must have been similar to the 5 billion pairs of LAION.) For example, the illustrative image in an OpenAI research paper (RAMESH et al. 2022) was generated from the phrase "a corgi playing a flame throwing trumpet". Here are the first few corgi examples and the first trumpet examples from the LAION database (cf. fig. 2 middle an lower part).



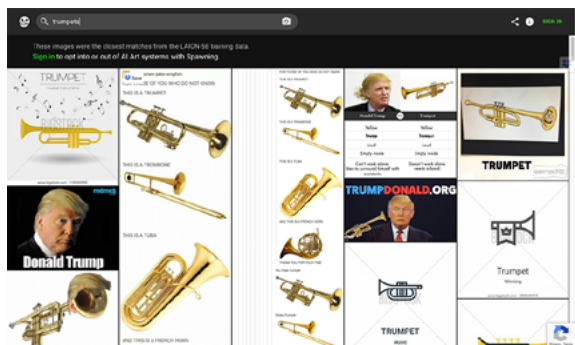


Figure 2: Screenshots from the LAION database for the queries “Keith Haring”, “corgis”, and “trumpets”

LAION and the haveibeen trained site are revelatory of the ontology of AI generated imagery. Viewing the initial images from the web makes it apparent that the process is one of remediation. Sophisticated algorithms create the models from these vast databases, and the models are tuned in various ways by human programmers. Nevertheless, without the original data the generated images would not be possible. In *Generative AI and the Collective Imaginary: The Technology-Guided Social Imagination in AI-Imaginesis*, Andreas Ervik (2023) argues that the AI generated images are becoming part of our collective imaginary. It is also important to remember that these images emerge from the prior collective imagery and then are added to it. We can call this process algorithmic remix or remediation. And when enthusiasts for AI generation claim that these systems make possible a new kind of art, their claim is similar to the familiar claim for audio and video remix as art. It is characteristic of new mediums to claim that their remediations constitute a significant new form of expression.

AI and Ekphrasis

Let’s return to the key feature of these new generative AI systems: the relationship of text to image. (The feature is addressed by almost every essays in our collection, but particularly by Feyersinger, Kohmann, and Pelzer and by Bajohr in his paper on what the latter calls “dumb” semantics, 2023a: 57) Text is crucial in both the encoding and the generation phases: The images that are scraped from the web all have captions, and the captions are encoded along with the images. This creates the encoding space called the prior that is used in the generation phase when the human user types in text that serves as a prompt. The user’s text

is a description of the image that is desired; the image cannot exist until the text is applied to the model.

The general popular reaction and the perspective of the makers of these systems seem to take the relationship for granted. Or rather, both users and makers seem most interested in a practical question: how to word the text so as to generate the desired image. This is clear from the OpenAI website, which includes a “prompt book”, a set of instructions about how to get various desired stylistic effects (cf. DALL·ERY GALL·ERY 2022). Here are samples OpenAI itself created and shows in its paper on the generative technique (cf. RAMESH et al. 2022, cf. fig. 3). Such images, emblematic of DALL·E 2, are interesting as cultural expressions. They are playful in an almost postmodern way. They have the quality of pastiche and a disinterest in stylistic coherence and are characterized by an absence of affect. We could almost imagine them as illustrations in Fredric Jameson’s *Post-modernism* (1991).

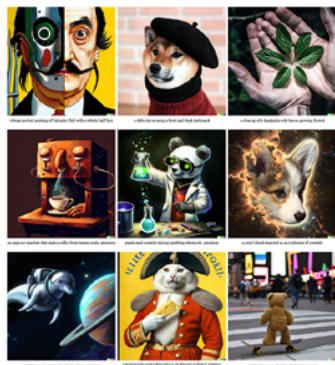


Figure 3: Various samples from OpenAI for DALL·E 2 prompts and their respective outcomes, taken from Ramesh et al 2022

I would argue that such text-image pairs constitute metapictures according to the definition by William J.T. Mitchell in *Picture Theory* (1994). Metapictures are pictures that are self-referential, and Mitchell distinguishes various classes, of which one is pictures that enter into a self-referential relationship with verbal text. This is the case for DALL·E 2 images. On the DALL·E 2 website, collections are displayed so that the images are visible and the generative text is obscured; the text only appears when the user mouses over. Mousing over reveals the text that stands behind and ontologically underneath the image. But what is the relationship here? The text is no longer exactly a caption. Does the text explain the image, justify the image? The indeterminacy of the text-image relationship is emphasized by the fact that multiple images are generated by the same prompt when repeated. DALL·E 2 shows you four, with the suggestion that many more are possible.

Mitchell devotes one of the essays in *Picture Theory* to the literary device of ekphrasis. The device dates back to antiquity, when poets would offer a vivid description of a visual scene or an art object. In many cases of ekphrasis, however, from antiquity to the present, the object being described does not exist. It is part of the literary fiction. The point of the ekphrastic description is to demonstrate the power of language to visualize with the implication that language can rival the visual arts at capturing reality. Ekphrasis has always been remediation in the sense of competition, seeking to show that the word can compete with painting at visual representation. The earliest example of ekphrasis usually cited is the description of the shield of Achilles in the *Iliad*. Let's test DALL-E 2 with this canonical example. We first see what kind of images there were in the database. To query LAION, we can type the phrase "the shield of Achilles as described by Homer in the Iliad" into the havebeentrained site. The result is a number of images as well as book covers; figure 4, upper part, shows a sample. If we then type the phrase "the shield of Achilles as described by Homer in the Iliad" into DALL-E 2, we get four results, all of which recall some formal aspects of the shield as Homer described it. Figure 4, lower part, is the most compelling.

Of particular interest is the appearance of text in the resulting image. (This also happened in the two cases above when we fed the title of essays from this collection into DALL-E 2, perhaps because the abstract vocabulary of the titles seemed to encourage the generative algorithm to produce abstract results.) Text appears, but not in the form of recognizable names or words. It seems as if the text has broken through to the surface of the image. As viewers or readers of this image, we might be tempted to try to make sense of it: Is this a blend of the names Homer, Achilles, and the *Iliad*? Or we might ask: Is the text here being used formally rather than symbolically? Is this what happens to language when it is absorbed into the neural layers of the model? It seems as if this imagetext is commenting upon, almost parodying the original prompt, again emphasizing the indeterminacy of the relationship between text and image. Are we witnessing the artificial or dumb semantics that Bajohr (2023a) discusses in his essay for this collection? In a separate paper, Bajohr has explored the text-image relationship in more detail and argued convincingly that these generative systems constitute a new kind of text-image relationship that he calls *operative ekphrasis* (cf. BAJOHR 2023b).

In his essay on ekphrasis, Mitchell speaks of the phenomenon he calls *ekphrastic hope*. He means the hope that a verbal description could succeed in bringing forth an image with perfect representational clarity – that the difference between word and image between the symbolic and the iconic, or perhaps in post-structuralist terms between the signifier and the signified, could be bridged. Mitchell also characterizes ekphrastic hope's opposite, ekphrastic fear. We can understand this as the fear that ekphrasis might in fact succeed. For if



Figure 4: Screenshot from the LAION database for the query “the shield of Achilles as described by Homer in the Iliad” and a DALL-E 2 creation prompted with the same text, generated in February 2023

ekphrasis fully succeeds, what happens to the status of the word and to verbal art and expression in general? Would the word be absorbed into the image and lose its identity? Perhaps we are witnessing the absorbing of the word into the image in the encoding processes of these AI models, in which the text captions are fed into the neural net and lose their semantic identity.

We could argue that Mitchell’s notion of ekphrastic hope is what the makers of these generative art systems are striving to realize when they cheerfully list all the uses of their systems for illustrating blogs and newsletters and for giving users the power to paint with words as never before. The *DALL-E 2 Prompt Book* is designed for this purpose, to empower users to tune their images, to paint with

words. In this sense, the prompt book is emblematic of the optimism of this AI moment in general; and then in turn, what Mitchell calls ekphrastic fear could be the backlash by those who resist AI image generation because it suggests to them that the larger project of generative AI could succeed and bring with it unforeseen and negative consequences for human creativity. These consequences could go beyond the economic loss to artists and designers through appropriation of their intellectual property and through automation of their skills and expertise. The ultimate threat would be the loss of the arts and crafts as autonomous human activities. The future almost certainly lies somewhere between the extremes of ekphrastic hope and fear. Most of the essays in this collection could be characterized as cautiously optimistic about the potential of AI generated imagery. They do not endorse the future that Open AI's CEO Sam Altman imagines for a world of AGI (Artificial Generalized Intelligence) (cf. LAWRENCE 2023), but they are still ready to engage with the theoretical and practical opportunities that AI affords in the realms of visual representation and art.

Bibliography

- BAJOHR, HANNES: Dumb Meaning: Machine Learning and Artificial Semantics. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of IMAGE: The Interdisciplinary Journal of Image Sciences*, 37(1), 2023, pp. 58-70
- BAJOHR, HANNES: *Operative Ekphrasis*, 2023. Manuscript provided by the author
- BOLTER, JAY DAVID: *The Digital Plenitude: The Decline of Elite Culture and the Rise of New Media*. Cambridge, MA [MIT Press] 2019
- DALL-ERY GALL-ERY (ed.): The DALL-E 2 Prompt Book. In: *Dall-ery gallery: Ressources for Creative DALL-E Users*. July 14, 2022. <https://dallery.gallery/the-dalle-2-prompt-book/> [accessed March 9, 2023]
- ELGAMMAL, AHMED: AI Is Blurring the Definition of Artist. In: *American Scientist*. December 6, 2018. <https://www.americanscientist.org/article/AI-is-blurring-the-definition-of-artist> [accessed March 9, 2023]
- ENGELBART, DOUG: Augmenting Human Intellect: A Conceptual Framework. SRI Summary Report AFOSR-3223. October 1962. In: *Doug Engelbart Institute: Strategies for a More Brilliant World*. No date given. <https://www.dougenelbart.org/pubs/augment-3906.html> [accessed March 9, 2023]
- ERVIK, ANDREAS: Generative AI and the Collective Imaginary: The Technology-Guided Social Imagination in AI-Imaginesis. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of IMAGE: The Interdisciplinary Journal of Image Sciences*, 37(1), 2023, pp. 42-57
- FEYERSINGER, ERWIN; LUKAS KOHMANN; MICHAEL PELZER: Fuzzy Ingenuity: Creative Potentials and Mechanics of Fuzziness in Processes of Image Creation

- with Text-to-Image Generators. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of IMAGE: The Interdisciplinary Journal of Image Sciences*, 37(1), 2023, pp. 135-149
- JAMESON, FREDRIC: *Postmodernism*. Durham [Duke University Press] 1991
- LAMERICHS, NICOLLE: Generative AI and the Next Stage of Fan Art. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of IMAGE: The Interdisciplinary Journal of Image Sciences*, 37(1), 2023, pp. 150-164
- LAWRENCEC: Sam Altman: "Planning for AGI and Beyond". In: *LESSWRONG Blog*. February 24, 2023. <https://www.lesswrong.com/posts/zRn6aQyD8uhAN7qCc/sam-altman-planning-for-agi-and-beyond> [accessed March 9, 2023]
- LESSIG, LAWRENCE: *Remix: Making Art and Commerce Thrive in the Hybrid Economy*. London [Bloomsbury] 2008
- MITCHELL, WILLIAM J.T.: *Picture Theory*. Chicago [University of Chicago Press] 1994
- RAMESH, ADITYA; PRAFULLA DHARIWAL; ALEX NICHOL; CASEY CHU; MARK CHEN: Hierarchical Text-Conditional Image Generation with CLIP Latents. *arXiv:2204.06125*. April 13, 2022. <https://arxiv.org/abs/2204.06125> [accessed March 9, 2023]
- SCORZIN, PAMELA C.: AI Body Images and the Meta-Human: On the Rise of AI-generated Avatars for Mixed Realities and the Metaverse. In: *Generative Imagery: Towards a 'New Paradigm' of Machine Learning-Based Image Production, special-themed issue of IMAGE: The Interdisciplinary Journal of Image Sciences*, 37(1), 2023, pp. 179-194
- STABLE DIFFUSION FRIVOLOUS: Stable Diffusion Frivolous: Because Frivolous Lawsuits Based on Ignorance Deserve a Response. In: *stablediffusionfrivolous.com*. No date given. www.stablediffusionfrivolous.com [accessed March 9, 2023]
- WIGGERS, KYLE: The Current Legal Cases against Generative AI are Just the Beginning. In: *TechCrunch*. January 27, 2023. <https://techcrunch.com/2023/01/27/the-current-legal-cases-against-generative-ai-are-just-the-beginning/> [accessed March 9, 2023]