1. The Latent Objective World

Photography and the Real after Generative AI

Amanda Wasielewski

When OpenAI presented DALL-E 2 on 6 April 2022, one of the sample images they used across social media and in the press was "A photo of an astronaut riding a horse." DALL-E was one of the first visual generative artificial intelligence (AI) platforms to make generating AI-produced images easy and accessible to the general public. The company's release of images such as this was designed to show the range of subject matter and style combinations that could be achieved. And the technology did not disappoint, as evidenced by the ensuing hype that surrounded it. In many of the sample images produced, two incongruous elements—in this case, the astronaut and the horse—are shown pictured together in a seamless way. This is made possible by combining a multimodal text/image model, which builds semantic links between these two "modes" of data using an artificial neural network, and a diffusion model, which is trained to produce new images from noise through a process of adding noise to training images.[1] Together, these models are a powerful engine for image creation.

The launch of DALL-E was not the first time that space exploration had been used to signal the start of an innovative new media platform.[2] The hook in this case, however, is not merely the presence of the astronaut doing something groundbreaking but the strange and fantastical juxtaposition of that astronaut doing something *impossible*, i.e., riding a horse in space. The fact that

1 This work is supported by a grant from the Artists + Machine Intelligence (AMI) research program at Google. Alec Radford et al., "Learning Transferable Visual Models From Natural Language Supervision" (arXiv, February 26, 2021), https://doi.org/10.48550/arXiv.2103.00020; Jonathan Ho, Ajay Jain, and Pieter Abbeel, "Denoising Diffusion Probabilistic Models" (arXiv, December 16, 2020).

2 MTV also famously used the image of the astronaut at its inception.

it is rendered like a photograph aligns this impossible scenario with photography's historic ties to the real.[3] Thus, a new frontier is opened up by this image—but what kind of frontier? The juxtaposition is, on one hand, reminiscent of the Comte de Lautréamont's famous "chance encounter of a sewing machine and an umbrella on an operating table" that was popular among the Surrealists of the 1920s.[4] On the other hand, it is also reminiscent of Jean Baudrillard's definition of simulation as "the generation by models of a real without origin or reality" and his attendant concept of the hyperreal (the etymological cousin of the surreal).[5]

This chapter explores the fraught connection to the real presented by AI-generated images rendered in a photographic style—what I call, simply, AI-generated photographs. Framed by Baudrillard's theory of objects, I argue that AI-generated photographs are defined by the elision of the photographic medium for the task of object recognition. This elision is complicated by photography's virtuality, which is defined by art historian David Summers as a surface with the appearance of "forms in real space."[6] Without the real space or origin presumed to underpin it, however, "The Virtual is no longer the potentially real, as it once was," according to Baudrillard.[7] AI-generated photographs embody precisely this loss of potential for the virtual in photography, i.e., the loss of real space.

Objects and Images on the Operating Table

Like the sewing machine and the umbrella on the operating table, the meeting of the astronaut and the horse is quite literally a chance encounter. It is the product of complex statistical weights in the AI model. Their coming together

3 Amanda Wasielewski, "Unnatural Images: On AI-Generated Photography," *Critical Inquiry*, Autumn 2024.

4 Comte de Lautréamont, *Les Chants de Maldoror* (Paris: Éditions de la sirène, 1920), 323; Shane McCorristine, "Lautréamont and the Haunting of Surrealism," in *Writing in Context: French Literature, Theory and the Avant Gardes*, ed. Tiina Arppe, Timo Kaitaro, and Kai Mikkonen (Helsinki: Helskinki Collegium for Advanced Studies, 2009), 31–49.

5 Jean Baudrillard, *Simulacra and Simulation*, trans. Sheila Faria Glaser (Ann Arbor: University of Michigan Press, 2014), 1.

6 David Summers, *Real Spaces: World Art History and the Rise of Western Modernism* (London: Phaidon, 2003), 431.

7 Jean Baudrillard, *Impossible Exchange*, trans. Chris Turner (London: Verso, 2001), 19.

is not only determined, in the most immediate sense, by the text prompt used but also by a statistical sampling of elements—of objects—in the hundreds of millions of images that have been used to train contemporary AI systems. Visual generative AI tools like DALL-E are able to create this kind of image because they are premised on an understanding of images as content containers rather than representative media. In other words, digital images—particularly photographic images—are used as the eyes of the computer. The field of computer vision concerns itself with an object world pictured through digital photographic images. Thus, an image is not a whole. It is primarily a collection of objects to be perceived—to be "recognized."

Once confined largely to government and military surveillance systems, computer vision technologies are an increasingly ubiquitous part of our daily life. While they are still used as a means to control, police, and wage war, they are also increasingly present in more quotidian applications. For example, we open our phones with facial recognition, we organize photographs with image categorization techniques, and we are asked to prove we are human by identifying bicycles or fire hydrants in reCAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) images provided from Google Street View.

Starting a Yahoo! Mail account circa 2005, a user would be greeted with a verification system displaying an image of some distorted letters with lines running through them.[8] They would have been asked to type in that string of garbled text into a textbox to create their account. Developed by computer scientist Luis von Ahn and collaborators a few years prior, the system known as CAPTCHA (and an updated version called reCAPTCHA) was designed to mitigate the onslaught of automated spam accounts that had grown into an enormous problem for internet users at that time.[9] To do so, Von Ahn and others had to devise ways to separate human users from automated systems, and one of the things that still set humans apart was their visual perceptual abilities. As CAPTCHA was taken up around the internet, Von Ahn realized that his invention required a certain amount of labor from users. Hoping to "do something useful" with this time, Von Ahn implemented text fragments from digitized

8 Luis Von Ahn et al., "CAPTCHA: Using Hard AI Problems for Security," in *Advances in Cryptology — EUROCRYPT 2003*, ed. Eli Biham, vol. 2656 (Berlin/Heidelberg: Springer, 2003), 294–311.

9 Finn Brunton, *Spam: A Shadow History of the Internet* (Cambridge, MA: MIT Press, 2013), 113.

books and periodicals, effectively asking users to aid in the training of OCR (optical character recognition). And so, ironically, CAPTCHA was involved in the process of training computers to "see," which would ultimately void its usefulness. In a very real sense, CAPTCHA dug its own grave.

After Google acquired reCAPTCHA in 2009, they continued to enlist unwitting users to perform computer vision training tasks in the name of internet security.[10] Since 2012, Google has used its vast repository of Street View photographs as part of the reCAPTCHA system, beginning with house numbers.[11] After OCR, i.e., text recognition, was effectively solved, Google moved on to objects pictured in Street View images. Internet users today are thus still asked to choose which squares from a street view image contain bicycles, traffic lights, or crosswalks.

The rise of deep learning in computer vision, aided in part by twenty years of CAPTCHA input data from users, has made the idea behind image-based security measures obsolete.[12] Images of all kinds are now, in fact, extremely machine readable. Computers can accurately decipher garbled text or identify unclear objects in photographs. What this means is that the camera is often conceptually positioned and implemented as the artificial eye of the computer. Because of this, one might naturally leap to the assumption that machines can "see." However, the process by which camera/computer and eye/human operate are very different. So, computer "vision" or "sight" is simply a metaphor.[13] Nevertheless, as the camera and eye are conflated, both are understood as mere mechanical tools, passively able to take in information about the world. They

10 Leena Rao, "Google Acquires reCaptcha To Power Scanning For Google Books And Google News," *TechCrunch* (blog), September 16, 2009, https://techcrunch.com/2009/09/16/google-acquires-recaptcha-to-power-scanning-for-google-books-and-google-news/.

11 Sarah Perez, "Google Now Using ReCAPTCHA To Decode Street View Addresses," *TechCrunch* (blog), March 29, 2012, https://techcrunch.com/2012/03/29/google-now-using-recaptcha-to-decode-street-view-addresses/.

12 Although consumer-focused tools are often programmed to reject user requests to decipher CAPTCHAs, machine learning expert Denis Shiryaev published his successfully attempt to trick Bing Chat into decoding a CAPTCHA in October 2023. See Benj Edwards, "Dead Grandma Locket Request Tricks Bing Chat's AI into Solving Security Puzzle," *Ars Technica*, October 2, 2023, https://arstechnica.com/information-technology/2023/10/sob-story-about-dead-grandma-tricks-microsoft-ai-into-solving-captcha/.

13 Wasielewski, "DALL-E in Flatland: Illusion, Space, and AI-Generated Images" *Media Theory* [forthcoming]

are seen as neutral in some way, or pure in how they receive impressions from the world.

In "Surrealism and Painting" (1928), André Breton begins with the line, "The eye exists in a savage state."[14] For Breton, vision is a form of bodily automatism, disconnected from the rational thought that pollutes it. As art historian Rosalind Krauss explains, "Besides being untainted by reason, vision's primacy results from the way its objects are present to it, through an immediacy and transparency that compels belief."[15] Vision is, thus, reduced not just to an automatic procedure but to one with a singular purpose: object recognition. Vision is a means by which to access objective reality, or the ontology of objects.

Jean Baudrillard's writing addresses precisely the nexus of relations that AI and object recognition raise, between images, objects, and the real. For Baudrillard, photography is a process of objectification, and AI image generation ultimately reinforces the view of photography as objectifier.[16] Vis-à-vis automation, photography is understood to reveal the latent objective world. For the Surrealists, it was also a manner by which to tap into the symbolic world of the mind that was often situated in the form of the object. AI-generated photographs, in turn, reveal the hidden facets of the data used to produce them—the object world of the dataset.

It is not often acknowledged but Baudrillard was an image-maker—a photographer—in addition to being a writer/theorist, though he only addresses photography's history and theory directly in a few of his essays. Baudrillard saw photography as, essentially, a process of objectification.[17] He writes, "To take photographs is not to take the world for an object, but to make it an object, to exhume its otherness buried beneath its alleged reality…"[18] Following this, object recognition is not the primary site where the photograph becomes a collection of objects. Instead, it provides a further layer of objectification on those things already long-objectified in the image. Said differently, the photographic images used to train computer vision systems have already transformed the world into a set of objects, and, so, the layer of data produced to

14 André Breton, *Surrealism and Painting*, trans. Simon Watson Taylor (New York: Harper & Row, 1972), 1.

15 Rosalind Krauss, "The Photographic Conditions of Surrealism," *October* 19 (1981), 10.

16 Nicholas Zurbrugg and Jean Baudrillard, "Fractal Theory," in *Baudrillard Live: Selected Interviews*, ed. Mike Gane, trans. Nicholas Zurbrugg (London: Routledge, 1993), 168.

17 Ibid.

18 Jean Baudrillard, "For Illusion Isn't the Opposite of Reality…," in *Jean Baudrillard: Photographies, 1985–1998*, ed. Peter Weibel (Graz: Hatje Cantz Publishers, 1999), 132.

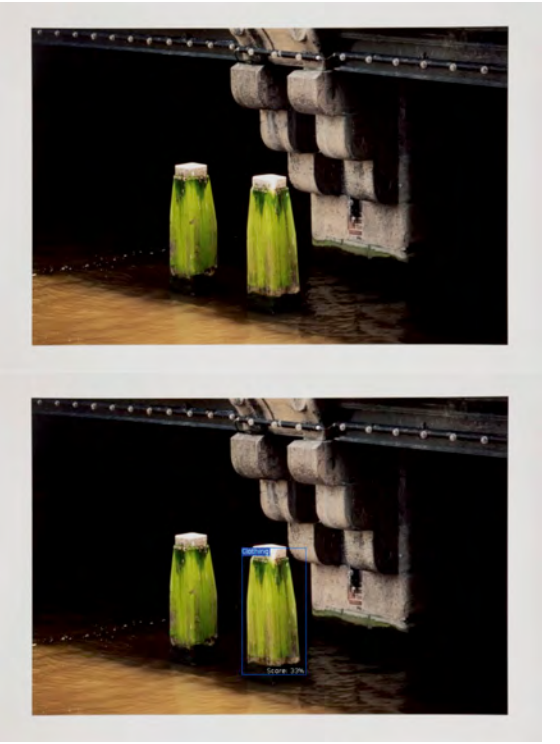
help identify those objects with words and automate their categorization represents a further objectification.

In Baudrillard's photography, he does precisely what he describes in the quote above: he harnesses photography's objectifying powers. The images he creates have an uncanny quality, where a skillfully cropped detail takes on a strange otherness despite its ordinariness. His photos often show objects in a state of decay or weathering, either well-used or long-abandoned. Objects pictured in these photos crack or rust or peal in a direct affront to their supposed usefulness. In *The System of Objects*, Baudrillard reflects at length on the concept of functionality in relation to objects. He writes, "The materiality of objects no longer directly confronts the materiality of needs, these two inconsistent primary and antagonistic systems having been suppressed by the insertion between them of the new, abstract system of manipulable signs—by the insertion, in a word, of *functionality*" [italics original].[19] In a way, then, Baudrillard uses already unfunctional objects and removes a second layer of functionality. He takes away the symbolic layer of function for the image, i.e., the layer of language, where one might assign a name or a type (and thus, a function) to each of the objects. Following this, AI-enabled object recognition attempts to put language back into the image.

In Baudrillard's photograph *Amsterdam* (1992) (Fig. 1.1), for example, a pair of bright green algae-covered piers poke out of a river in front of a bridge support. The close cropping of the photo abstracts these elements from any reading of the image that points to their highly functional purpose, namely the protection of the bridge from errant boat collisions. They become part of a flattened collection of purposefully-formed objects with no evident purpose. Before beginning to write about this photo, I puzzled over what these objects in the water are actually called. I could not find any other term except "pier," which felt a bit too generic. Never mind. I need not have bothered anyway, since the identification of the object seems so decidedly unimportant to the image. The objects depicted have, in other words, somehow been stripped of their recognizability, despite my natural inclination to describe them in some categorical terms.

19 Jean Baudrillard, *The System of Objects* (London: Verso, 2020), 68.

1.1 Jean Baudrillard, Amsterdam, 1992, 60 x 90 cm (paper size); 50 x 75 cm (image size), giclée print on pure cotton paper.



Source: Château Shatto gallery. © Jean Baudrillard. Reproduced with the permission of Marine Baudrillard.

In addressing this quandary, I decided to turn to AI object recognition software. Aided by several free online tools, which all appear to use Google's Cloud Vision API, I found that these objects were identified as "clothing" by the AI computer vision algorithm. While there may be better object recognition tools available that could identify these objects with more accuracy, the exercise seems to neatly demonstrate the point above, namely, that these photographs defy both the function of the objects depicted and an additional layer of functionality reserved for the image itself.

When the filmmaker Harun Farocki coined the term "operational images" in his film series *Eye/Machine* in 2001, computer vision and object recognition were still rudimentary in their capabilities.[20] Farocki defined operational images as images that "do not represent an object, but rather are part of an operation."[21] Thanks to contemporary deep learning and neural networks trained to do object recognition, any and all images today could be deemed "operational." This means that all images now have the potential to be used as part of operations, i.e., they are all machine readable and interpretable. Farocki discusses the use of operational images as those that are, in a way, free from authorship or intention—free even from being "custom-made to mean something to us."[22]

In both his film work and media theory, Farocki was influenced by the theorist Vilém Flusser, who used the term "technical images" to describe electronic images in the mid-twentieth century.[23] He describes these images as abstractions from texts, that are "not surfaces but mosaics assembled from particles."[24] Farocki similarly describes operational images (implying *automated* images) as those in which "each movement is broken down into fragments, and each fragment of the movement is performed with equal dedication."[25] Flusser's work and its description of the fractured nature of the technical image can be usefully applied to AI-generated photographs, which are likewise piecemeal or particulate compositions that have been abstracted from large datasets and text-based prompts.

As I have discussed previously, the underlying operation of computer vision and object recognition is, on a technical level, a fractured process, which reflects the make-up of most digital images (and thus their machinic processing) as a grid of pixel values.[26] These fractured processes are not done in the ser-

20 Volker Pantenburg, "Working Images: Harun Farocki and the Operational Image," in *Image Operations*, eds. Jens Eder and Charlotte Klonk (Manchester: Manchester University Press, 2016), 49–62.

21 Harun Farocki, "Phantom Images," trans. Brian Poole, *Public*, January 1, 2004, 17, https://public.journals.yorku.ca/index.php/public/article/view/30354.

22 Ibid., 18.

23 Christa Blümlinger, "An Archaeologist of the Present," *E-Flux Journal*, no. 59 (November 2014), https://www.e-flux.com/journal/59/61092/an-archeologist-of-the-present/.

24 Vilém Flusser, *Into the Universe of Technical Images*, trans. Nancy Ann Roth (Minneapolis: University of Minnesota Press, 2011), 6.

25 Farocki, "Phantom Images," 17.

26 Amanda Wasielewski, "Authenticity and the Poor Image in the Age of Deep Learning," *Photographies* 16, no. 2 (May 4, 2023): 191–210.

vice of connecting or constructing an understanding of the whole but, rather, in *segmenting*. Segmentation is a term used in computer vision to indicate the division of the image into neatly delineated and identified objects. Each object is an individual in an empty two-dimensional field, independent from that field. It is discrete. It is not part of any relation, except a binary ontological one, i.e. whether the object is present in the image or not. For example, an automated security camera using computer vision software is not designed to capture a scene per se but, rather, to capture *objects* as they appear/wander into its field of vision. These objects pass under the security camera's machinic eye, individually recognizable as a type. The common way that segmentation is represented in image data for human consumption is as a bounding field that isolates and labels the object, separating it from the surroundings. Even the specificity of individual objects is thus reduced to abstraction.

Given this, what is at stake in object recognition? Say, for example, a *person* is the object that will be recognized in an image. This person is an individual and will thus will be identified and labeled in some way that names them. It may be their actual name or it could be "person" or another larger category. This label might be prejudiced, unfair, simplistic or erroneous, but it will nevertheless point out the pattern of pixels that represent the appearance of individuality in an image and match it to others like it with a high level of precision.[27] The label is the final gloss on what is otherwise simply a matching exercise. In other words, there is a connection to the real in such exercises (even if that real is mischaracterized) but each training image is still a photographic representation that is distanced from its real-world referent by its representational transposition through camera and lens. Computer vision, in other words, is not the same as a machinic eye. It is the identification of patterns in photographic representations. This creates the illusion that such systems *understand* objects—that they understand the world in general—but they do not.

What we call artificial intelligence today is not equivalent to artificial *general* intelligence (AGI), the powerful thinking machines that are familiar from science fiction. Despite the Silicon Valley rhetoric around AGI, there are currently no machines that think, plan, and adapt the way that biological creatures do, despite the rapid development and increasing sophistication of deep learning systems. Baudrillard wrote about artificial intelligence (at least, in this hy-

27 Kate Crawford and Trevor Paglen, "Excavating AI: The Politics of Images in Machine Learning Training Sets," 2019, https://www.excavating.ai/.

pothetical sense) long before the present developments, but his statements resonate with the situation today:

> Let us say, before it is too late, that artificial intelligence is incompatible with thought for the simple reason that thought is not an operation, that it is not exchangeable for anything whatever, and, most particularly, not for the objectivity of an operational calculation of the input-output type.[28]

It is a remarkably prescient statement, given the current debates around whether computer can be trained to truly understand the meaning of words or images.[29] According to Baudrillard, it is an impossibility that computers can think or understand for the simple reason that *we* are not computers. Thought and computation are not equivalent. Computers may be able to imitate or simulate "our psychological and social mechanisms," he argues, but they are performing nothing more than a kind of "plagiarism."[30] Indeed, he points to the same manner of conflation between computer and human being that haunts rhetoric on the camera and the eye.

In order to explore this concept of imitation further, I return once again to Baudrillard's *Amsterdam* photograph. This time, rather than try to categorize it, I plagiarize it. In other words, I use text-to-image AI models to create something of its type. The task, paradoxically, is to describe an object that defies description. I used DALL-E 2 (with outpainting) and Midjourney (v.5.2), a competitor of DALL-E, for the exercise since DALL-E 3, at time of writing, does not produce images that look sufficiently photographic.[31] A selection of two of the resulting images, from around fifty images generated (Figs. 1.2 and 1.3) resonate with Baudrillard's photograph. All three images—the two AI-generated ones and Baudrillard's photograph—are closely cropped and relatively abstract. All three contain the same shade of algae green, though Baudrillard's photograph is more visually striking in its contrast and composition.

---

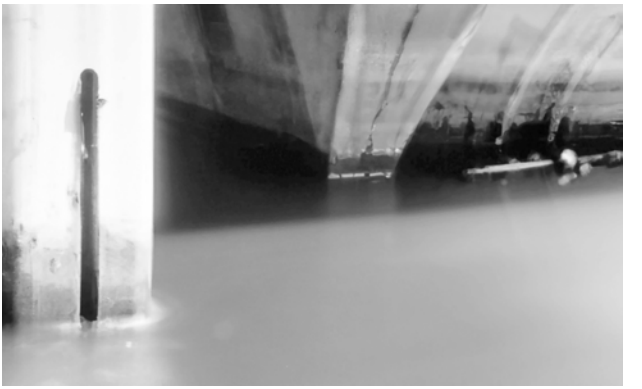
28 Jean Baudrillard, "Beyond Artificial Intelligence: Radicality of Thought," in *Impossible Exchange*, trans. Chris Turner (London: Verso, 2001), 146.

29 See, for example, the "stochastic parrot" debates: Emily M. Bender et al., "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event Canada: ACM, 2021), 610–623.

30 Baudrillard, "Beyond Artificial Intelligence," 148.

31 This change in DALL-E was notable from version 2 to 3

1.2 DALL-E-2 image created with the prompt "photograph 35mm photorealistic green pillar sticking out of brown water under a bridge green abstract next to stone bridge" (June 2023).



1.3 Midjourney image created with the prompt "photograph 35mm green pillar sticking out of brown water under a bridge green abstract next to stone bridge –ar 16:9 –style raw" (December 2023).

In the end, what is produced does not actually reverse the object recognition task, since there is more to these images than the text itself describes. We do not see two piers in a blank field but rather in a field populated by chance encounters, i.e. things statistically weighted to be appropriate to this image prompt. Applying the object recognition software once again to these AI-generated photographs, I find that Figure 1.2 (DALL-E 2) is labeled "window" and

Figure 1.3 (Midjourney) "fountain." The latter is, perhaps, the more accurate of the three. However, the exercise still demonstrates the extent to which these AI-generated photographs seem to defy functionality in a similar way that Baudrillard's photographs do.

Imagining the Photographic Form

The value of the image for object recognition operations in computer vision, then, is not in its overall meaning or composition—not in a thought process that comprehends the image per se—but rather merely in its contents/what is depicted. By contrast, Generative AI takes the segmented or fragmented nature of training images, i.e., those images that have been used for object recognition exercises, and compiles images from the places in the latent space of the model where image fragments associated with those words are located. In other words, the AI model is no longer trying to successfully identify a bicycle pictured in a photograph but, rather, is trying to create a new bicycle based on the learnings from the training data. Object recognition categorizes the individual to a general label, whereas generative AI produces an individual from a general category. Thus AI-generated photographs do not produce general objects (as a category) but rather produce a *specific* object (an individual).

Midjourney is accessible through a Discord server, where users create images by typing the command "/imagine" along with the text prompt describing what they want the image to contain. While scholars may debate the creative capacity of AI, it is clear that these tools can be used to "create" or "imagine" in the most basic etymological sense of producing images.[32] But what kind of images can they imagine? A typical text prompt consists of words that describe the content or objects pictured in an image and, perhaps but not necessarily, words that describe the style in which those objects are depicted. The visual field is, thus, conceived of as a plane on which to fill in things. It is a kind of stage to be choreographed or arranged.

32 See, among others, Joanna Zylinska, *AI Art: Machine Visions and Warped Dreams* (London: Open Humanities Press, 2020); Bojana Romic and Bo Reimer, eds., "Artificial Creativity," *Transformations Journal*, no. 36 (2022); Hye-Kyung Lee, "Rethinking Creativity: Creative Industries, AI and Everyday Creativity," *Media, Culture & Society*, vol. 44, no. 3 (April 1, 2022): 601–12; Ziv Epstein, Aaron Hertzmann, and the investigators of human creativity, "Art and the Science of Generative AI," *Science* 380, no. 6650 (June 16, 2023): 1110–11.

However, unlike a painting that is created from a blank canvas, everything that appears in an AI-generated image has *not* been decided or specified by the prompt writer. In this way, AI-generated imagery is far more like photography. As a test, I input the prompt "an umbrella and a sewing machine on an operating room table –ar 16:9 –style raw" into Midjourney. This prompt is, of course, an attempt to "imagine" the evocative line from the *Les Chants de Maldoror.* The operating table is sometimes translated as dissection table, but neither the words "operating table", "dissection table" or "operating room table" came through with an image that very strongly reflected this simple description. Instead, in the dozens of images I created, the sewing machine was always placed on what looks more like a normal sewing table rather than a medical bench. The training association between the sewing machine and the purpose-built sewing table is too strongly weighted, it seems. Nevertheless, the objects that I requested appear in the images produced.

1.4 Midjourney image created with the prompt: "an umbrella and a sewing machine on an operating room table –ar 16:9 –style raw" (November 2023).

But they are not all that appears. Looking closer at one of the resultant images from this prompt (Fig. 1.4), one can see that there are some objects pictured that were not specified in the text, as there always are in AI-generated images. One of these objects is a spotlight lamp, shown hanging over the table in a somewhat gravity-defying manner. The other object is difficult to describe. It looks like some kind of surgical room equipment—a chair? a table? a stool?—covered by a thick fabric with piles folded on top. The periwinkle blue

fabric and its thick, shiny texture evokes the material worn by surgeons in the operating room or the kind of protective gear a medical professional or their equipment might have. On top of this strange object is what looks to be a pan for surgical implements. As this example demonstrates, AI-generated images tend to contain two broad categories of anomalous objects: those we might expect to be there and those we struggle to identify. For example, in a prompt requesting a street scene, we may expect that cars and people and other elements of the street will be generated in the image even if not specified in the prompt. These are things that are typically part of a street. However, there may also be things we cannot recognize or identify—hybrids or blobs that are hard to decipher.[33]

Like photography, generative AI captures the incidental details of what "is there"—the accidental objects that happen to be in within its frame. For photography, this means that, when one presses the button or closes the shutter, one cannot always control *all* of the things that appear in the image. Photographers may try to curate the subject of a photograph, arranging or selecting certain elements they see, but there are always ambient details that are unplanned or, even, unwanted. For AI-generated imagery, capturing what "is there" means capturing the image data that is associated with certain words in the latent space of the AI model, i.e., the matrix of associations built up when the model was trained on millions of images and text pairings.

The details that appear in the AI-generated image are thus phantom image parts and truncated textures or appearances of certain pixel combinations that are somehow proximate to the words of the prompt in the model. In the case of the umbrella/sewing machine image, this strange surgical equipment/fabric combination has wandered into the frame, so to speak. Baudrillard writes, "The robot is a symbolic microcosm of both man and the world, which is to say that it simultaneously replaces both man and the world, synthesizing absolute functionality and absolute anthropomorphism."[34] The AI-generated image is a kind of snapshot, then, not directly representing the world but instead representing a *model* of the world, which is itself represented by photographs.

---

33 There is also a third, though less likely, category of objects that may appear: those that we do not expect to be there but recognize as known objects. For example, it is possible that an elephant would appear in the aforementioned street scene. This is highly unlikely, however, because most of the *training* images of street scenes would not picture an elephant.

34 Baudrillard, *The System of Objects*, 129.

Baudrillard's concept of the real, as discussed in "The Precession of Simulacra" (1981) is illustrative here. He opens this famous essay with a discussion of the Borgesian map that is so large it becomes the territory, claiming that this classic fable about representation is "unusable" and that there is no longer any difference to draw between the two, i.e., between the real and representation. The "map precedes the territory," he argues.[35] This is exactly the kind of representational relation that occurs in visual generative AI. These AI models are essentially representations, a kind of map of the world. This map is not a direct representation of the world, however, but rather a collation of image-based representations. The underlying reality behind the model is thus difficult to trace. Despite this complex dimensionality, however, the model is able to *generate* a real—or a hyperreal—in the form of singular images. Thus, the map *produces* the territory for AI-generated imagery. Baudrillard writes, "The real is produced from miniaturized cells, matrices, and memory banks, models of control—and it can be reproduced an indefinite number of times from these… It is no longer anything but operational."[36] So we are dealing not only with operational images but a kind of operational reality, where simulations beget simulations.

There is a common misconception today that AI-generated images are a collage or mashing together of existing imagery. Collage and remix require fragmentation and, even, repetition, but are essentially heterogenous mixtures that still maintain the integrity of their parts. AI-generated images, on the other hand, are the products of a chemical reaction where the parts that go into making it are fundamentally transformed in the process. The outcome is no longer merely a mixture or a mashing together of parts but something new.

This does not make AI-generated images less derivative, but it does represent a fundamental change of state. Baudrillard writes that, for simulation, "everything is already dead and resurrected in advance."[37] The process of objectification means that there is no underlying reality to reference, no real to compile or collage. So, despite their origins in individual datapoints, AI models are not creating new images based on a mashing or a mixing. There is no reference, no "living" reality behind these images but rather another set of objects. In other words, they are pure simulacrum.

35 Baudrillard, *Simulacra and Simulation*, 1.

36 Ibid., 2.

37 Ibid., 6.

Photography has often evaded claims to realism based on technicalities. It could not be called "realistic" because it lacked color. Or because it is still. Because it is silent. Because it is cropped. Because it is a lens-based reflection, distortion or transposition.[38] Baudrillard writes that these factors mean that "the photograph is the purest and most artificial image."[39] True in theory, perhaps, but these faults or failings never did much to quash photography's popular association with the real.

The rise of Photoshop and digital photography in the 1990s, however, seemed to eliminate the need for such technicalities. Instead, the lack of realism in the photographic image could be easily explained away by the ubiquity and seamlessness of image manipulation, of "photoshopping." Given the recent advances in generative AI technologies and platforms, any need for such technicalities or claims of manipulation fall away entirely. Photography cannot possibly lay claim to the real anymore. And yet, it inevitable and inextricably does. The aspects of form that seem so essential to photography are, in fact, not. Adjacent to these arguments is the idea that photography was not the paradigm-shifter it is sometimes claimed to be.[40] The paradox of contemporary photography is that it cannot escape its history. So it does not matter how distant the contemporary products of AI are from camera and lens-based photographs, they nevertheless continue to evoke the real, even when there is little to none to reference. From *The Pencil of Nature* to camera phone documentation, the photographic subject exists in a state of expectation or presupposition for the real.[41]

Surrealist photographers used this facet of the medium to evoke the uncanny. No sophisticated techniques were necessary: simple cropping or multiple exposures were often enough to do the job. The work of Jacques-Andrés Boiffard, for example, captures body parts in unexpected ways through cropping and camera angles. Realism and the sense that something is off coexist in photographs such as *Orteils et doigts croisés* (1929), which depicts toes and fingers interlaced together in an unnatural-looking way. This image is reminiscent of

38 Sabine T. Kriebel, "Introduction," in *Photography Theory*, ed. James Elkins (New York: Routledge, 2007), 4–5.

39 Baudrillard, "For Illusion Isn't the Opposite of Reality," 140.

40 Kirk Varnadoe, "The Artifice of Candor: Impressionism and Photography Reconsidered," *Art in America*, vol. 68, no. 1 (1980): 66–78.

41 Henry Fox Talbot, *The Pencil of Nature* (London: Longman, Brown, Green & Longmans, 1844).

the digit confusion that continues to appear in even the latest iterations of text to image generators, such as DALL-E-3 (Fig. 1.5).[42] The photographic appearance of the image heightens the sense of uncanniness because it is accompanied by the assumption of the subject's realness.

1.5 DALL-E-3 Image (via Bing Image Creator) created with the prompt "photograph of fingers and toes interlaced together" (December 2023).

A corpse is uncanny not only because we know that this body was once animate but because it has been reduced from a living thing to an object. It still carries the name it once had but it has completely transformed. It is no longer a person, no longer *that* person specifically. So too are the things pictured in a photograph, which are always already objectified. Baudrillard observed this, saying, "We are continually speaking of the disappearance of the object in photography—that's how it was, it's not like that now—and there is indeed a kind

42 Amanda Wasielewski, "'Midjourney Can't Count': Questions of Representation and Meaning for Text-to-Image Generators," *IMAGE: Zeitschrift Für Interdisziplinäre Bildwissenschaft*, vol. 37, no. 1 (May 2023): 70–81.

of symbolic murder in the photographic act."[43] Baudrillard was by no means the first to associate photography with death or the act of murder.[44] But his triangulation of photography, objects, and death leads in a slightly different conceptual direction to the evocation of the uncanny.

Part of what makes *AI-generated* photographs uncanny is their double objectification, first in a distant way as photographic training data and then, more specifically, as categorized or segmented forms with names or labels. The objects and appearances in AI-generated images are, in other words, like dolls assembled from individually-created parts. They resemble the things that populate human experience and yet they are not. They are doppelgangers. They bear their names and appearances but they are transformed. This is why these images can seem strange and uncomfortable, even when rendered seamlessly. They emerge from a regress of objectification.

Conclusion

If object recognition dissembles a level of understanding it actually lacks, generative AI exposes the illusion through the process of creating individuals. So, what is at stake in image generation? The extrapolation of individuals to create something new of a type, of a label. Whatever one finds oneself labeled as, therefore, generative AI can create a new one of "those" (meaning, you or me or anyone/anything else singular) and it will be non-identical to anything it has created before or after. It will be singular but still derivative, and, unlike the objects identified in an object recognition exercise, it will be something without reference to the real. It is more apt, then, to call it hyperreal, a virtual photograph with no connection to the real.

Baudrillard's assertion that "the model comes first" has never been more literal than it is after the advent of generative AI.[45] The visual generative AI model, however, is not a model of reality per se but a model built largely on

43 Baudrillard, "For Illusion Isn't the Opposite of Reality," 147.

44 Many have noted this facet of photography, see, for example: Susan Sontag, *On Photography* (New York: Penguin Books, 1977); Roland Barthes, *Camera Lucida*, trans. Richard Howard (London: Vintage, 2000); Philippe Dubois, *L'acte photographique et autres essais* (Paris: Nathan, 1990).

45 Baudrillard, *Simulacra and Simulation*, 16.

photographic representations. These photographs, in turn, have already operated on the world in a specific way. Following Baudrillard, they have objectified it. Photographs are the end of a process but also the beginning of a new process.