

Speculations for a Future Garden

Jayne Wilkinson

There are three broad areas of inquiry related to artificial intelligence and art that I'm interested in better understanding.

One: How can we chart the increasingly unstable relationship between images and language produced by generative AI? Does the rise of consumer AI—with its growing capacity for astonishing deep fakes and the widespread mistrust of images and fake (algorithmic) news—mean that the poststructuralist division between sign and referent is complete? That is, can we no longer trust that images are what they say they are?

Two: The history of photography is replete with aesthetic and technological developments where science and art overlapped, often in politically contestable partnerships with the military industrial complex and corporate think tanks. What can we anticipate about the critical reception of AI-generated or AI-produced artworks through the broader contexts of photographic histories and the reception of photography as an artform?

Three: How is machine vision coming to influence what we know about “nature”? As the world continues its descent into climate crisis hell with mass extinction ever on the horizon, generated images of the natural world have become increasingly valuable and available in extremely high definition. More real than real, they portray worlds far beyond what one could reasonably expect to see or experience. Will computational images eventually replace documentary ones? Or have they already?

In some sense, these are rhetorical questions since they can't be answered directly, and perhaps the answers are implied by posing them, but it seems imperative to think through ways of contextualizing the work of artists who are both critical of artificial intelligence and engage with it to advance our understanding of its technological and epistemic limitations. These questions are not separate from the ways that racial, gender, ethnic, and ableist biases have already been built into automated systems and are being intensified by new generations of artificial intelligence, particularly related to military and police applications. Critiques of in-built biases are often led by artists, who can demonstrate limitations and possibilities through their practices—Stephanie Dinkins, for example, is widely regarded for projects that address how to make artificial intelligence and data technologies better representative of Black experiences¹—and the work of undoing or rebuilding technologies equitably is itself a growing interdisciplinary field in which art is increasingly playing a significant role.² Still, art as such rarely makes change possible; decades of environmental art, relational aesthetics, and social practice have demonstrated this. What transforms society is widespread protest, the dismantling of violent regimes, the collective power of agitation to fight for legislative change.

1 “Stephanie Dinkins on AI, Artistic Expression, and Making Criticism Count,” *Bellagio Bulletin* (The Rockefeller Foundation), August 2023, <https://www.rockefellerfoundation.org/bellagio-conversations-in-ai/stephanie-dinkins-on-ai-artistic-expression-and-making-criticism-count/>. For example, Dinkins’s projects have demonstrated how AI applications fail to generate realistic images of Black women or have developed new apps that can train ChatGPT to respond using accurate Black vernacular language. <https://www.stephaniedinkins.com/>.

2 Kate Crawford, Sarah Myers West, Meredith Whittaker, *Discriminating Systems: Gender, Race and Power in AI*, AI Now Institute, April 1, 2019. <https://ainowinstitute.org/publication/discriminating-systems-gender-race-and-power-in-ai-2>.

However, on a smaller scale, art does have the capacity to influence one’s thinking and perspectives and can point to the social, political, and aesthetic limitations of entrenched discourses. Art loosens the boundaries, particularly around the ethical and ecological implications of coming intelligent technologies, whose narratives thus far are largely dictated by corporate interests.



As a title, “unseen garden” is a misdirection: there is no garden. A garden is an ecosystem, typically manufactured and maintained, but in Chris Hamamoto and Federico Pérez Villoro’s *Unseen Garden*, there are only single floral specimens that appear to grow, bloom, and wilt through time-lapse photography. That too is slightly misleading, as the rapidly advancing images are actually stitched together from stock photographs of different flowers—pink peony, charmer orchid, French tulip, yellow rose, purple dahlia—and combined with captions autogenerated from NeuralTalk, an early language model designed to write sentences that describe images and which the artists adapted to work with contemporary computational languages. The captions are rarely accurate. Almost all of them mention a vase, which has a certain logic but is

nonetheless incorrect, as no vase appears. Some of the generated phrases are less plausible—“a small bird sitting on a branch of a tree”, “on a table,” “with a glass of water”—and others are absurd. For example, why would “a man holding a tennis racquet on a court” be used to describe an orchid?

In the *Unseen Garden*, these conventionally beautiful flowers, with their long histories of cultivation and genetic manipulation, are symbolically loaded, conceptual rather than representational. They are registers of a different world, plucked from their ecosystems and produced as signifying concepts: beauty, growth, decay, value, genetic modification, and even psychological attributes, such as hope, serenity, or passion. Where scientific photography once bore a documentary function, to document the details of a specimen to allow further study of a species, our current image regime requires a different curiosity for reading images, where the certainty of looking is not assured. Previously, one could assume a direct, if complex, relationship between the subject (what is in front of the lens), the apparatus (the camera), and the photographer, or image-taking person. Conventionally, those three actors all existed within the same space and time, a simultaneity that contributed to photography’s magic: the three were forever connected by the print, the evidence of their encounter,

which remained after the conditions that conspired to unite them had long since dispersed.

Many of the frustrations, or fears, being levied against contemporary applications of artificial intelligence—that it takes away the human subject, that it cannot make art as beautifully or impactfully as a human can, that it is a fad, that it co-opts artists’ work, that it destroys the original, infringing on copyright—were also levied against photography, and its reception, and both analog and digital images are frequently still made to defend themselves as artworks. Ideas of misdirection, mislabelling, the slippage between an image and its referent, the leap of logic required to understand what is being seen versus what one’s mind already projects into the image—these are all concepts familiar to photography, where the relationship of text to image comes through the charged role of a caption.

In postmodern photographic theory, the caption, and language more broadly, was highly contested. Theorist and writer Susan Sontag famously argued that photographs often failed to import political meaning since it was too easy to become desensitized to what they depicted. For Sontag, “no caption can permanently restrict or secure a picture’s meaning.” Text accompanying an image is always at risk of being “undermined by the plurality of meanings that every photograph carries, or from being qualified by the acquisitive mentality implicit in all picture-taking—and picture collecting—and by the aesthetic relation to their subjects which all photographs inevitably propose.”³ Prescient arguments made well before the advent of digital images: first, that a caption (or text) cannot secure meaning for an image given the wide range of possible interpretations within the contents of an image itself; and second, that all photographs propose an “acquisitive mentality” that suggests we own or control a subject through its image. The desire for acquisition that compels the taking, collecting, sharing, and organizing images and videos is even stronger now than in the era of analog photographs and it’s a key driver in how social media influencers monetize their accounts.

The context that has shifted is how language both produces *and* contextualizes images. Rather than a camera, what creates an image now is data: metadata, tags, large language models, and image descriptors generate the images that fill our screens. Despite frequent mislabelling or inaccurate applications, the text-image relationship is vital in determining how images are made and organized, and it’s restructuring how culture is produced and shared. This is the realm of computational images, which generally refers to using processes of image-capture and algorithmic measurement instead of optical lenses or sensors to take images.⁴ Software extends the visual capabilities of the user (no longer a photographer in the conventional sense) and in so doing offers a departure point for images with no human referent. In parallel, machine vision, the automated extraction of information from digital images, enables

3 Susan Sontag, *On Photography* (New York, NY: Dell Publishing Co., 1977), 109.

4 Early examples of computational images would be applications such as MRIs (magnetic resonance imaging), ultrasound imaging, or remote sensors used to determine seismic information. Computational images now describe the vast range of images that result from software algorithms made to automatically enhance images as the user takes them; these include high-dynamic range (HDR), super zoom, panorama and automatic stitching, night mode or low-light functions, and astrophotography features.

5 Joanna Zylińska, *The Perception Machine: Our Photographic Future between the Eye and AI* (Cambridge, MA: MIT Press, 2023), 97.

In addition to the modelling of intelligence systems through language, we are also adjusting and refining our language to better refine the generated results.



So, it seems clear that text has become a necessary component of contemporary image-making, not interpretive, but generative. For images produced by and labelled with descriptive language, the relationship is inherently destabilized; the datasets on which image generators are trained seem to reproduce errors that would be obvious to a human eye, and in both cases, requires a subjective, interpretive function. In the example of *Unseen Garden*, the disassociation between what is depicted and what is described is highly contestable, since the limits are clearly parsed with the obviously incorrect and somewhat pedantic labelling. By using simple subject matter, the artists intentionally expose the frayed edges of the machine, where interpretation breaks down, to reveal the limitations not only of the machine but of taxonomic language itself—if language is not always capable of determining meaning in an image, why should a machine be able to do so? Are we destined to be awash in images that remain at the level of fantasy or imagination?

machines to perform tasks otherwise requiring human sight more efficiently, using sensor-based cameras, processing hardware, and algorithmic software.⁵ Artist and filmmaker Harun Farocki is widely regarded for his works that analyze such images as operational images, describing them as “images without a social goal, not for edification, not for reflection.”⁶ As early as 1984, French philosopher Paul Virilio was theorizing such images related to “the military use of space, whose conquest was ultimately the conquest of the image – the electronic image of remote detection.”⁷ A remote image with an operative function, “trouble[s] what an image is, as far as it shifts from representational to nonrepresentational, from the primacy of human perception of bodies, movement, and things to measurement, pattern analysis, navigation, and more.”⁸

This field of image analysis is rapidly expanding, as machine learning systems capable of generating new images and image combinations are further refined. The consumer use of generative AI relies on the language skills of its users: the drafting, editing, reworking, and revising of the prompt produces the image. A unique prompt is what helps create difference in modelling software that is designed to flatten and find averages in the aggregate. In

6 Harun Farocki, *Eye/Machine I, Auge/Maschine I* (2001), two-channel video edited to single channel, B+W and colour, 23 mins. During the 1991 Gulf war, television channels were showing striking images of bombings taken by cameras placed on the heads of the missiles, creating a new kind of visual reportage. “The eye of the camera creates an idea of subjectivity and where there is a subject there is intelligence,” said Farocki. <https://www.macba.cat/en/art-artists/artists/farocki-harun/eye-machine-i>.

7 Paul Virilio, *War and Cinema: The Logistics of Perception*, trans. Patrick Camiller, (London: Verso, 1989), 111. Originally published in French by Cahiers du cinema/ Editions de l’Etoile in 1984.

8 Jussi Parikka, *Operational Images: From the Visual to the Invisual* (Minneapolis, MN: University of Minnesota Press, 2023), preface. <https://manifold.umn.edu/projects/operational-images>.



Where does this leave current cultural entanglements with AI? For many, a generally unnerving aspect of technology is its otherness to human experience and capabilities. Writer and cultural critic Mike Pepi outlines this as a problem of recognition since technology is always framed as “somehow alien to our

Linguistic ambiguity is challenged by the absolute need for context that the political dimensions of image systems demand. For technology and art critic Nora Khan, the anxiety around artificial intelligence and aesthetics is not about the question of authorship per se,⁹ it’s about the ethics of representation that are produced when the responsibility for creation is dispersed or seemingly authorless. Writing in 2019, she articulated this imperative for living with computational images: “How we see or unsee is the primary ethical question in a culture and computational regime that privileges vision. And how we see, name, and know the world is increasingly influenced and shaped by how machines see, name, and know; machines read images and then produce a matrix of knowledge that deeply shapes how humans read images on the same platforms.”¹⁰ This presupposes another, important question around reception: How do we receive and critique AI-produced artwork when our tools for critiquing images remain embedded in discourses connected to earlier vision regimes?

Images are no longer required to bear indexicality; there is no requirement that the subject of an image or the person in it is “real”. In photography, extended debates around what constitutes the original and the copy were an aspect of discourse that was heightened by the development of a robust commercial art market for photography and photo-based work. And questions about aura, and the value associated with that aura, have never really gone away. Despite the worlds of virtuality that we inhabit, art (generally) seems to be a singular domain where the human touch, the handmade, retains a value that borders on the sacred. Interestingly, artists working with AI tend to be protective of the prompts they use and the processes they develop for engaging with the systems, the prompts being perhaps the last human, auratic, function of a machine process. Against the backdrop of digital societies and “frictionless” forms of communicating, working, and living, art seems committed to valorizing the production of objects and ideas that are decidedly human.

9 Many scholars and advocacy groups have raised concerns around authorship and copyright related to authors' and artists' rights to their work. In Canada, CARFAC has started developing recommendations related to AI and the rights of artists. See their recent report, *Generative AI and Visual Artists*, January 30, 2024, <https://www.carfac.ca/news/2024/01/30/carfac-raavs-recommendations-regarding-ai-and-visual-artists/>.

10 Nora N. Khan, *Seeing, Naming, Knowing* (Brooklyn, NY: The Brooklyn Rail, 2019), 9.

11 Mike Pepi, "Tool Time: How Technology Uses Us," *Manual Override* at The Shed, New York, December 2019, 2.

12 Ibid., 20.

13 Ibid., 21.

humanity” and the attendant panic over its application is predictable.¹¹ The task for the “techno-critical artist” is to balance a resistance to the subject’s infrastructure and an artist’s complicity within it.¹² He identifies this tendency in artists who utilize technology to express doubt: “The artist can be critical of technology, while at the same time understanding its utility in demonstrating what is at stake.” Artists can remind us that “art is not powerless against the tools of platform capitalist monopoly.”¹³ One of the driving impulses of Chris and Federico’s collaborative practice is to “become defamiliarized with a larger technical imposition” as they continually assess such positionality, asking of themselves and their peers: “How can our practices both detach and disobey technical expectations while also using our positions (and technical literacy) tactically, to bend the limits of these unescapable tools?” We can be realistic about the impact of art in helping to contextualize our relationships to technology, even ones that appear to threaten our definition of what it is to be human. There is a real urgency in understanding how the tools of artificial intelligence operate and what is at stake in using them and what their limitations are.

As both form and genre, in still and moving imagery, documentary was once tasked with representing narratives, events, and significant social and political movements to broad audiences. Now, we are witnessing a transformation from documents that, if unreliable, were proffered as evidence into images that are resolutely not documents, not evidence of reality but merely suggestive enough of “realness” to be a convincing simulation. What does it mean that machine images seem poised to replace documentary images? This question intersects in important ways with how representations of nature—the disappearing birds, animals, plants, landscapes—are increasingly rendered in CGI and computer-assisted modelling to produce incredibly high-resolution imagery.¹⁴ We can “see” more than what is optically visible, because we are seeing with machines. Visual studies and image discourse scholars in the coming years will be tasked with addressing this slippage between the indexicality of visual evidence, the facticity of data-based imaging and the “material witness,”¹⁵ and the speculative imagery of AI still and moving images created through text prompt.

There are many precedents in modernist photography for attempting to see what is too small or large or fast or slow to be perceptible, unassisted, by human sight, and such experiments were often related to investigations of nature. The scientist and photographer Harold Edgerton was, for most of

14 For example, the “nature documentary” series, *Prehistoric Planet* (2022–), narrated by Sir David Attenborough and produced by the BBC Studios Natural History unit. Depicting the planet some 65 million years ago through hyperrealistic CGI-created images of various dinosaur species on land and undersea, the series uses the same types of filmic conventions as actual nature documentaries—including voiceovers imparting human emotions and narratives onto animal eating, grooming, and mating routines, and sweeping vistas of landscapes unpopulated by humans—to convey an immersive reality based on current research but obviously without photographic precedent.

15 Susan Schuppli, *Material Witness: Media, Forensics, Evidence* (Cambridge, MA: MIT Press, 2020). Schuppli’s concept of the “material witness” addresses the “evidential role of matter as both registering external events and exposing the practices and procedures that enable matter to bear witness.” Unlike visual evidence, where the image is analyzed or dissected to provide clues, the material witness is inscribed through technical and ecological agency and interpreted through non-visual means.

his career, preoccupied with the unseen. He was well-known for inventing the electronic flash and strobe light, allowing him to take photographs with split-second timing, to effect the “freezing” of time. In 1939, he published a book of such photographs, *Flash! Seeing the Unseen by Ultra High-Speed Photography*, that was astonishing to contemporary audiences. Edgerton’s contributions were technical and scientific—he was a professor at MIT in electrical engineering, working in the school laboratories for most of his life—but also curiously aesthetic. He used photography to demonstrate the locomotion of the body in a tennis player’s swing or a runner’s gait; he photographed the tiny corona of a drop of milk and the exact second of a bullet piercing an apple. His skill was in capturing microscopic events, such that they could be appreciated and studied with visual acuity. His slowing down of time is the opposite of time-lapse photography, which aggregates and speeds up such moments, but the fascination is similar. There is a perpetual human curiosity for things that our eyes alone cannot see: the rapid unfurling of a seed into a stem into a flower has such power to captivate.



As I was drafting this text, OpenAI released its latest consumer application, following the popularity of its DALL-E and ChatGPT models: Sora, an AI diffusion model that creates short videos from user-submitted text prompts. “Sora is able to generate complex scenes with multiple characters, specific types of motion, and accurate details of the subject and background. The model understands not only what the user has asked for in the prompt, but also how those things exist in the physical world.”¹⁶

Although some of the examples are quite incredible, like many examples of generative AI, there is an obvious gap between inputs and outputs. It becomes an exercise in close looking, though to what end is unclear. These videos are a significant advancement from early versions of DALL-E in their convincing qualities, but it doesn’t take serious analytic skill to discern that these are artificially generated—and slightly weird as a result. A video showing a Victoria crowned pigeon with realistic eye movements and the familiar pigeon neck wobble was generated using this detailed prompt: “This close-up shot of a Victoria crowned pigeon showcases its striking blue

16 OpenAI, "Sora," accessed February 16, 2024, <https://openai.com/sora>.

17 OpenAI, "Flower Blooming," accessed February 16, 2024, <https://cdn.openai.com/sora/videos/flower-blooming.mp4>.

plumage and red chest. Its crest is made of delicate, lacy feathers, while its eye is a striking red color. The bird's head is tilted slightly to the side, giving the impression of it looking regal and majestic. The background is blurred, drawing attention to the bird's striking appearance." The language is self-describing. If you image search this type of bird using its name alone, you'll find photographs that

show its blue plumage, red chest, delicate feathers shaping its head, and features that could reasonably be described as "striking" (used twice in the prompt) "regal" or "majestic". While we easily categorize some animals as regal, others as dirty, thinking some are majestic while others ordinary, these are human interpretations, imposed by our concepts of beauty or behavior or even of racial or social bias, not observable characteristics of the species itself. When the caption produces the same image that it also annotates, describing what it has made using the same terminology, it embeds those ideas into the very process. Rather than act in a taxonomic way, the generated "natural" image produces something between a fantasy and a hallucination. The title of this essay is partly flippant but partly genuine: Will our future gardens be digital reconstructions that only slightly approximate the species that flourish on earth now?

One of Sora's demonstration videos could easily be a component of *Unseen Garden*.¹⁷ It shows a flower slowly opening, on a windowsill, in a ceramic pot (rather than a vase), and in a direct beam of sunlight. It is also, obviously, generated: the space doesn't make sense, the sun seems to be shining at night, the bloom is too large for the squat stem. It's not surprising that OpenAI would make such generated videos as examples of Sora's capabilities, playing on the human fascination of watching plants and animals in minute detail, or of taking in the wide sweeps of a beautiful landscape that could rarely be seen in person.

There are still so many questions—about copyright, appropriation, originality, technology, humanity, and even more old-fashioned ideas about the role of beauty and aesthetics—being tested against the technologies of artificial intelligence and the increasing prevalence of machine vision. The social and political implications of these shifts between real, imaginary, and simulated image worlds—spaces that we will soon inhabit along with all our other digital daily tasks—are in flux. That we are now capable of creating images with the same language we use to describe them is a somewhat frightening loop. Ultimately, it should remind us of the necessity and urgency of visual literacy as a vital tool for cultivating the seeds of future knowledge.

