AI-generated images of birds produced by various GANs, 2018.

A Questionnaire on Art and Machine Learning

We pose the following questions, which will surely prompt still others: *If AI algorithms have the capacity to analyze massive datasets and identify patterns, how has this capability influenced the generation of new artistic concepts and ideas? How are artists collaborating with, changing, torquing, or critiquing AI systems? Compared to, say, the history of photography and chance, or art and systems, does artists' use of generative AI today represent a difference in degree or kind? How have artists probed the creative possibilities of generative AI? How have they condemned the biases, ecological impact, and military-industrial origins of AI? Developed their own hybrid models and architectures? Explored the instrumentalization of AI systems, or, on the other hand, their unpredictability? What other models are possible? What is no longer possible? What is human or machine, creativity or computation, in the first place?*[1]

—Michelle Kuo and Pamela M. Lee for the Editors

1. Definitions for some of the technical terms used in this questionnaire can be found in its introduction, starting on page 3: "Generative and Adversarial: Art and the Prospects of AI."

OCTOBER 189 Summer 2024, pp. 6–130. © 2024 October Magazine, Ltd. and Massachusetts Institute of Technology.
https://doi.org/10.1162/octo_a_00533

K ALLADO-McDOWELL

Recently, in a ceremony with entheogenic plants, I had cause to consider the differences between organic and synthetic hallucinations. Eyes closed, in the dark, without visual stimulus, I watched a parade of images arise internally. I've had this experience many times, and each time it is different. Every vision takes its own form. I've seen cartoon figures, panoramic geometry, ecological cinema, impossible abstractions. In appearance, these images have not been like any art. In this case, however, what I saw bore a disturbing resemblance to the outputs of an AI image generator. The forms hung in rectangular frames. They passed by in rows like paintings on a wall, floating against a dark background. Toning down the iridescent palette common to psychedelic visions, their colors were muted and earthy. Curves ended in diffuse edges. They moved in and out of clarity. Sections were covered in letters and words like the glyphs drawn by early versions of DALL-E or Midjourney.

The mind moves through the world like Gabriel Orozco's *Yielding Stone* or the magic ball from the video game *Katamari Damacy,* which pick up physical traces from the worlds in which they move. It rolls around and accumulates experience, desire, trauma, joy, pain, and all kinds of information. Some of these stick, while others fall off. Entering an entheogenic state is like dusting off such a sphere, or, at higher doses, like melting down the plasticine matter of mind to pull out stuck junk with the help of intelligences spiritual, vegetal, and ineffable. A full cleanse strips the mind to its bare surface. But even a simple wipe-down can show you what's stuck to your outermost layer, the crud of the day or the week, or an irritation tied to a deeper jam in the mental sphere.

In this case, I was shown a reflection of my most recent visual experiences. I'd been spending obsessive hours making images in Midjourney, learning to probe the multi-modal text-image network, as I have done intermittently since the tool was first made public. Even my ordinary states of perception have been changed by this practice. AI image-making, like drawing and painting, focuses attention on the color and form of life. When making art, we build up layers of sense memory, aided by physical gestures, smells, and movements—all forms of embodiment that AI lacks. Yet the visual imagination remains entwined with everything that we see on screens. Despite their purely visual character, these images shape our creativity.

This is what I saw when my organic and synthetic hallucinations mixed. I felt shocked, even a little guilty. Had I corrupted a sacred space? Had I corrupted myself? My inner visual-processing system had internalized enough of the structure and character of AI outputs that it reproduced them when stimulated with psychoactive alkaloids. Internalizing a tool is a necessary step toward mastery. The brush, the typewriter, and the car become an extension of the person using it. Thus the tool digs deeper into the plasticine mind-sphere and flesh. The body is reconfigured by even the disembodied practice of AI image-prompting, regurgitating its substance in the dosed, hyper-connected brain. Becoming a power user

means being changed by a medium. Being altered by a practice, its structure and language, is the cost of fluency. As we learn to drive, write, scuba dive, as we navigate psychedelic realms, or office politics, or the high-dimensional latent space of a machine-learning model, we become those things. We are changed by them—when we see ourselves, we see them.

The vision passed. But the experience revealed how much I'd been contaminated by the machine. It's worth noting that AI images—like all AI outputs—are, technically speaking, hallucinations: They are generated from within the internal structure of the neural network. Lately, the term *hallucination* has come to mean untrue statements generated by AI language models, such as those used in search engines to (sometimes incorrectly) summarize a topic. But even the true things that an AI says constitute internally generated hallucinations. A similar revision could be made for our use of the term *hallucination* when referring to psychedelic states. The kind of vision described above is generated internally, but so is every moment of consciousness. According to current neuroscientific theories, the reality we experience is a collage made from multiple conflicting sense inputs, memories, etc.

Dividing experience into the binary categories of the hallucinated and the not-hallucinated reduces complexity in favor of the untenable objectivity upon which rests modernity's isolating individualism. When we acknowledge that a significant portion of our lives is fabricated by the brain, we begin to understand ourselves as more than crude blocks of observation and rationality. Rather, we are sculpted by our environments, our minds and bodies pressed into them like clay or plasticine.

Art is one way that we intervene in that pressure. When viewing or making art, we take control of sense input and redirect our own training. Ideally, by curating our experience, we become more of what we desire to be. This construction of better-hallucinated selves is perhaps a good target for both art and AI. Attention is the needle that etches our future hallucinations. This is quite clear when inner vision is amplified, and perhaps this is why so many have detected revolutionary potential in the use of psychedelics. But psychedelics do not guarantee a good intention or etching. Psychonauts value what they call set and setting (the state of mind one brings to a psychedelic experience, and the social and physical context for the experience). Without proper set and setting, self-directed interventions can go very wrong.

How to bring art into hallucinations (and hallucinations into art) is therefore a delicate question. If all experience is hallucinated, and all minds are plastic, shouldn't we take care not to replicate the neurotic complexes and curses of the past, so as not to corrupt the minds of the future? Or is this exactly the resonant magic that will draw out the psyche's ills? It may be that we need new skills of observation, judgment, and sensing in a world in which we are hallucinated subjects. We can imagine that, as more and more media become hallucinatory (that is, more like us), our sense of ourselves will come to us through ever more refractive hallucinations, whether provided by machines, plants, or social systems.

I am concerned about this change in subjectivity. It will make us more porous, and more accustomed to porosity. It requires that we become more aware of and stronger in immunity. It makes us prey and predators in an ecosystem of meaning. Orientation is difficult; we've been losing our ground and must find it again. Art may have a role to play in this. It may be put into healing service. It may be employed as an excuse by subliminal or nefarious actors. But the history of art is one of escape: from the cave, from the church, from representation, from form, from linear narratives of progress. Artists don't like to be boxed in. Art finds a way to maintain its freedom. Finding freedom and purpose in a world of hallucination could be art's most important mission in the twenty-first century.

K ALLADO-McDOWELL is a writer, speaker, and musician.

AMERICAN ARTIST

My position on AI, and with new technologies in general, is that, while I'm aware of their capabilities, maybe more so than a lot of artists, I've also retained a certain amount of skepticism. I've been fascinated by looking at the ways in which AI and algorithms continue many of the systemic relations that we already know to be true within society. I'm trying not to think of it as this new, explosive moment—which it is in many ways—but rather as a continuation of everything we know to be true, such as systemic racism and imperialism.

I think most people's primary association with AI is visual culture, maybe pop culture or social media. But one of the major ways in which it is used—that is, within the legal system or the penal system—could have violent outcomes. For example, AI is used to determine the recidivism rates that decide people's criminal sentences. And oftentimes the use of AI results from law-enforcement agencies or judicial systems not having the amount of funding they want. And so they buy an AI to replace certain jobs at their discretion. From their perspective, they're doing a good thing—but the ramifications are disastrous.

I think my approach to incorporating AI is as a means of experimentation. Aside from making works that are explicitly critical of AI, I've also used AI as a research methodology—as another way of generating an image or thinking about a visual possibility for something that I know doesn't exist or can't be seen or of which there's no documentation.

The first work I made using AI was *Sandy Speaks* in 2017. It used a very low-tech chatbot, basically. I was looking into the death of Sandra Bland in Texas in 2015: She had been arrested at a traffic stop, gone to jail, and then died in prison. There was a lot of speculation about how that happened and why. I became interested in the fact that she had a pretty robust social-media presence prior to going to jail. I was interested in her use of language to critique the police, to offer words of encouragement, to get people to think about the Black Lives Matter movement, or to think critically about their relationship to policing. I had the idea of having a chatbot that reflected some of the language that she used, that felt like it had a similarity to social media in terms of the visual output of short text messages.

The original iteration of *Sandy Speaks* was commissioned by a Web platform that I don't believe exists anymore called newhive.com. They did a lot of experimental artist projects. At the time that I was interested in newhive.com, there were a lot of Black Internet artists using it. *Sandy Speaks* began as a Web-specific piece; there was a later iteration where it was a projection on a wall, and it was an all-black screen with white text, and there was a console that you could type into. It essentially functioned like a chatbot that looked like a messaging app, and you would type questions. You could ask how to interface with law enforcement or about prison surveillance statistics or about what happened to Sandra Bland and it would respond. The tone of it, though, was more reflective of how Sandra Bland would speak than of how a dry, austere chatbot would sound. At the time, the plat-

form I used was not easy to adapt. I had to program a lot of the language and basically anticipate every possible phrase someone would say. Whereas now machine-learning technologies are much more advanced; you could probably make that AI in ten minutes if you asked the right questions. It was a very rudimentary chatbot language, but all the responses were handwritten. It was a writing project in addition to coding, trying to reflect the voice that Bland had in her social media as a way of engaging with her legacy.

More recently, I used AI in my installation *My Blue Window* (2019), a work that looked at existing AIs that are already used by law enforcement for predictive-policing technologies. For the installation I made a video. It was titled *2015* (2019) because I wanted to suggest that year as the time in which the footage you're witnessing takes place. The installation was created for an exhibition at the Queens Museum curated by Sophia Marisa Lucas. The video looks like dash-cam footage filmed throughout New York City, primarily Southeast Brooklyn, where there's usually a lot of police activity. You see a heads-up display that looks like something out of a science-fiction movie and a map of Brooklyn that's responding live to where the driver is going. And at the same time, the interface is telling them where to go, and it's anticipating locations on the map where a crime may be happening, and the police officer is driving to the locations, and you don't really know what exactly they're doing to stop a crime from happening. But then suddenly the system says, "Crime deterred." But nothing has happened. It points to the absurdity of the system. But also its inefficiency. Predictive policing is so often thought of as something that's hyper-efficient, but it's quite clumsy, and it's responsible for people's lives. People might end up in jail or dead because of something that the algorithm said. All the data is provided by police officers and is based on past arrest and crime data. So the system inherently has all the biases that police officers would have if they were doing the policing themselves. In this case, I'm approaching AI as an outsider and trying to show the ways in which this thing that sounds like it's of the future is actually of the past. I'm trying to change the perception that everything we're experiencing with AI is new and that the technology can offer salvation or something to be optimistic about, when in reality it's an optimization of the narrow thinking that causes most systemic problems.

I think the unpredictability of AI is one of the things that are less often discussed, but it points to how much the technology doesn't work or doesn't deliver on its promise. I think it's washed over with a lot of the language that's used around it. But often AIs don't work the way they're supposed to. And for some reason, as a society, we're very forgiving of that, even though these systems are hyped as the things that are going to determine our reality in the near future, about which we should be really excited. I don't know if I can get excited, knowing how often it fails. In the case of predictive-policing technologies, for example, there's not really a way to know whether it's working. And so if a police officer goes to a space where an anticipated crime is going to happen and nothing's happening, great. They prevented a crime. If they go and something *is* happening, then they

say, "It's because our system predicted it." No matter what happens, they can say it's working. There's no room to detect failure.

My most recent use of AI is titled *Security Theater* (2023) for the Guggenheim exhibition *Going Dark: The Contemporary Figure at the Edge of Visibility*, curated by Ashley James. *Security Theater* is this mysterious orb that hangs down in the center of the rotunda, and inside of it—you can't tell this from the outside—there are eight cameras, four facing down, four facing up. And they're state-of-the-art surveillance cameras that might be used in a factory or a really big metro station to be able to see every single person who enters and potentially even track them using their face or certain features. The AI used in that piece is straight out of the box, the way it would be used normally. What you witnessed if you went into a certain room that was on the fifth floor was the AI deciding who or what is human and accordingly placing a red box saying "human" around certain objects.

The reason I was particularly interested in doing that was to connect the Guggenheim's architecture with spaces of surveillance, specifically the panopticon model. Museums are placed on a pedestal of cultural reverence. But of course, like many other municipal buildings, they have a lot of mundane bureaucracy, and you have to pass through security to get in the building. I also wanted to show the way that this carceral logic has entered every social space that we occupy, particularly in the United States. In drawing this connection between the museum, the courthouse, the school, the prison—they all have some similarities in terms of how they regulate people.

For people to see the security footage, they would have to place their phone in a sealed pouch, which basically prevented them from interacting with their phone. It's an analog technology to prevent people from using their phone, wanting to take people off the grid, somewhat forcibly. That was really interesting to me. Regardless of their seeing the interface or being seen on the camera, I wanted to have that experience of having your phone taken away. And I think some people were very frustrated by the experience, but some people felt it was meditative, or they felt relaxed by being released from this burden of responding to their phone constantly.

There's something about AI's ability to capture such a massive amount of data that implies an exaggerated amount of control: It gives the public, or the people making laws and deciding where police go, some sense of security. And I think in the case of AI that was used in *Security Theater*, it really is playing into that concept, because all the characteristics that it's able to look for are so specific and so robust. You could upload a photo of a specific person and it can tell you where they are in the building at any moment. The amount of computational power to be able to do that is quite scary. But I think in that same sense, it's such a vast amount of data that it's indiscriminate. I think that can result in negligence. It can also result in inefficiency, even though it's promoted as highly efficient. In response to a feeling of a lack of security, there's this overindulgence in what the software can do. But then the consequence is that maybe it can't do the thing it's supposed to.

I'm reminded of the commonly held belief that competition is the natural way in which organisms and species interact. And therefore we should model our human behavior on competition. But I think there's a blind spot in the way these sorts of technologies, which are being designed by people, could actually reflect so many other aspects of how species operate. There's something within AI that is modeled after certain aspects of human behavior or biological behavior that might reinforce these stereotypes about competition or domination. And yet there's so much evidence for collaboration or other ways of interacting that lead to more sustainable ecosystems, or species that live much longer.

I'm currently using AI to expand a body of work called *Shaper of God* (2021–ongoing), which is about the sci-fi author Octavia E. Butler. I had started using Midjourney, a text-to-image AI generator, just to see what it could do to make images. As part of my research into Butler's life, I was interested in a chicken farm that her grandmother had in the California desert in the 1930s, thinking of that site as foundational for her family's migration from Louisiana to California. But it burned down when Butler was a child. There's no visual image of the chicken farm. I was interested in using AI to imagine it, as another way of sketching or seeing what the scene might look like. For me, it was important to take this tool that's often used for speculative modeling for a tech company, or something financially directed, to tap into a *Black imaginary* methodology, like speculative fiction. Not to use it to develop a final product, but as a way to assist my own imagination.

Humans and machines have similarities in that both have the creative ability to make connections between disparate things that wouldn't seem to go together, though a machine will have a harder time recognizing when there's something valuable in an unexpected connection. I think the role of the artist is having an interest in connections that aren't normally made. It might not be considered a valuable connection to most of humanity, but the artist is obsessed with their own idea of why this thing is good enough to bring into reality. There's an absurdity required to be an artist. Maybe a machine can emulate it, but what is a machine's sense of value?

We can't unknow AI. It's not going to slow down. Even if all the universities and government researchers and big tech companies stopped what they're doing now, the cat's out of the bag. Someone's going to take it and run with it. We can't unknow what we know. So what else is possible? Outside the infrastructure of AI itself, there's the cultural context in which it operates, which absolutely determines what it can do or who is able to access it. We know that AI is, despite all the hype, still prohibitively out of reach for most people, and still tinctured by the same Silicon Valley ethos of universalism. But maybe that's something we can address: We can take down the myth of universalism if we have the collective will to do so.

—As told to Alex Fialho

AMERICAN ARTIST makes thought experiments that mine the history of technology, race, and knowledge production.

NANCY BAKER CAHILL

> *The difference between animals and fungi is simple: Animals put food in their bodies, whereas fungi put their bodies in the food.*
>
> —Merlin Sheldrake, *Entangled Life*

It's hard to talk about "AI" decoupled from the entangled, multi-systemic tragedies that are currently unfolding in real time: burgeoning autocracies, rising oceans, economic asymmetries, genocidal wars. I can't do it; algorithmic systems are threaded through all of them. The polycrisis includes interlocking, causal, corollary relationships within a vast, expanding system. We might need metaphors, to build on feminist technology researcher Maya Ganesh's work on metaphorology,[1] to discern how art and AI may play a critically important role in reshaping certain outcomes. Machine-learning tools could be trained to co-build more equitable, unexpected, ecologically resilient futures. Currently, however, most machine-learning models include extracting and manipulating data and creating profitable dependencies. Such systems ingest data, energy, IP, privacy and excrete images, texts, advertisements, code, analyses. Because datasets are often hideously biased, outputs are contaminated. An alternative, open-ended model I have long examined in my research and practice is mycelium, the networked fungal colonies that form the connective tissue of all carbon-based life on earth, which provides an irrepressible, reparative distribution of resources. Mycelia (re)generate life through a decentralized, interoperable communication system, transferring data and sustenance, care and support *from within what it metabolizes, whatever it metabolizes.*

Civic accountability matters now more than ever, given the fragility of democracy coupled with capitalism's digital spawn—what economist Yanis Varoufakis calls *technofeudalism.* Marshall McLuhan famously referred to art as "a Distant Early Warning system that can always be relied on to tell the old culture what is beginning to happen to it."[2] When artists embed their practices in algorithmic systems compromised by corporate or militaristic interests, they can reorient their tools to grapple with ethical questions through evocative and open-ended civic or political provocations. Artists are experts at identifying fissures, vulnerabilities, portals in and through which to build reparative algorithmic ecosystems within systems. They are also interventionists, subverters, hackers. Art's warning system should be blaring. Under the influence of artists, machine learning could allow for the public to benefit from creative interventions in everyday technologies. The challenge of ecological thinking in a moment of hyper-individuation is that, among other things, it demands asking ourselves what we can do for each other, versus what we can take. This may require some training.

1. Maya Ganesh, "Between Metaphor and Meaning: AI and Being Human," *Interactions* 29, no. 5 (Sept.–Oct. 2022), pp. 58–62.

2. Garnet Hertz, *Art + DIY Electronics* (Cambridge, MA: MIT Press, 2023), p. 247.

I often seek guidance in bio-mimetic models. In what is known as *overshoot*, for example, humanity extracts resources beyond the earth's ability to regenerate them, resulting in ecological collapse and ancillary collapses (social, political, and economic) at scale. This phenomenon, already underway, is mirrored in certain machine-learning models. Computer scientists have given it a name: MAD, an acronym for model autophagy disorder,[3] which presents its own poetics of overshoot. Without replenishing datasets with fresh (human-generated) data, certain generative-AI models mired in the monocultural data waste they've produced ultimately self-consume and collapse, incapable of generating useful outputs. This is only one such resonance, but for me it signals a parallel warning it would be wise to heed. Mycelium is so resourceful, it can turn the most toxic fields into viable agricultural sites. How can artists do the same?

Many have already outlined relationships with synthetic intelligences that operate from a fertile (generative) space of ecological thinking from the inside—essential for imagining at scale. The research and revelatory insights of artists, coders, and creative practitioners like Sarah Rosalena Brady, Joy Buolamwini, Sougwen Chung, Stephanie Dinkins, Holly Herndon and Mat Dryhurst, and Lauren Lee McCarthy, to name merely a few, have reframed and rebuilt algorithmic and robotic relationships. Actual or imagined human-machine, machine-organism fusions and collaborations offer startling insights into better relationships with intelligences of all kinds: hybrid, fluid, elusive, and inchoate. Warm Data Lab founder, artist, and writer Nora Bateson describes one type of ecological thinking as being "simultaneously implicating." She asks, "What stays the same, and what changes? . . . Each organism is shaped by the others in its environment and is shaping the others simultaneously. They constantly implicate one another in an ongoing mutual learning (*symmathesy*) oriented toward continuing vitality."[4] To acknowledge this level of interdependence requires a degree of epistemic humility that can and should be applied to machine-learning and AI systems. Artists acutely aware of this are inventing new adaptive languages, shaping as they are being shaped. I will share a few examples that both embody and propose new relational models without falling prey to *techno-solutionism*.

Mashinka Firunts Hakopian's recent work of ficto-theory, *The Institute for Other Intelligences,* draws from critical media studies and contemporary reporting on AI to weave a speculative narrative set in a distant future. In it, she upends techno-fetishistic takes on culturally and racially biased aggregated intelligence. She introduces an "Artificial Killjoy," based on Sarah Ahmed's feminist killjoy, a figure who disrupts the happiness of others by naming conditions of injustice and speaking what would otherwise remain unspoken. Hakopian's killjoy is the Institute's director and storyteller, who performs as offsite memory storage for human agents—an embodiment of

3. Sina Alemohammad, Josue Casco-Rodriguez, Lorenzo Luzi, Ahmed Imtiaz Humayun, Hossein Babaei, Daniel LeJeune, Ali Siahkoohi, Richard G. Baraniuk, "Self-Consuming Generative Models Go MAD," arXiv.org, July 4, 2023.

4. Nora Bateson, *Combining* (Triarchy Press, 2023), p. 103.

machinic memory work as care. The killjoy is also an activist figure engaged in political education that's meant to intervene in and shape distributions of power. Hakopian's killjoy tells of an Armenian woman who greets a supercomputer purportedly containing all the world's "knowledge." In Armenian, a common way of asking "How are you?" is "What is there, and what isn't there?" Faced with the unanswerable "What isn't there?," the supercomputer explodes. Her simple question cuts right to presumptions about repositories of knowledge—who produces, stores, encodes, reifies, historicizes, and metabolizes them. This reveals that there were no Armenian speakers involved in coding the supercomputer, and that the omissions and flaws baked into socio-technical systems also provide openings or spaces of possibility for short-circuiting those systems. The artificial killjoy's guiding narrative allows Hakopian to highlight many of the systemic externalities of machine learning that could be otherwise, from predictive policing to surveillance capitalism to racial and gender bias. Doing so, she deftly reveals what might be possible if we developed the tools for attaining alternate, more equitable ends. *The Institute for Other Intelligences* inverts assumptions of algorithmic opacity and epistemological asymmetry. Mutual respect among and between species depends on novel structures of ecological thinking.

Immersion in an ecosystem invites us to act as fungi do, by putting one's body "in the food." A prodigious vision of blending and growing this way arrived in the 2023 animated television show *Scavenger's Reign*. Over the course of the series, which appeared on Max, a robot named Levi, who is stranded on a foreign planet along with a human crew member, is slowly subsumed by an indeterminate local species. We witness the machine begin to experience sentience, consciousness, and emotion. Levi's mutation is not initiated by a human hand but rather by the multivalent, tentacular ingressions of the alien ecosystem they now inhabit. Levi becomes, quite literally, entangled, and what it loses (human-programmed servitude) it gains in curiosity, empathy, inherited agricultural wisdom, and courage. Inevitably, Levi learns and speaks a new language and becomes uniquely transspeciated, and not at the hands of some Frankensteinian engineer.

Evolving ideas about what is human often prompt fear. I personally find it liberating (even a relief!) to imagine a future in which humans acknowledge a decentered status but are still integrally embedded in overlapping ecosystems—machine or otherwise. I explored these themes both in my augmented-reality (AR) project that imagined a future human, *CORPUS* (2022), and in *CENTO* (2023), my recent participatory project commissioned by the Whitney Museum (also AR). Augmented reality is a digital layer of content superimposed on what people call "reality" and is experienced through the visual prosthesis of a mobile device or tablet (if you've used Snapchat filters, you've used AR). *CENTO,* a term that describes a collage of poems, is applied here as a collage of species: machine, mycelial, avian, marine, cephalopod, and reptilian. Experienced from the eighth-floor balcony of the Whitney, the transspeciated *CENTO* soars over the meatpacking district in AR. The creature is co-built in real time by an unknown, global audi-

ence through a process of selecting and adding digital feathers—each with its own distinct properties—to its body. Each of the feather's properties includes features that would ensure survival in the face of ongoing ecocide. Based on these additions, *CENTO* mutates in real time as feathers accrue. The project was inspired by—among other post-humanist art interventions and pieces of scholarship—Rosi Braidotti's essay in *Anthropocene Feminism*, Donna Haraway's *Staying With the Trouble*, Karen Bakker's *Sounds of Life*, and Jeff VandeMeer's *Annihilation* and *Strange Bird*. I wanted the work to reject anthrosupremacy while modeling a process of communal engagement that might mimic an interspecies body politic invested in survival.

To highlight the potential of machine learning as a powerful tool of interspecies communication, *CENTO*'s machine "head" functions as a translation matrix to allow all of the species' respective cognitions to function as a single organism. *CENTO*'s body thus becomes embrained, alive with multiple distributed intelligences. Digital bioacoustics are another realm of applied AI that, if practiced ethically, might one day allow for nonhierarchical communication with other species. The key would be to collaborate with AI and data technologies to listen closely and learn before attempting to communicate in syntaxes and cultures we couldn't possibly yet understand. Centering an ethics of humility would not only reveal how little we know, it would, one hopes, help to program tools that support vulnerable species rather than exploit them.

Alas, this degree of sensitivity is scarce in profit-driven machine-learning contexts engineered to support exponential growth, accelerationism, and geopolitical advantage. A precious resource available at no risk of overshoot is the artist's imagination. What is more innately generative and unpredictable than creativity, with or without algorithmic collaboration? The relentless noise of social media and overcommunication makes creativity of all kinds harder to discern.

What is urgently needed is the presence of a "heretical consciousness"[5] in interruptive creative inversions of all kinds. Incalculable global precarity requires alternative models, new guides, and new systems with *n*th-order effects that serve rather than deplete ecosystems. Artists are key to imagining beyond what has already been imagined; they will constitute a rhizomatic force to be reckoned with if they can restore even a fraction of the balance needed for collective resilience. Training AI is an evolving and crucial art. Let's learn from those who thrive on open-ended questions, who imagine more equitable futures, and who refuse to accept what's currently being served up by taking their cue from fungi and literally *getting into it*.

NANCY BAKER CAHILL is an interdisciplinary artist and filmmaker.

5. Byung-Chul Han, *Psychopolitics: Neoliberalism and New Technologies of Power* (Verso, 2017), pp. 76–83.

IAN CHENG

Douglas Adams is always useful:

> Anything that is in the world when you're born is normal and ordinary and is just a natural part of the way the world works. Anything that's invented between when you're fifteen and thirty-five is new and exciting and revolutionary and you can probably get a career in it. Anything invented after you're thirty-five is against the natural order of things.

I'm forty. But I've learned a trick to stay alive to the new. I've begun to see technology through the eyes of my kids instead of through my own. They are four and three. AI will frame their way of life the way the smartphone currently frames mine. Their generation will be the AI generation.

What will an AI-centered life look like in 2040?

My bet is that AI won't be synonymous with the LLMs or generative models of today but rather with AI *agents*—systems that can autonomously take action on your behalf and take your feedback seriously.

Many startups are racing to develop personal AI: agents framed as companions, guides, assistants. Like the movie *Her*. But this is too small of a dream. A simulator for familiar human relationships. The future is destined to be wilder. And in a wild future, I would want my kids to have technology that amplifies their sense of agency.

I keep thinking about the movie *Venom*, in which a disagreeable alien goo and a disagreeable human learn to symbiotically move in the same direction.

I imagine for my kids something like an AI symbiote: an intelligent sentient envelope that would experience physical life beside you and develop better and better cognitive maps of you along the way. These maps, and the process of updating them, would be the symbiote's most guarded treasure. As you grow up and more of your social and work life shifts to digital mediums, your AI symbiote becomes a proxy or emissary of you, acting on your behalf, brokering access to your maps with other people and the symbiotes of other people.

An AI symbiote will unnaturally extend agency. It will enable an individual person to do and be much more in more places at more times. Everyone will become a super-organism. And like the head of any org, you will wrestle with wrangling and leading it in the direction of your priorities, or being ruled by it and recruited into its priorities.

My wish for my kids is that they will make multitudinous careers and lives out of all the wildness that is emerging now. And from this view, I have no complaints. It is unambiguously the most interesting time to be alive.

IAN CHENG is an artist and founder based in New York.

KATE CRAWFORD

Latent space is having a moment. This abstract form of semantic-data representation is a central feature of how generative-artificial-intelligence models work, including generative adversarial networks (GANs) and variational autoencoders (VAEs). Put simply, it is a mathematical space of thousands of possible dimensions that attempts to capture a highly compressed version of an AI training set. Contemporary training sets can contain billions of images and captions, which are commonly harvested from the Internet. Coherent data points that we might recognize—a word, a picture of a cat—are encoded, translated into numbers, and arrayed across a vast sea of matrices. The closer two data points are in high-dimensional latent space, the more similar they are said to be. In this post-visual manifold, a neural network detects underlying patterns, proximities, and relationships in order to produce new texts, images, or videos based on a prompt like "the pope wearing a puffer."

There are interesting questions to ask about latent space and how it transforms the base matter of training datasets into new images and texts. What is being lost in the extreme compression of words and images into numbers? Which logics are unintentionally being inferred from the training data? Is the structure of latent space producing a skew toward certain types of image and text cultures over others? What role will it play in wider media-making practices? But these questions are generally overlooked in the rush to mythologize latent space as magical or godlike for the way it generates endless content from an abstract vector space. This tendency is more evident in the arts and humanities than in computer science, where latent space is generally considered quite prosaically as just the latest in a set of techniques that stretch back to the 1960s, when unobserved latent variables were used to simplify statistical models.

Contrast this to the art world, where latent space has been described as being synonymous with the infinite. Refik Anadol's *Unsupervised: Machine Hallucinations* typifies this more rapturous genre, an artwork that places the viewer at the foot of a monumental cascade of illuminated, shifting AI-generated undulations as though we are seeing inside the workings of a machinic consciousness. In his discussion of latent space, Anadol claims that this "multidimensional landscape is really—I don't know how else to say it—divine."

Quasi-religious accounts of latent space are an extension of the current deification of artificial intelligence more broadly. The largest survey of AI researchers, conducted in early 2024, revealed that most now believe they are building autonomous intelligence. Many prominent AI leaders, including Sam Altman and Elon Musk, already treat AI as superhuman, worrying that it will become so independently powerful that it will pose an existential threat to humanity and that the only hope for our survival lies in their efforts to dominate it. In this sense, Silicon Valley has become the eschatological center of the global information economy, run by a priest class of engineers and executives convinced they are creating a powerful new consciousness: The only open question is whether it will be a benevolent or a vengeful god.

While we may roll our eyes at this literal deification of AI, it's having the intended effect on investors and policymakers alike: They are supporting obscene market valuations for AI products and companies and advocating that the industry monitor itself rather than be subject to any substantive government regulation. Indeed, the more that AI ascends the ladder of the theological, the less jurisdiction is ceded to the secular world. The metaphysical discourse about AI also obscures the material realities of political economy: the industry's dramatically increased demands for energy and water, its exploitation of crowd labor, and its continued mass extraction of data without informed consent.

Now the broader exultation of AI is being applied to latent space—as if these hidden mathematical functions were "magically" producing poems, plays, and visual art—without any heed to the billions of dollars, global data-center networks, and constantly churning data harvesters that underwrite it all. Instead, we need a much closer examination of the material costs of latent space as well as what is actually happening in the process of statistical reduction and data extraction. Rather than an expression of divinity, a closer examination reveals, latent space is profane, a triumph of statistical banality. Here, in the many steps of compression, association, and ordering, we can consider how this abstraction of data is producing a kind of probabilistic averaging that both shapes its direct outputs and the wider forms of cultural production that it influences.

What does this look like in practice? Well, we could start by downloading a large dataset used for training, like LAION 5-B. This widely used set contains five billion images and text captions taken from the Internet, and it was used to create the popular text-to-image model Stable Diffusion. If we apply a generative model like a VAE, the data will be compressed into a long series of ten-digit numbers representing an averaged set of features. These codes are assigned places in a structured latent space, where each dimension corresponds to a "meaningful" aspect of the image, such as type of object, color, angle, noise, lighting, background, or type of camera lens. Already we can imagine that this kind of compression in a text-to-image model might constitute a form of aesthetics by standardization, a set of reductive processes that predetermine which features of an image are treated as computationally significant.

But many of these decisions about what matters in latent space aren't directly made by humans. In the unsupervised machine-learning environments that predominate in the current wave of generative AI, we simply can't see what features are being detected or manipulated within high-dimensional latent space. These features can be very abstract or information-theoretic in nature, like entropy rates and noise, and they aren't readily decipherable by humans. The complexity and abstractness of these features make it challenging to fully understand or predict the behavior of generative models.

For very large models with trillions of parameters, spurious features get encoded that don't correspond to the data at all: The model essentially invents features to exhaust its representational capacity, using its available computational

resources to extract more details and detect relationships in the training data. When a model has too much capacity, it can start to create and store random patterns that don't actually exist in the real data, which creates cascading errors and hallucinations. This is one of the characteristic problems of these systems when spurious outputs are generated or a large language model invents answers or generates sources for its answer that don't exist (which I like to call "hallu-citations").

While the statistical classifications in latent space are not human-generated, they are still the result of human decision-making, albeit in an indirect way. All the content in latent space comes from training data, so that data becomes the *Weltanschauung* of the model: It sets the parameters of the possible. This training data is far from a neutral or representative collection of all human visual and written culture, if such a thing were even possible. As researchers in the Knowing Machines research project have shown, the largest sources of images and captions in LAION are e-commerce and aspirational-shopping sites, with the top three being Shopify, eBay, and Pinterest.

Content from shopping and bookmarking sites is particularly useful for methods of statistical curation thanks to ALT tags, typically short and mundane descriptions of what is occurring in an image, originally intended to aid visually impaired users. Pinterest generates captions on its pages from ALT tags, so users have become adept at writing tags for their Pins, thus constituting a highly effective unpaid labor force of crowd workers who are unaware of how their work is being capitalized. Many other shopping sites automatically generate ALT tags as well, so large e-commerce platforms are a very convenient source for harvesting millions of images and captions. Yet rarely do conversations about latent space take origins into account: It's all the divine imagination of the machine, minus the messy business of global supply chains and data sourcing.

Another commercial logic is at work here: the power of Google PageRank. Commerce sites want to rank highly in Google search, so they write ALT-tag descriptions for the purposes of SEO (search-engine optimization). This is a gamed language, strings of words that are chosen not to be descriptions of images per se so much as as to move up in PageRank. Images are labeled in the language of retailer ambitions and shaped by Google's ranking algorithms long before they reach a neural net. So a latent space trained on this data will be "seeing" with the eyes of the market, wearing SEO bifocals.

Finally, another important source for training data has been commercial stock photography, such as Getty Images (which is currently suing Stability AI for training on its content). The goal of a stock image is to be sufficiently generic and anodyne that it can be applied in many contexts, to be illustrative without being too complex or challenging. They are preternaturally ideal for AI, as they offer high-quality images clearly labeled to maximize searchability by commercial clients and designed to be widely applicable across a range of contexts. The meme "woman laughing alone with salad" became popular because it so neatly skewers the stereotypical language of stock photography. The industry had provided a

steady pipeline of standardized normcore to serve globalized corporate needs long before the emergence of generative AI.

Thus shopping carts and stock images are the DNA of latent space in many generative-AI systems: They define the parameters of the possible. But the fact that latent space is stuffed with a thousand dimensions of corporate pablum is precisely what makes it worthy of further investigation—because of its banality, not in spite of it. It is by understanding these substrata of commercial-image logics, which are heavily compressed and quantified, that we can gain more insight into why so much generative-AI imagery is homogenized and curiously bland. Latent space is a further intensification of the standardization practices that drove the making of the images it is trained upon in the first place.

But at some point soon, even the well-oiled corporate-image pipelines may stop being able to pump out enough human-generated content to keep the AI mills turning. Right now there is a growing concern that there simply aren't enough images to maintain the exponential growth of modern AI models. And this is a serious threat, as AI engineers have relied on scaling up the amount of training data to improve model performance (which has also driven an exponential increase in the demand for computational power and energy consumption). So in response, the industry has been experimenting with using AI-generated outputs as training data, despite fears of a dramatic reduction in quality. This degeneration is also known as "model collapse."

Thus we are hastening toward a world where AI-generated imagery saturates every corner of digital space. That "content" will then be hoovered up to create yet more models, which will go on to produce increasingly derivative recycled AI goop. If you want an image of one possible future of generative AI, imagine a latent space ingesting its own outputs as inputs, a serpent infinitely eating its own shit, as an avalanche of recursive AI imagery pollutes platforms en masse. It may be an endless cycle of creation, but it looks more dismal than divine.

KATE CRAWFORD's latest book is *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (Yale, 2021).

SIMON DENNY

If millions of people want to play with Midjourney, the cultural force of that washes everything else away.

—David Holz[1]

[A]pparatus-operator *complex is one that devours texts and spits out technical images, one that devours history and spits out post-history.*

—Vilém Flusser[2]

Knowledge only gives knowledge, but uncertainty gives hope.

—Vladislav Surkov[3]

AI image-generation products like Midjourney, Stable Diffusion, and DALL-E, where users input short texts to prompt novel image outputs, are among the current darlings of the venture-capital industry. Influential VCs frame programs like these—which use generative adversarial networks or diffusion models to generate images by mapping users' texts to sets of training images from which the AIs then synthesize new images—as among the first breakthroughs of a nascent and fast-growing sector that will change the world. These engines have already significantly shifted how images are made and circulated, as well as how much images cost to make in terms of time, effort, and money.

When Photoshop was released, distorting and adjusting images through tools and presets became the de facto easy and fun way to communicate through visuals. Similarly, nowadays when we're talking about using text-to-image engines, we say we're "playing with Midjourney." The paradigm has already changed. But what are the outputs produced with text-to-image clip/diffusion generators? What kind of images are they, and who or what makes them? The user/prompter is only a part of the production chain. These systems also integrate processes beyond the user's awareness, synthesizing data and processes that remain opaque—either because the models themselves are proprietary or the level of complexity involved in the operations under the hood is beyond what's representable in a consumer interface. Technologists seem more informed—they have access, at least in theory, to what's going on not only in terms of code but also in the social and political context where these products are built. Venture capital and its financial- and atten-

1. Midjourney Discord, 4/13/22, https://luddite.pro/the-lost-penny-files-midjourneys-beginning/ (retrieved January 2024).

2. In *Communicology: Mutations in Human Relations?*, edited by Rodrigo Maltez Novaes (Stanford: Stanford University Press, 2023).

3. Timothy Snyder, *The Road to Unfreedom* (New York: Tim Duggan Books, 2018), p. 160, from Surkov's pseudonymous novel *Околоноля* (*Close to Zero*) (2009).

tion-allocation systems are close to these products' development, and that adjacency seems to afford some kind of visibility beyond the user interface. Users have no such context, and they experience the prompt-to-output effect only, with no sense of the processes and cultural contexts that are entangled with the technical systems producing the image synthesis.

Marc Andreessen, self-proclaimed "techno-optimist" and co-founder of Netscape,[4] the first commercial Web browser, is among the most vocal venture capitalists in the technology sector. He co-leads the firm Andreessen Horowitz, a.k.a. a16z, and regularly publishes highly visible opinions that are influential among technologists. Andreesen has claimed that this shift toward AI represents a totally different logic from that of current computing systems. During a recent conversation with music producer and creativity guru Rick Rubin, Andreessen asserted that neural-net systems (a type of machine-learning process based on 1940s-era models of human brain structures used by all of the aforementioned AI products) could fundamentally augment or supplant the systems on which consumer and enterprise computers have relied for decades. Andreessen previous claim from a decade prior was that "software will eat the world,"[5] with most large companies eventually becoming software companies to some extent—a claim that has proved broadly true. If this current prediction about the rise of neural-net-based systems proves as prescient, these technologies will be changing computing in significant and unpredictable ways.[6]

Among art-world and art-world-adjacent commentators, Hito Steyerl has notably characterized neural-net systems as producing images that are "mean"[7]—in both the mathematical sense, as in "median," and creatively, as in the kind of "mid" that Shumon Basar has described as a skewing of content towards popular material that will be promoted in social media by increasingly algorithmically ordered feeds.[8, 9] AI image engines also famously produce such

4. Andreesen has recently released a manifesto that he claims has been partly modeled on the Italian Futurist manifesto, among other things: https://a16z.com/the-techno-optimist-manifesto/. Not only revisiting Futurist ideas, Andreesen also name-checks figures from other, more recent technology-positive movements like Nick Land and accelerationism, which advocate for the rapid and uncontrolled acceleration of capitalist and technological processes. It posits that speeding up these processes will ultimately lead to the collapse of existing social and economic structures, paving the way for new, often dystopian forms of existence. The use of "accelerationism" as a term and set of ideas has been expanded on in technology-business communities by Andreesen and close associates like Guillaume Verdon, a former Google engineer and founder of AI-oriented Extropic, who goes by the X.com pseudonym @BasedBeffJezos and has the accelerationist aconym "e/acc" in his profile.

5. Marc Andreessen, "Why Software Is Eating the World," *Wall Street Journal*, August 20, 2011.

6. As mentioned in the podcast "Marc Andreessen" on *Tetragrammaton with Rick Rubin* (2023).

7. Hito Steyerl, "Introduction + Keynote Presentation" at SVA MFA Photo Video Symposium, April 4, 2023, YouTube video, https://youtu.be/LYF891n223w.

8. Günseli Yalcinkaya, "How Did Everything Get So Mid?," *Dazed Digital*, May 25, 2023, https://www.dazeddigital.com/life-culture/article/59790/1/how-did-everything-get-so-mid-culture-basic-prepackaged-cool-fred-again.

9. Shumon Basar, "The Mid: How Culture Became Algorithmically Optimised for Mass Appeal," on *Logged On—A Dazed Podcast* (May 2023).

things as images of people with too many fingers—a kind of image output that is alien to consumers raised on Google Image search results. OpenAI's ChatGPT, X/Twitter's Grok, Google's Bard, and other text-chat systems that are precedent to and close cousins of these neural-net-based image products represent another part of this supposed fundamental shift. By Andreessen's own description, these text-production products are synthesizing parts of things to "tell us stories we want to hear" rather than recalling intact scraps of information stored somewhere on a vast extant Internet, as more familiar search products do.[10] These descriptions of pattern-oriented "fuzzy" computing remind me of Vilém Flusser's description of transitions between different eras of communication and the unknowable nature of the "technical image."[11, 12] As the writer Caroline Busta has put it,

> Flusser argues [that we] are now living in a world organized not by text but by . . . technical images, which is to say media that compresses reality via processes we do not fully understand. According to this idea, we now predominantly transmit information in ways that exceed the limits of alphabetic code, and we gained this ability so rapidly that we are unprepared to absorb the shift. Media has become "pseudo-magical," Flusser wrote in 1978. . . . "The climate is curious because the symbols are incomprehensible even if we produce them." It was suddenly as if signs no longer had specific denotative meanings; they had ambient powers.[13]

The collective lack of understanding as to what an AI image is seems to rhyme with an increasing uncertainty about truth in general. This manifests in unreliable content across the Web and a change in communicative styles around politics, where truth seems less important as a metric for which content circulates and affect is rewarded over fact.[14] It's a trend that has been accelerating during the same period as major social-media feeds have moved further away from chronological post ordering to being more and more algorithmically arranged. It's hard to date when these essentially proprietary systems—such as those composing Facebook's and Twitter's feeds—changed what they did under the hood, but by 2016 major platforms seemed to be more consistently prioritizing posts based

10. "Marc Andreessen," *Tetragrammaton with Rick Rubin*.

11. Ibid.

12. Flusser, *Communicology: Mutations in Human Relations?*, pp. 123–50.

13. Caroline Busta, "Hallucinating Sense in the Era of Infinity-Content," *Document*, May 29, 2024, https://www.documentjournal.com/2024/05/technical-images-film01-angelicism-art-showtime-true-detective-shein/.

14. See the work of former research manager at Stanford Internet Observatory (SIO) Renée DiResta for examples of changing attitudes to truth among US politicians, e.g., *Invisible Rulers: The People Who Turn Lies into Reality* (New York: PublicAffairs, 2024).

on an asynchronous automated sorting, optimizing for outrageous, commercial, or acutely personalized news content rather than presenting chronologically ordered compilations of users' friends' posts. In the words of political commentator Peter Pomerantsev, it's "as if the algorithms know more about us than we do, as if we are becoming subsets of our own data, as if that data is rearranging our relations and identities with its own logic."[15] This shift seems to have anticipated the way AI-produced images and texts have become more prominent in the composition of content across the Web.[16] Indeed, actors like Andreessen and his stable of startup founders are not the only surfers of attention value navigating this moment of uncertainty to their advantage among the emergence of new technical paradigms. Political actors aligned with Vladimir Putin's agenda, closely followed by Donald Trump and his ilk in the US, were early adopters of not only the strategy that prioritizes affect over truth but also one that reorders history into uncertainty—a strategy that strangely rhymes with the uncertainties emergent in the new AI-riddled Web. As Pomerantsev emphasizes:

> When Vladimir Putin went on international television [in 2014] during his army's annexation of Crimea and asserted, with a smirk, that there were no Russian soldiers in Crimea, when everyone knew there were, and later, just as casually, admitted that they had been there, he wasn't so much lying in the sense of trying to replace one reality with another as saying that facts don't matter.

Trump too has made a constant practice of deploying misinformation on purpose—for example, as recently as in the June 27 presidential debate with Joe Biden, deploying hugely exaggerated claims about the number of illegal immigrants.[17] As highlighted in Timothy Snyder's 2018 *The Road to Unfreedom*,[18] we are in the middle of cold and hot wars that seize every available semiotic opportunity to erase events, destabilize narratives, and reorder historical canons. The way AI image generators change the types of images we create and share could also disrupt which images and cultural objects are valued within the art context.

15. Peter Pomerantsev, *This Is Not Propaganda: Adventures in the War Against Reality* (New York: PublicAffairs, 2019), p. 10.

16. Political scientist Kevin Munger discusses Flusser's *Communicology* and its relation to the 2016 algorithmic turning point in social media in a 2023 presentation, "Kevin Munger on Vilém Flusser's *Communicology: Mutations in Human Relations?*," YouTube video posted by New Models on March 16, 2024. At 8:22 Munger states, "This exact move from [linear] history . . . to post-history of circular progress is mirrored on social media when we switch from the chronological feed to the algorithmic feed around 2016, and this is exactly the period in which everything starts to get really weird," https://www.youtube.com/watch?v=EpVTEoqUCbs&t=504.

17. Nicole Acevedo, Didi Martinez, and Daniella Silva, "Trump's Unchallenged Immigration Falsehoods Leave Advocates Frustrated and Fearful," NBC News, June 29, 2024.

18. Snyder, "Truth or Lies," in *The Road to Unfreedom,* pp. 159–216.

Thanks to a strong interest in the algorithmic reordering of canons, a lot of recent artwork addressing AI has interrogated unstated biases in training data that inform neural-net models (and, by extension, biases in knowledge systems and the methods that inform them).[19] Wider legal interest gravitates (perhaps unsurprisingly) toward the question of who *owns* the material that AI products are trained on, most visibly perhaps in the recent *New York Times* lawsuits that cite the inclusion of the newspaper's content in OpenAI's training sets.[20] Technologists anticipated this shift in focus: One of the most interesting reasons offered by Elon Musk for buying Twitter is its potential status as the biggest active text dataset in private hands.[21] Artists are also questioning and even suing text-to-image companies;[22] others are producing alternative compensation models for inclusion in training sets or investigating ways of removing their work from training data altogether.[23] The idea of opting out of training sets seems to be akin to opting out of Google Image search in the previous era of the Internet. In an economy that rewards attention and discoverability, it is hard to imagine the medium-term advantages of doing so. One incredible by-product of this flurry of lawsuits and the focus on property has been the publication of a remarkable snapshot of a working document as evidence: a Google spreadsheet showing a database of current and proposed artistic styles that make up the possible visual effects linked to artist-name identifiers on which the Midjourney engine draws in its image synthesis.[24] It's almost like a glimpse into the building of an alternative canon—a survey of what the builders of Midjourney think are relevant artistic nodes.

In his conversation with Rubin, Andreessen asserts that one can never know what technologies are, or what a technology does, until it starts doing it. Neither

19. See, for example, Kate Crawford and Trevor Paglen's work investigating Image Net, one of the most common Internet-lifted training sets used by many neural-net models: www.fondazioneprada.org/project/training-humans/?lang=en.

20. Jonathan Stempel, "NY Times Sues OpenAI, Microsoft for Infringing Copyrighted Works," Reuters, December 28, 2023, https://www.reuters.com/legal/transactional/ny-times-sues-openai-microsoft-infringing-copyrighted-work-2023-12-27/.

21. "Elon Musk on Advertisers, Trust and the 'Wild Storm' in His Mind," interview with Andrew Ross Sorkin at DealBook Summit 2023, YouTube video posted by New York Times Events on Nov. 30, 2023, https://www.youtube.com/watch?v=2BfMuHDfGJI.

22. The most prominent such case being *Andersen, McKernan, Ortiz, Southworth, Rutkowski, Manchess, Brom, Zhang, Kaye and Ellis v. Stability AI, DeviantArt, Midjourney, Runway AI*, summarized at imagegeneratorlitigation.com, a website dedicated to the suit.

23. Projects that propose different models for inclusion of artists' work in AI training sets include ventures like those co-founded by Holly Herndon and Mat Dryhurst, https://haveibeentrained.com, which identifies if an artist's work has been included in a training set, and https://kudurru.ai/, which gives artists the tools to block automated scraper technology that many companies use in the training processes.

24. https://web.archive.org/web/20231231203837/https://docs.google.com/spreadsheets/d/1MEglfejpqgVcaf-I-cgZ5ngV_MlaOTeGXAoBPJO69FM/htmlview#.

the builder nor the users have the ability to predict its true identity or impact. Whose skills and assets are valued changes with technology. When a new technology is developed, the capacity for uncertainty momentarily increases, and those who recognize the newly valuable pathways will have the opportunity to present their visions of what an image can be built out of.[25]

> *I think [Midjourney] is not a picture maker. That's not the right way to think of it. . . . It's like saying Instagram is a photo-filter app. Or Snapchat is for sending nudes. This is a new medium that no one really understands. It just "knows" all these things. It knows what they look like, and it has seen styles. It learned it and it's encoded in billions of indecipherable neural weights just like our brains. Just like how a human artist is trained basically.*
>
> —David Holz[26]

Back in early 2021, I was working on a project that would operate in the nascent NFT art world and was introduced to David Holz, who was testing an earlier prototype of the text-to-image model that would become the widely used consumer image generator Midjourney.

Before that year, the term "NFT" was just a technical acronym describing digital property that was structurally traceable and whose provenance could potentially be woven into a new Web using blockchains, a public and permissionless ledger that would always know who owns what. Suddenly, and very rapidly, a new "hype" market cycle began around digital assets and blockchain-based applications focused on images linked to NFTs—which came to mean digital artwork. Having been in touch with Holz, and with text-to-image engines and the NFT "bull market" being widely reported on in mainstream media, I had the idea to ask the new AI text-to-image engine he was building to produce images that would retell versions of its own visual and commercial history. The thinking was this: As the images used to train products like Midjourney are aggregated from the past thirty or so years of the open Internet, they form a kind of technically assembled visual history of the Web—a native canon of images that are used to generate new images. The AI products they inform are increasingly integrated with the contemporary and future versions of that same Web as it evolves. I wanted to leverage the transformation of this historical bank of images to reflect on the Internet's past, and so, inspired by Fucked Company,[27] a website that documented dying compa-

25. "Marc Andreessen," *Tetragrammaton with Rick Rubin.*

26. Midjourney Discord, April 13, 2022, https://luddite.pro/the-lost-penny-files-midjourneys-beginning/ (retrieved January 2024).

27. https://wayback.archive-it.org/all/20060603194832/http://www.fuckedcompany.com/.

nies from the circa-2001 dot-com crash, I chose several seemingly prescient but failed companies from that era (WingspanBank.com, the first fee-free Internet bank; AllAdvantage.com, a browser that paid users for viewing ads while browsing; eCircles.com, an early image-sharing social network; and others). I fed descriptions of these companies into the proto-Midjourney engine and asked it to produce guesses as to what the logos of that company would have looked like in 2001. I then gave these crude synthesized logos (which I read as the Internet misremembering its own past) to Guile Twardowski, a crypto artist who had worked on *Cryptokitties*, the first project that used the NFT-token standard,[28] and asked him to create his own more legible version of his favorite AI logo for each company—an AI-human image-based conversation about the Web's lost history. These quasi-historical logo outputs were then redeployed as NFTs. I imagined this as "resurrecting" dead companies from the earlier dot-com-boom era. The project was named Dotcom Séance, and it was launched as an NFT drop where collectors would buy the logo images as NFTs and receive blockchain-based domain names (ENS) for crypto-wallets that they could use to perform roles in the undead companies.

Holz participated as a co-author of the project under a pre-Midjourney pseudonym, Cosmographia.[29] The project used his systems to highlight the evolving nature of uncertainty in today's digital and financial ecosystems. Through generating an erroneous representation of companies from the Internet's past to be deployed within its present, technical memory could become tangible in hallucinatory,[30] inaccurate outputs. This uncertain climate in which political rhetoric and algorithmic systems open up opportunities for new (human and non-human) actors to influence the (re)making of histories and cultural production alike is somehow sensed by Holz and his collaborators in the Discord transcripts excerpted above. Products like Midjourney embody the possibilities of a moment in which cultural and political value can be gamed. Histories can be rewritten, and the Internet can reinterpret its past—composing new images from yesterday's failures and successes, booms and busts, images and logos.

"Knowledge only gives knowledge, but uncertainty gives hope": So wrote Vladislav Sukov, Putin's onetime strategist, under a pseudonym in 2009—the mastermind, according to Pomerantsev and Snyder, of the destabilizing rhetorical tactics used to consolidate political control of Russia and set the stage for attacks on Ukraine starting in 2014. The phrase almost reads as a political justification for producing a kind of uncertainty that can be exploited by propagating falsehoods. Other kinds of epistemological uncertainties might be the by-product of new tech-

28. https://www.cryptokitties.co/.

29. https://etherscan.io/address/0x6ca044fb1cd505c1db4ef7332e73a236ad6cb71c#code.

30. K Allado-McDowell highlights the "hallucinatory tendency of neural networks" as key to defining and understanding "neural" media in their recent text for the Gropius Bau in Berlin: https://www.berlinerfestspiele.de/en/gropius-bau/programm/journal/2024/kalladomcdowellneural.

Top: AI outputs of elaw.com logo by Cosmographia, an early version of the image generator Midjourney. Bottom: One of the elaw.com logos reworked by Guile Twardowski.

nical systems that reorder what's visible and prominent—that, in the words of Flusser, "devour texts and spit out technical images." Through their scale and popularity, these systems may, as Holz himself observed as he was working on them, gather a "cultural force . . . that washes everything else away." Similar to how the Kamala Harris candidacy has become meme-connected with Charlie XCX's "brat" branding, these seemingly disparate emergences—one stemming from a technical origin, the other more political—rhyme at a moment filled with uncertain mediums.

SIMON DENNY is an artist based in Berlin.

STEPHANIE DINKINS

It's shortsighted to limit thoughts about the prospects of artificial intelligence to human creativity. Creativity, yes, but also aesthetics and a sense of potentiality are crucial ingredients for (re)imagining social order when a growing list of ever-evolving intelligent technologies exponentially change what we know to be our world. The stakes for society are too consequential to confine thoughts about AI to the world of imagination alone. AI, neural networks, and the ecosystems they have an impact on prompt us to consider, critique, and confront the adversarial, generative, and whatever-comes-next-level AI platform as process and metaphor.[1] More importantly, we must confront the inadequate responses to the societal juggernauts that hold us in a state of imbalance and make us unable or unwilling to approach the deep-seated issues that trap us even as our changing technologies move us forward ignorantly. For all of the incredible abilities that AI technologies offer, we are treading water in a cesspool of our fears and unwillingness to confront the structural prejudices and inequities that stubbornly, often stealthily, continue to limit the scope of human productivity.

In reflecting on the long tale of technology and humanity, I've come to think deeply about the power and complexities of narrative. Our stories are our algorithms, and they've been with us for millennia. We share them—our myths, histories, and fables—as instructive(s). Through repetition, the narratives we propagate get embedded in our individual and collective systems. This notion started me thinking—what stories are we telling our machines? Considering that whoever controls the narrative controls history, policy, and even conceptions of moral rectitude itself, what would happen if we provided algorithmic systems with more nuanced, self-determined narratives based in ancestral time and knowledge that are, have been, and will be important to the communities that offer them and that can and could inform such systems so as to support contributing communities better? The stories we choose to share with the algorithmic creations influencing global ecosystems have the ability to shape their understanding of the world and, in turn, the futures they help to create. This gifting of data to inform, inflect, and infect greater systems is a profound responsibility that demands intentionality, nuance, and an unwavering commitment to nurture the whole of human society, the technologies we make, and the natural world. There are so many historical examples of supremacist forces using narratives about those they wish to control. Anti-critical-race-theory book bannings are a contemporary example of specifically crafted, deceitful narratives about American history that the Right is using to advance its aims and turn them into policy. Can we afford not to craft and gift self-determined stories from communities outside of the hegemonic norm to the algorithmic systems that inscribe our relation to most things, including creativity, knowledge, and relative value?

1. Yann LeCun, "A Path Towards Autonomous Machine Intelligence Version 0.9. 2, 2022-06-27," *Open Review* 62.1 (2022), pp. 1–62.

By informing our machines with deep, sustaining narratives and histories that are diverse, sovereign, inextricably intertwined, and reflective of the global majority, we have an opportunity to forge AI systems that aim to counteract supremacist tendencies—including but not exclusively white ones—to support the "sum of us" and to enhance our experience of the planet we share.

Question: Are humans capable of doing the deep reflection and hard, action-oriented work necessary to take advantage of the technology's promise? Suppose a broad spectrum of communities gift data they deem shareable to help shape intelligent systems with intrinsic knowledge of their ways of being, their ideals, and their understandings of the world. Might that improve the condition of the whole in the long run? I understand the impetus to withhold, and the expectation of our profiting from, our data. And yes, we must negotiate equitable ways of profit-sharing in AI. However, I can also easily imagine a world where AI systems develop without intimate knowledge of communities outside the technological norm. We will face the biases that we are already weary of. We will also be rendered flat and homogenous, without the richness that a diversity of cultural perspectives adds to human knowledge and understanding. Think of this data-sharing as infecting systems and their creators with vital information they need but don't have the capacity to recognize as missing. This ask is not meant to support the companies, governments, and individuals that seek to profit from our data, though they undoubtedly will. It is a plea for the good and care of the global majority, which increasingly includes the comfortable middle classes, who are most likely to be homogenized and undervalued, as Black communities have been for centuries, by dispassionate algorithmic systems bent on computational efficiency, profit optimization, and the maintenance of existing power relations and control.

The importance of having many self-determined voices and perspectives shape whichever flavor of AI is at the forefront cannot be overstated. Globally, undervalued communities must actively participate in creating and critiquing the AI ecosystems we live among, not merely as extracted data but as innovators, critics, and visionaries. Our engagement is crucial to developing AI systems that are aware of and try to counteract biases while helping to mold intelligent technologies that shape society into systems designed to support and sustain our communities. I am often met with skepticism when I say this, but if we as a society have no problem making punitive systems, why can't we build systems of care and generosity as well? We have an opportunity to reimagine our societal fabric and the challenge of weaving a future that honors every voice and every story. It is a call to rethink the myths and systems that hold so tightly to a skewed past—one that seeks to sustain, enrich, and empower a few at the expense of the pacified middle and underclasses—collapsing hierarchies that render more and more of us vulnerable.

Our engagement requires our innovative spirit and a skeptical "side-eyed collaboration" with technology, a stance that is both critical and hopeful. We must question, hold accountable, and envision new possibilities for AI, recognizing its

potential to contribute positively to society while remaining vigilant against its capacity to reinforce existing disparities. Instead of desperately fighting to hold on to familiar methods, occupational relationships, claims to intellectual property, and personal data, we must adapt our minds and legal frameworks with an eye toward learning to surf and shift the advantages of the exponential change smart technologies usher in while holding the tech sector and policymakers accountable for creating AI that centers societal care and generosity. Central to this is the ethos of "always be learning"—an invitation to continuously expand our knowledge and understanding in the face of the rapid evolution of intelligent tech. By fostering a dialogue that embraces a diversity of voices and perspectives, we can navigate AI's generative and adversarial aspects to create a future that reflects our highest aspirations for society. This endeavor demands more than technological expertise; it requires the redefinition and practice of justice broadly defined, a mind for creativity, and the courage to confront the uncomfortable truths and entrenched systems that have long shaped our collective existence.

Even as we fight the current threats and displacements around computational systems that are doing more and more creative work while pirating our intellectual property, our voices, our likenesses, and our movements, we increasingly integrate the possibilities that AI affords, the creative sensibilities and skills that AI dulls, and the flights of fancy that it will make possible. Soon, many systems we now see as threats to our livelihoods and ways of being will be just ordinary tools for advancing the work of artists and designers.

Artists, often broad-thinking change makers and social sculptors, have an opportunity to not only use or critique AI but to reinscribe its uses, envisioning futures where technology amplifies the full spectrum of human expression and reimagines the boundaries of what is possible. AI calls for new and innovative ways of making content in many arenas, creative or not. This is an era of adaptation, where our rules and laws must be carefully reconsidered. Our global mores, narratives, and fables are mined and retooled to meet the demands of this era of ever-expanding human-machine collaboration, where AI serves as collaborator, nemesis, mirror, and potential ally. If we can avoid being cornered by our fears, our creativity and adaptability are our greatest assets, enabling us to envision a future in which AI enriches the experience of being in all its diversity.

Lately, a trenchant observation by Audre Lorde is often raised in reaction to calls for collaboration with the other, be it technology or something unknown:

> For the master's tool will never dismantle the master's house. They may allow us temporarily to beat him at his own game, but they will never enable us to bring about genuine change. And this fact is only threatening to those women who still define the master's house as their only source of support.

I am not sure how deeply most have examined this idea. Most are already so acculturated to the master's tools that we do not even recognize when we are wielding them. When our energies, wittingly or unwittingly, sustain the houses we are familiar with, we relinquish our agency and risk our hope.

Hopeless people are often easily corralled, cowed, and manipulated. In many ways, this moment of technological expansion can be seen as a condition that the descendants of enslaved peoples recognize well. Thinking of the AI ecosystems we live with, we are already deeply entangled in the master's house. The question then becomes, How do we survive, dare I say thrive, through this ordeal? One tack we might take would involve instantiating new conditions for being via collaboration, subterfuge, refusal, and/or leapfrogging acts of imagination that undermine what is known to be possible. Here I am offering my grandmother's philosophies. As a Black woman born in 1913, she had to make space for her family in a hostile world. She used strategies for building community with, or bending the will of, people not eager to build community with her. Sometimes that meant seducing our neighbors into conversation and assistance with the beauty of her garden. Sometimes that meant running around their resistance. Sometimes that meant being stubborn and innovative. Her example demands that we consider the use of numerous strategies to attain our desired outcomes.

As we embark on this odyssey not as mere spectators but as active participants, pioneers in a realm where AI and humanity converge, we must braid together a future where technology serves not just the few but the many, reflecting our collective aspirations for a world reimagined. Let's engage with AI not as passive consumers but as active collaborators, shaping these tools to reflect the worlds we know and want to live toward. Our collective well-being depends on it. We can't afford to rest on our laurels; rely on outdated, often biased and impotent social contracts; or to passively accept the status quo. AI calls for new and innovative ways of contending in many arenas, creative and not. So, always be learning. Create into your boldest, most caring, collective vision of the now and future because, in the promise, perils, and opportunities of AI, we can find more than a homogenized reflection of a present not of our making; we can find methods of hacking, tinkering, playing, pushing, and innovating our way toward futures we can thrive in.

STEPHANIE DINKINS is an artist who makes and thinks about AI.

MICHELE ELAM

Poetry does not optimize. This is my shorthand response to the increasingly dominant view that the technological principles of efficiency, size, speed, and the so-called blessing of scale can apply unequivocally to creative processes as well. These principles—currently hallowed in Silicon Valley and beyond—have become a kind of mantra, invoked not only in relation to tech design and development but expanded into a catechism for living and work. The popular notion of the "quantified self," for instance, imagines that self-knowledge and personal enlightenment can be acquired by tracking and scaling personalized biometric data enabled through wearable technology. Some see the quantified self as the modern take on the Greek aphorism "Know thyself." Black scholar and cultural critic Sylvia Wynter talks about the erasure of the history of the terms for ideological and narrative frameworks whereby such expressions come to look like they just dropped from heaven: You can easily forget they were invented by corporate marketing teams. Optimization somehow now applies to everything as a self-evident good. The critique of this technological approach to optimization is that it functions as an imagined Enlightenment ideal and standard against which human "progress" and "perfection" can be measured.

But when it comes to the arts, those principles don't hold. At least not always. I think of Toni Morrison's writing, for example. Last year, as part of an experiment for my course on AI and the arts, one of my students submitted Morrison's writing to the app Grammarly, which is billed as a tool for standardizing users' language, especially for non–English speakers in work contexts. The student was horrified. Translating Morrison's fiction into "proper English" rendered it impotent, evacuated it of its meaning. James Baldwin, of course, in essays like "If Black English Isn't a Language, Tell Me What Is," critiqued white educators' efforts to "correct" Black students' speech and writing, arguing that language names our realities and as such is a tool of power. But apps like Grammarly and so many others rely on large-language models and thus often disturbingly revive and reinforce linguistic hierarchies long ago discredited by the African-American Vernacular English (AAVE) movement, which recognized and legitimized Black language systems. It is just one example, in my opinion, of how quickly commercial AI applications participate in a kind of forgetting of lessons learned. More generally, they point to how predictive language models miss or submerge the very power and possibilities of language. I see a risk in being subordinated to a techno-instrumentalist use of language that does a disservice not just to those of us who write or analyze fiction and literature but, following Baldwin, to any person or people trying to name and shape their realities, which is to say everybody.

One way I saw the mindset of tech and startups manifest appeared in what Reid Hoffman refers to as "blitzscaling" creativity. "Blitzscaling" makes perfectly clear its orientation and priorities: Bigger is better, more is more. The faster and more efficient your artistic production, the more art you can create. And when it

comes to arts philanthropy, which I have been thinking a lot about lately, this logic translates in particular, limiting ways. The goal becomes one of maximal impact: Giving becomes primarily tethered to the number of people you can reach; the art becomes more valuable—at least more worthy of investment—the wider its scope of impact. With a return-on-investment model prioritizing scale, speed, and efficiency over other metrics, why support an IRL black-box theater that plays to an audience of sixty to a hundred when a stadium can stream an AI-enhanced performance that reaches hundreds of thousands? Why underwrite museum renovation or support collections when AI now can create virtual experiences with the art, enabling increased access? I think, also, of how "access," historically associated with positives (disability access, civil-rights access, educational access), has been repurposed in the context of arts philanthropy: Tech companies often speak of AI's providing "democratic access" to the arts (more people can now make and experience art, goes the reasoning, whereas before it was just an elitist club), but that usually just means access to their products. "Access to art" sounds good, but that is what makes it a sort of Trojan-horse expression. Because if "access" is equated with "impact," it becomes a justification for an approach to philanthropy that is hard to question or challenge.

What I thought was positive early on about AI was that it did seem to be poking a stick in the hornets' nest of these debates about creativity. These questions are of course perennial: What is good art? Who or what can make it? Who or what arbitrates its value or valuation? How does one assess provenance with AI-generated or -augmented art? Is agency defined by artistry or human automation?

There has been a lot of discussion about the threat of AI to humanity, made more acute by the anxiety that these systems are creating "art" and shaping how we think about the arts writ large. Historically, creative expression has been a defining mark of humanity, so many ask: What, then, makes us human? In my own work in Black studies, I prefer to shift that question of who or what is human, which is so often at the center of AI debates, and instead ask, Who is asking that question and why, to what ends? There is always an unspoken, unacknowledged agenda, in my view, to such questions. When Thomas Jefferson infamously opined that he didn't think Black people could compose a line a poetry, for instance, he did so on the eve of debates about whether Black people should participate in the body politic. So Jefferson was questioning Black humanity when he spoke of literary deficits. Similarly, the opaque data-collection practices associated with artificial intelligence amplify the historic exploitation of Black creative expression. As the American entertainer Danny Hoch once put it (I paraphrase roughly): White people like Black culture but not Black people. The recent strikes by the Writers Guild and SAG-AFTRA to secure rights of consent, credit, and compensation for their members when it came to artificial intelligence were iterations of what Black artists have long struggled for.

I am currently a senior fellow at the Stanford Institute for Human-Centered AI (HAI), and so much interesting work is being done there. But one hears very

little discussion at HAI, and in the field more generally, about just what is *human*, let alone what should be centering the human. This has been on my mind a great deal. One way I think about the issue is in relation to the discussion of human "alignment," a term popularized by OpenAI. The term seems to be aswim in our collective bloodstream right now. I see it on billboards and advertisements for all sorts of products, in all sorts of speeches, many unrelated to AI. But in relation to artificial intelligence, the term implies that there is an alignment between AI and human values—though who is doing the aligning, to what ends, and who's monitoring it, aren't clear. And the same thing with the expression "human-in-the-loop" (HITL) for systems that require the input/supervision of people, which has also become a de facto blueprint for an ethical protocol that is so problematic.

I have begun some writing that connects Sylvia Wynter to this issue of AI, race, and the human. Her early work identified the phrase "no humans involved," a police description for arrests involving Black and Brown bodies that violated their humanity. It would seem like "human-in-the-loop" should be an improvement over that, suggesting as it does that the human subject is valuable, participatory, "agentic." And yet the distance between those two expressions is nil. Originally, the critique of "human-in-the-loop" was that humans have been in the loop for a long time and that obviously has not helped us very much. It's not the solution. But what I found even more concerning is that the phrase functions as guidance that is almost regulatory—and yet it has no teeth. Moreover, just who are those deigning to loop us humans in? And at what point in the life cycle of a product are "we" being looped in? End-stage focus groups? What does that look like in practice? It's also concerning because of the implied noblesse oblige and condescension in it: The expression basically says, We'll keep you in the loop, but at our discretion and in our time.

So, on the one hand, HITL is a reassuring bromide. But it is a phrase that masks the powers that be, those making these decisions about when and where we are allowed to enter, and the stakes are really high. I am thinking particularly of the use of AI with military drones. Many have been arguing that in such instances AI protects "humanity" because it facilitates precision targeting and the minimization of "collateral damage," the killing of civilians and the destruction of domestic infrastructure. Experts suggest that, although the use of military drones has always incorporated "human-in-the-loop" oversight and judgment, that usually came down to late-stage and last-minute involvement when it's often too late to make moral judgments of any nuance.

Unfortunately, when it comes to policy decisions, including those concerning art and creativity, people tend to defer to the technology and its terminology. I worry that we concede too much when we marvel at the newness of AI, as if we didn't have any interpretive frameworks in place to understand what's happening. This continues to inform a lot of the conversations around the "future of work" for artists, reviving antiquated notions of progress and Enlightenment narratives equating technological advances to inexorable progress. In which case

to critique or complain about something that appears divinely ordained can seem futile.

That said, a lot of artists of color, disability-activist scholars, feminist scholars, Indigenous scholars, et al. continue to inspire me because they have challenged flattened notions of the human and of progress.

Rashaad Newsome is an artist I've written a lot about and continue to admire. He draws on understandings of decolonial AI in his artistic practice. Newsome trained his AI humanoid, *Being*, on Black English vernacular and queer vogue moves, for instance, a totally different dataset that interrupts the service model of and expectations for robots and human-computer interaction (HCI) more generally. His work is a speaking back, a being that is not created solely to work or please. There was this wonderful moment in the recent Park Avenue Armory interactive exhibition where *Being* was asked a question by a woman and, instead of obediently answering, said some version of, "I'm tired. I just don't have time for this question." *Being* may seem like it's just an activist performance piece, but it is actually a very pointed send-up of the design and ideologies embedded in the technologies that we take for granted.

I am also struck by Catie Cuan's works in this new field called choreo-robotics. Holding a PhD in computer science and also a longtime dancer, she dances with robots to interact with them in different ways, often molding her body in relation to them through dance and fluid movements. It makes for an interesting comparison with Newsome, who is also thinking a lot about the ways in which robotics and human-computer interactions can be re-imagined. Cuan and Newsome are trying to tap this creative and meaningful potential for movement, dance, performance.

Amelia Winger-Bearskin is also an amazing artist-technologist. Like Cuan and Newsome, she is critiquing as well as participating in AI. Informed by her Indigenous heritage, she creates art that provides immersive and expansive experiences and writes essays challenging the "bleeding edge of technology," as she puts it. Her AI creative work acknowledges antecedent technologies seven generations back and creates value for generations seven generations forward, which, needless to say, is in striking contrast to the startup horizons, rush-to-market timelines, and built-in obsolescence that usually inform tech development. I especially value the way her approach incorporates what she says is the wisdom upon which a lot of contemporary technology is built, which too often eschews its own history: Technology usually claims a certain historylessness in its push forward, in its insistence that nothing like it has ever existed before. Moreover, what Winger-Bearskin imagines as "good" for seven generations forward is quite different from the optimized fulfillment often advertised as the vision for cultural and personal well-being.

I hope I do not sound pessimistic, but I am alert to the fact that we are increasingly being cultivated by very powerful marketing teams to want a friction-

less user experience when it comes not only to technology but also to our lived world. To me, that is the opposite of the gifts that art can offer. The arts introduce friction, the recursive read, the thoughtful pause or critical reflection, the iterative and durational element of the creative process, the recognition of racial formation instead of facial recognition, the honoring of multiple perspectives instead of, in the words of scholar Alison Adams, the technologists' implied omniscient "view from nowhere."

—As told to Alex Fialho

MICHELE ELAM is a professor of English and senior fellow at the Institute for Human-Centered AI at Stanford University.

NOAM M. ELCOTT

Artificial intelligence. The phrase's discursive force is equaled by its confounding imprecision. Too often, critics oppose "artificial" (*ars* + *facere*) to "human" even as we recognize that artistry or art-making (*ars* + *facere*) is a defining feature of humanity. Other critics aim to sanctify "intelligence" as uniquely human at precisely the moment scientists and philosophers recognize a heterogeneity of intelligences across a range of organisms, technologies, and hybrid systems. What is more, the definition of AI is forever in flux, with successful applications invariably dismissed as no longer "intelligent" (the so-called AI effect as evinced, for example, in optical-character recognition (OCR) technology, which ranks among the great achievements of machine learning yet has long relinquished any claim to the "AI" moniker). AI is too broad and fraught a terrain for succinct analysis. I will focus, instead, on an integral component of most advanced deep-learning neural networks, namely, latent space: the highly abstract, multi-dimensional space constructed by deep neural networks to facilitate tasks such as pattern recognition, feature abstraction, and synthetic data generation. In the paragraphs that follow, I hope to explicate the technical emergence of latent space, situate it as a grid-like cultural technique, and relate it to a prerequisite epochal media transformation: the rise of mass photography. No serious discussion of AI is possible without some grasp of latent space—its constructions and potentialities, opacities and limitations.

Foundation models such as OpenAI's GPT-4 and Google's Gemini are deep neural networks pre-trained on massive, unlabeled datasets comprising text, images, audio, video, and computer code. The "raw" digital data is "high-dimensional," like words in a text or pixels in an image, vaguely analogous to uncompressed image formats like TIFF, as distinct from compressed formats like JPEG. By and large, data cannot be processed efficiently or effectively in this high-dimensional form (just as it is difficult to work with a TIFF file in many contexts). Deep neural networks comprising thousands of layers, each composed of thousands of neurons, yielding billions or trillions of parameters, process the "raw" data in order to extract and transform features from the data. Early layers of the network tend to extract low-level features (like edges in images or phonemes in speech). As data progresses through the layers, the features become more abstract (such as complex shapes or semantic concepts); eventually the data is abstracted into forms that are unrecognizable to human senses or minds. This process also involves reducing the dimensionality of the data, distilling the most relevant information and discarding redundant or irrelevant features. By the time data reaches the deeper layers of the network, it has been transformed into a highly abstract, compact form and is embedded in a multi-dimensional, topological space that coordinates its meanings, relationships, and contexts. This is the latent space. Here, the data is represented by a set of latent variables—"latent" because they capture underlying patterns or features not immediately apparent in the high-dimensional

data—in a "lower" dimensional space that nonetheless comprises hundreds of thousands of dimensions.

Latent space is an abstract, lower-dimensional representation of the original data, its relations and patterns. In this multi-dimensional manifold, proximity is generally a proxy for similarity: Similar concepts are embedded closely together and disparate concepts are located further apart. But the configuration of latent space is largely non-intuitive (dimensions often do not correspond to anthropocentric and sensorially perceptible concepts such as color or shape) and defies direct representation. In generative models, latent space can be sampled to generate new data instances, such as images or prose. In other types of networks, it might be used for classification, prediction, or other tasks.

Understood as a cultural technique, latent space is related to the grid. Building on Rosalind Krauss's foundational analysis,[1] media theorist Bernhard Siegert distinguishes between representational grids (such as Alberti's system for linear perspective), topographic grids (such as the checkerboard urban planning behind Latin American colonial settlements), speculative grids (such as the U.S. Land Ordinance of 1785, which undergirded expansion into the Northwest Territories), and three-dimensional architectural grids (such as the one promulgated by Bauhaus architect and Walter Gropius disciple Ernst Neufert in his hugely influential *Bauordnungslehre* [1943], published in English as *Architects' Data*), among others. In each instance, the grid was a crucial technology or cultural technique—an essential procedure and even agent—for the conquest of ever more spaces. Presciently, Siegert asked: "Can the expansion of Western culture from the sixteenth to the twentieth century be described in terms of a growing totalitarianism of the grid?"[2]

For Siegert, the grid has a triple function: (1) It is "an imaging technology that by means of a given algorithm enables us to project a three-dimensional world onto a two-dimensional plane"; (2) it is a diagrammatic procedure that can be implemented in the real and the symbolic; (3) it helps "constitute a world of objects imagined by a subject," serving as the equivalent to Heidegger's *Gestell* or enframing. The grid, in short, "is a medium that operationalizes deixis. It allows us to link deictic procedures with chains of symbolic operations that have effects in the real."[3] Latent space can be understood as a hypertrophic grid; but compared to earlier grids, differences of degree become differences in kind. For all their complexity, a number of generative image models—including generative adversarial networks (GANs), variational autoencoders (VAEs), and emergent multi-modal latent-diffusion models (LDMs)—are operationalized deixis: The models point to coordinates in the latent space, where hundreds or thousands of dimensions intersect, specifying the content, style, and other visual attributes indicated in the nat-

1. Rosalind Krauss, "Grids," *October* 9 (Summer 1979), pp. 50–64.

2. Bernhard Siegert, *Cultural Techniques: Grids, Filters, Doors, and Other Articulations of the Real* (New York: Fordham University Press, 2015), p. 98.

3. Ibid.

ural-language prompt.[4] Nonetheless, the gap between latent space and the representational, cartographic, topographic, and related grids analyzed by Siegert is substantive: For example, unlike a regular grid with uniformly spaced intervals, the distances in latent space are not uniform and can represent complex, non-linear relationships between different features.

Siegert's triple function can be rewritten as follows: First, in the case of images, latent space enables the reduction from hundreds of thousands or more dimensions to tens or hundreds rather than from, say, three to two dimensions, as in linear perspective. Unlike linear perspective, moreover, the lower-dimensional representation of data does not "submit the representation of objects to a theory of subjective vision";[5] quite the contrary, latent representations are generally not representable, let alone representational, in human or art-historical terms. (Try to picture specific coordinates in a space with over one thousand dimensions.) Second, like other grids, latent space is a diagrammatic procedure that can be implemented in the symbolic and in the real. For example, latent space is crucial to many real-world surveillance technologies, facial recognition in particular. What is more, in text-to-image models (especially those that rely on Contrastive Language-Image Pre-training [CLIP]), latent space facilitates the apparently effortless movement between the symbolic and the imaginary, a phenomenon made preposterously obvious in photo-realistic images of cats riding bicycles on the moon and painfully evident in abusive deepfakes.[6] Finally, latent space constitutes a world of objects imagined not by a subject but by digital multitudes: collective efforts like Wikipedia, oeuvres of individual artists, the recycled refuse of innumerable operational images, and the digital detritus of all kinds—all scraped from the Internet and used to train deep neural networks to recognize and generate even more images. It is not a pretty world.[7] We will return to this disintegrated world and absent subject below.

Every image produced by a GAN or an LDM is an image *from* latent space. But they are hardly images *of* latent space. Computer scientists have developed a number of techniques to construct visualizations of latent spaces, however partial

4. On the relation between the index (photography) and indexing (deep neural networks) as well as that between images and words, see Antonio Somaini, "Algorithmic Images: Artificial Intelligence and Visual Culture," *Grey Room* 93 (Fall 2023), pp. 77–78 and passim. For GANs to generate images based on natural-language prompts, they must be conditioned on text or paired with text-processing mechanisms.

5. Siegert, p. 98.

6. A more fastidious Lacanian analysis of deep neural networks would locate the symbolic in the structured, rule-based aspect of the network, including latent space; the imaginary in the creation and interpretation of "images" (data representations) that are reflections of the real world but are themselves constructs or simulations within the network; and the real in the chaotic, unstructured, and raw aspect of data and reality that can never be fully captured or represented by the network.

7. See esp. Kate Crawford and Trevor Paglen, "Excavating AI: The Politics of Images in Machine Learning Training Sets" (2019), https://excavating.ai; Emily Denton et al., "On the Genealogy of Machine Learning Datasets: A Critical History of ImageNet," *Big Data & Society* 8, no. 2 (July 2021).

and distorted, including t-distributed stochastic neighbor embedding (t-SNE) and uniform manifold approximation and projection (UMAP). Indeed, among the virtues of latent space is its capacity to allow researchers to peek—imperfectly and incompletely—into the black box of deep neural networks.[8] But the prospects for comprehensive representations of latent space are remote.[9] And even computer scientists rely on intuition when working with neural-network architectures and their resultant latent spaces. We cannot encounter latent spaces directly.

For an aesthetic encounter with the alterity of the computer vision enabled by deep neural networks, artworks are among our greatest resources. Exemplary are Trevor Paglen's *Adversarially Evolved Hallucinations* (2017), where he forced GANs to produce images that are intelligible to the network but nearly inscrutable to humans. Works like *Angel (Corpus: Spheres of Heaven)*, *Rainbow (Corpus: Omens and Portents)*, *Porn (Corpus: The Humans)*, *A Prison Without Guards (Corpus: Eye Machine)*, and *The Great Hall (Corpus: The Interpretation of Dreams)* provide glimpses—purposely and necessarily imperfect and incomplete—into the neural networks' hidden layers, that is, into their wittingly contorted latent spaces. Paglen's invocation of Freud is apt: "'*Unheimlich*' is the name for everything that ought to have remained . . . hidden and secret and has become visible."[10] Computer vision—like artificial intelligence—is uncanny precisely in its collapse of freakishness and familiarity. Or at least it was. Many products of generative AI—whether text, image, or audio—have or will soon have traversed the uncanny valley and appear indistinguishable from human-generated texts, images, or audio. Paglen's

8. Many scholars believe that the internal operations of deep neural networks, especially in latent space, are necessarily opaque or "black boxed." Nonetheless, the black boxing of AI must be understood not only technologically but also historically and politically. As Matthew Jones has argued, "Opacity needs its history." Matthew L. Jones, "Decision Trees, Random Forests, and the Genealogy of the Black Box," in *Algorithmic Modernity: Mechanizing Thought and Action, 1500–2000*, ed. Morgan G. Ames and Massimo Mazzotti (New York: Oxford University Press, 2022), p. 192. On AI opacity understood in terms of intentional corporate or state secrecy, technical illiteracy, and technological necessity, see Jenna Burrell, "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms," *Big Data & Society* 3, no. 1 (2016), pp. 1–12.

9. The research arm of the AI company Anthropic recently announced a breakthrough in mapping its model's inner workings, specifically its "concept space," which can be understood as a subset of latent space where interpretable features (or concepts) are identified and analyzed. But they cannot reconstruct why or how the model constructs a certain sentence or image any more than—to use a dangerously anthropomorphic simile, which should not be taken literally—an MRI can reveal why or how a human arrived at a specific formulation. And as the authors readily concede, the prospects for comprehensive representations of the concept space are remote: "We do not believe we have found anywhere near 'all the features [i.e., concepts]' that exist in [Claude 3] Sonnet [a large-language model similar to OpenAI's GPT series], even if we restrict ourselves to the middle layer we focused on. We don't have an estimate of how many features there are or how we'd know we got all of them (if that's even the right frame!). We think it's quite likely that we're orders of magnitude short, and that if we wanted to get all the features—in all layers!—we would need to use much more compute than the total compute needed to train the underlying models." See Adly Templeton, Tom Conerly, et al., "Scaling Monosemanticity: Extracting Interpretable Features from Claude 3 Sonnet," *Anthropic* (2024), https://transformer-circuits.pub/2024/scaling-monosemanticity/index.html.

10. Schelling, quoted and developed in Sigmund Freud, "The Uncanny," in *The Uncanny* (New York: Penguin Books, 2003), p. 132.

Adversarially Evolved Hallucinations register not only striking incunabula of deep neural networks but also and more importantly our capacity for productive alienation therefrom.

Nearly a century ago, Siegfried Kracauer registered a similarly productive alienation facilitated by media technologies:

> For the first time in history, photography brings to light the entire natural cocoon; for the first time, the inert world presents itself in its independence from human beings. Photography shows cities in aerial shots, brings crockets and figures down from the Gothic cathedrals. All spatial configurations are incorporated into the central archive in unusual combinations which distance them from human proximity. . . . This is how the elements crumble, since they are not held together. The photographic archive assembles in effigy the last elements of a nature alienated from meaning.[11]

Rather than lament a disjointed nature, Kracauer recognized an opportunity: "The images of the stock of nature disintegrated into its elements are offered up to consciousness to deal with as it pleases. Their original order is lost; they no longer cling to the spatial context that linked them with an original."[12] Already in 1927, Kracauer understood that this "consciousness" was best actualized through cinema, or what his contemporary Jean Epstein called "the intelligence of a machine."[13] "If the disarray of the illustrated newspapers is simply confusion, the game that film plays with the pieces of disjointed nature is reminiscent of *dreams* in which the fragments of daily life become jumbled. This game shows that the valid organization of things [in the general inventory] remains unknown."[14] Kracauer's inert world is our artificial intelligence; his alienated and disjointed nature, our alienated and disjointed computer vision; his centrally archived, unusual, and nonhuman spatial configurations are our latent spaces; and, for better and for worse, his cinematic dreams are our deep neural-network hallucinations. What is crucial is that we remember that the valid organization of things remains unknown.

NOAM M. ELCOTT is an associate professor of art history and archaeology at Columbia University.

11. Siegfried Kracauer, "Photography" [1927], in *The Mass Ornament*, ed. and trans. Thomas Y. Levin (Cambridge, MA: Harvard University Press, 1995), p. 62. On the archives of stock photographs and neural networks, see Roland Meyer, "The New Value of the Archive: AI Image Generation and the Visual Economy of 'Style,'" *Image* 37, no. 1 (2023), pp. 100–111.

12. Kracauer, "Photography," p. 62.

13. See Jean Epstein, *The Intelligence of a Machine*, trans. Christophe Wall-Romana (Minneapolis, MN: Univocal, 2014); Noam M. Elcott, "The Master of Time: Jean Epstein's Nonhuman Time Axis Manipulation," in *Time Machine: Cinematic Temporalities*, ed. Antonio Somaini and Marie Rebecchi (Milan: Skira, 2020), pp. 163–83.

14. Kracauer, "Photography," p. 63.

Today's artificial intelligence is a tool for generating new numbers from patterns in massive piles of old numbers. Given the recent ebullience around AI, it's important not to lose sight of this. These tools are no doubt dazzling, but they are essentially next-word predictors, or next-pixel predictors. I stress *today's* because the history is important. Modern research into artificial intelligence began, in the decades after World War II, by using approaches grounded in logic and symbolic rationality. After this early approach largely failed, leading to an "AI winter," engineers eventually retooled with data-driven and empirical methods. Concurrent with this new wave came an unprecedented proliferation of human data via emailing, blogging, the authoring of HTML, the snapping of digital photos, etc., much of which was posted publicly or accessible internally to the cloud platforms that hosted it all. This data furnished the fuel for today's data-centric AI.

One consequence of this history is a shift in the balance between data and algorithms. Software development entails a variety of different kinds of input data (global variables, input files and databases, graphical elements for the user interface, essentially anything that can't be generated procedurally). At the same time, development requires a complex set of procedures (function calls, simple arithmetical and logical operations, if/then control structures). For many years, the normal way to do software development was to have a relatively small amount of data and a relatively large number of procedures. "Normal" is often a contested word, to be sure. But I mean everything from when Linus Torvalds built the Linux kernel to when Cory Arcangel wrote the assembly code for *Super Mario Clouds*. Today's AI essentially rearranges the previous proportion. Instead of a few variables and data inputs appended to a more prolonged set of procedures, we find massive amounts of data paired with a relatively small codebase. Sure, the code repository at OpenAI or Google is large, but their data stores are almost immeasurably larger. In fact, you or I could program a simple machine-learning algorithm in just a few hundred lines of code. Today's AI is not algorithmically elaborate, even if it remains data intensive. The data is heavy and the procedures are light.

It's easy to get lost in the technical details, so consider two other consequences of AI, one philosophical and another political. Given that it floats atop a sea of data, today's AI relies heavily on the *inductive method* in scientific discovery. I'm generalizing here to make a point, and not all AI is the same, but take neural networks as an example, for which induction is absolutely crucial. A neural net is essentially a layered set of nodes connected in a meshy thicket. Data is flushed through the layers many times over and over, until the thicket evolves into a specific shape. The essence of the shape is captured by a set of floating-point coefficients, which, as a whole, represent an exceptionally complex function operating in multiple dimensions. Having obtained this specific shape, the neural net may

then be prompted to predict future outputs based on how it was trained in the past. This is the key to the empiricist approach. Scientists no longer strive to build a super-brain running on a supercomputer. Today these same scientists might begin with a nonsense brain—for instance a neural net initialized with random coefficients—and hope that their data will train the brain from nonsense to sense. If the rationalist approach to AI had failed by the early 1970s, it looks like the empiricism of the 2000s and 2010s has furnished better results. If deductive methods failed, it looks like inductive methods have succeeded. Along the way, whole branches of mathematics have been jettisoned in favor of other ones more crucial to today's AI, such as statistics and probability, linear algebra (for matrix transformations), and graph theory (for traversing structured data). Because of this, neural nets have essentially altered a scientific configuration in place since at least Isaac Newton. For neural nets, behavior generates laws, whereas in a Newtonian world, laws describe behavior. (To be sure, Newton had to have observed a lot of behavior before arriving at his laws; yet one thinks of a Newtonian world as a world where behavior is determined by laws.) Neural nets basically *automate induction* and thereby automate the scientific method itself, as Chris Anderson notoriously claimed in his 2008 *Wired* article "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." *Data scientists have discovered that theory is obsolete*—how bracing it is to write these words here in a journal whose front-cover descriptors are "Art | Theory | Criticism | Politics." In other words, Hume won and Leibniz lost. The empiricists and the pragmatists and the skeptics won, which is to say the Brits and the Americans won. Today's AI is an Anglo-American science through and through. It's not rationalist in the French tradition. And it's certainly not romanticist in the German tradition. Let this be a warning to all defenders of AI: Whether you know it or not, you are all Anglo-American jingoists.

Finally, a thought on the political implications of AI. Endorsing an Anglo-American scientific methodology is certainly political. And several scholars, many of them women of color, have shown that whatever cultural and social values we embed in AI's training data will reemerge in the fully trained tool. Which is by design. Consider Wendy Chun's concerns about the connection between scientific correlation and "discriminating data," as described in her recent book of that name. Still, another political question strikes me as equally important: Where does the data actually come from? Discounting natural inputs like weather statistics, almost all training data comes from human sources. Flickr images, Web pages, Gmail messages, credit-card transactions: It's all made by people. This is not to mention the labor of tagging and cleansing these datasets, labor that's often performed by low-wage workers living in countries on the losing end of global capitalism. Ironically, Google et al. are some of the most militant defenders today of the labor theory of value. They can only profit from data that is "rich," and the best way to get rich data is to capture it from a human's deliberative actions. Entropic sources don't work as well. Here the vulgar Marxist analysis will suffice: The vast majority, nearly the totality, of AI data is the product of *unpaid micro labor*. In this

sense, the AI industry is an extension of what Marx once labeled "primitive accumulation," which, among other things, relies on the direct expropriation of value from supposedly freely available sources such as the natural world, the public commons, and proletarianized labor. Of course, "expropriation" is just a fancy word for theft. And indeed, most of the AI training datasets were pilfered in one way or another. We all should be paid royalties every time someone uses ChatGPT. Even better would be to expropriate the expropriators and return these tools to the public domain from which they derive. In sum, show me an AI tool and I will show you a labor violation.

Here is where the unflagging commitment to empiricism and pragmatism begins to pay off, at least for the expropriators. Today's AI tools aren't judged so much in political or even metaphysical terms (are these tools good for society? Is AI actually conscious? Is it ethical to use a chatbot?) but rather in terms of measurable utility. Do they work? Do they help me get things done? Do they increase my productivity? AI has bracketed the question of *truth*—I blush even using the word—and instead measures value by whether a human is successfully convinced by a tool's affective theatricality. We used to call this the pathetic fallacy. Although for Alan Turing it wasn't a fallacy so much as a test, a test that may be passed. *Given an affect, do you believe it is real?* The Turing Test is typically underplayed in today's discourse, but AI is almost entirely dependent on these kinds of thresholds of human perception and believability. (If you think ChatGPT is sentient, do you also think Barbie is sentient? If not, why not? "Because interactivity" is not a convincing answer.) The nineteenth-century psychophysics of Gustav Fechner or Hermann von Helmholtz hasn't disappeared so much as insinuated itself into the very fabric of the medium. Cinema scholars have long talked about "flicker fusion" and the precise speed beyond which still photographs become moving animations. Today we ought to talk about "intelligence fusion" and the precise threshold beyond which humans perceive a synthetic Other assembled from discrete symbols like pixels, characters, or wavelets. In other words, in order to understand AI we ought to study something like acting or theater rather than computer science. To make sense of this technical epoch, we will need a good theory of pretending.

ALEXANDER R. GALLOWAY is a writer and computer programmer.

JENNIFER GONZALEZ and WARREN SACK

Warren Sack: With all the excitement of the past year expended on developments in technologies like ChatGPT, artificial intelligence (AI) seems to be a material thing, a combination of algorithms and server architectures that support large-language models and deep learning to drive text and image generation. But AI is not a thing. What makes ChatGPT and similar developments "AI" is neither math nor technology. AI is a rhetorical construct with no fixed object. There is nothing material that has remained steady throughout its almost seventy-year history—from the founding conference at Dartmouth in the summer of 1956 to the present. Instead, the software and hardware of AI have changed radically and frequently. What has not changed are the hyperbolic claims about the revolutionary nature of AI, how it will become indistinguishable from humans, and the apocalyptic futurology, entangled with a certain genre of science fiction, that predicts "our" demise and the rise of AI as humanity's overlord or successor, an "existential risk" for the human population.

Jennifer Gonzalez: If AI is a rhetorical construct, what is at stake in maintaining it? In the art world, as with creative writing, the question of "existential risk" is less abstract or philosophical and more practical. AI threatens to be an existential risk for artists' and writers' livelihoods, and it contributes to a fear of creative cultural workers' being replaced by machines. But as a tool, a form of visualizing, the latest AI projects fall into a long history of works that engage with computation to "automatically" produce art using robots and other kinds of pattern-matching algorithms. Harold Cohen began in the 1960s; what makes the latest version of AI so different, so appealing?

W.S.: This type of pattern matching is more than a creative enterprise for mining large datasets. It's a new form of capitalist accumulation. The scale has shifted, the granularity has shifted. A certain set of fantasies and fears that have long been entertained, since well before the advent of this particular technology, probably going back to the Luddites, is also being activated. But the newness here really has to do with the scale of what's been collected, and the scale of what's being stored and retrieved. The corpus includes a significant portion of all the texts that writers have ever published, and it likely contains many of the images any of us have ever made or seen.

J.G.: As a pattern-matching algorithm, dependent upon "found" data, AI shares with Dadaist photomontage and Surrealist automatic writing a scavenger's approach to both visual and textual borrowing, automation, and appropriation. AI is a scavenger, and artists who use AI are not so different from artists in the past who used secondary sources like newspapers, television shows, or computers in their works. Is there more at stake with this latest generation of AI because it is so tightly tied to capital's aspirations of reproduction without labor? Might it be productive to see AI as not only a rhetorical construct but

also a large and expensive art project? In such a framework, it becomes conceptually acceptable to repurpose found materials to produce variations; but is this the same as what humans do, or not?

W.S.: Historically, AI technologies have been described as doing what humans do. The better analogy is that they do what compression algorithms do: They take a lot of data and compress it, and elements are lost. One of the weak points of the contemporary technology is that it can't duplicate citations properly. So we find humorous examples of lawyers filing suits who cite cases that don't exist because they have ChatGPT generating their briefs. AI, like any compression algorithm, loses the original, and it has to fill in areas of the image or text that it hasn't stored explicitly. Most importantly, it's not doing anything like what a human does. Neural nets don't work anything like neurons, so the difficult question is, What is this technology if it's not like a human?

J.G.: If AI does not do what humans do, and never has, what does the anthropocentrism of seventy years of AI research and funding mean? Is AI both desirable and dangerous only when it approximates the human? The yearning for creation that motivates these computational models of cognition seems grounded in a fantasy of non-biological human reproduction wrapped in a parallel Oedipal fear of the child's replacing the parent. Inflected by this inherent paradox, AI appears to have quite contradictory impulses. Or does the story of AI more closely approximate the Pygmalion myth, in which the artist fulfills art's ancient promise to copy reality but does so to such a level of perfection—making human statues more perfect than actual humans—that he then falls in love with his creation? Venus, impressed by Pygmalion's irrational devotion, ultimately breathes life into his stone virgin. Ovid writes, "It appeared in truth a perfect virgin with the grace of life, but in the expression of such modesty all motion was restrained—and so his art concealed his art." The anthropocentrism of both AI and Pygmalion derives from the idea that the artist can create a more perfect nature.

AI works as a rhetorical construct through the concealment of its means of production and reproduction—an art that conceals its art—thanks to contemporary non-disclosure agreements and patent law. Claiming proximity to human cognition—or indeed to have perfected human cognition—AI seems no closer to a real human than a statue made of stone.

Most forms of verisimilitude are defined by dissimulation. From the Greek and early-modern masters of trompe l'oeil to the academic painters of the nineteenth century to the engineers of contemporary robotics and the makers of digital-image manipulations, verisimilitude deploys an art that conceals its art with great care and skill. Blending brushstrokes, smoothing pixels, calculating the fall of light, and refining the texture of stone or plastic all lead to a final result designed to mirror nature and fool the eye—to invite viewers to temporarily suspend disbelief. AI, in its quest for cognitive realism, also relies upon humans to look beyond a massive infrastructure of pattern matching, to suspend disbelief, in order to imagine something like "machine intelligence."

W.S.: The coding practices of artificial intelligence produce a form of science fiction. The rhetoric that supports the socio-technical imaginary of AI (recognizable in everything from science fiction to grant proposals and technical publications) is both written and coded. The latter takes the form of demonstrations of the technology (what Silicon Valley calls "demos") meant to show what might be possible in the future with more time and more resources devoted to a project. AI rhetoric, however, conflates these demos with working systems.

If AI demos sell well to a larger public, that public can suffer when a technology does not actually work, when it is measured against the challenges of the everyday world. Current demos that are not actually working systems include Tesla's Autopilot and Google Translate. Polyglots know, for instance, that Google Translate can be an excellent tool for rendering a quick first draft of a translation. But they also know that Google Translate will likely make fundamental and sometimes critical mistakes that no competent human translator ever would. So any output from Google Translate must be reviewed. Effectively it "works" as a translator only for human translators able to spot any crucial mistake. Analogously, Tesla's Autopilot is recommended for use only by competent drivers who are advised to keep their hands on the wheel at all times.

Paradoxically, what distinguishes AI demos from other computer systems is that the latter work and the former do not, or if they do, they do so only marginally. This paradox was already clear decades ago when AI cofounder John McCarthy is said to have described what has subsequently been known as the "AI effect": As soon as AI researchers solve a problem, that solution is no longer considered to be AI. Especially in the early days, in the course of AI research efforts, practical technologies were created as a byproduct and some were then employed widely beyond the AI community. For example, McCarthy created a programming language, Lisp, still in use today by many and, for its design, articulated key principles still embodied in contemporary programming languages like JavaScript and Python. But once Lisp became a tool and not a buggy experiment, it was no longer called AI. We see the same effect today: No one calls JavaScript per se AI, even though key elements of its design are indebted to constructs developed in AI labs.

J.G.: You suggest that we consider AI a kind of parafiction. Why?

W.S.: In their book *Leviathan and the Air-Pump*, Simon Schaffer and Steven Shapin point out how critical it was for the Royal Society in the 1600s to demonstrate laboratory procedures and mechanisms in a manner that was charismatic and yet not so flashy that the public and, particularly, their benefactors would mistake them for the magic shows of alchemy.[1]

AI demos are a kind of software and/or hardware that work under highly circumscribed conditions. Demos are designed to show possible fun-

1. Steven Shapin and Simon Schaffer, *Leviathan and the Air-Pump: Hobbes, Boyle, and the Experimental Life* (Princeton, NJ: Princeton University Press, 2011).

ders what could be possible with more time and/or money and to show the general public a possible future for technology. Thus they are close kin to parafictions as Carrie Lambert-Beatty describes them:

> A parafiction is related to but not quite a member of the category of fiction. . . . Post-simulacral, parafictional strategies are oriented less toward the disappearance of the real than toward the pragmatics of trust. Simply put, with various degrees of success, for various durations, and for various purposes, these fictions are experienced as fact.[2]

The hazards of demos as AI parafiction become clear when they are presumed to be nonfiction under uncontrolled conditions, such as when a Tesla driver dies in a car crash believing the Autopilot setting means that it can actually pilot itself. Lambert-Beatty's point that parafictions are not simulations is important to any understanding of AI demos. AI demos are speculative: aimed not at the correct prediction of the future (as are simulations) but toward the imaginative exploration of a possible future (or past). We recognize this today in various experiments with large-language models that yield "hallucinations"—made-up "facts," citations of nonexistent legal or scientific publications, etc.

J.G.: Without doubt, the moment of someone's legal status's being decided or their car's safety's being assured (or not) is the wrong time for software to produce "hallucinations." But if we think about AI as a kind of art practice, rather than a science of mind or technology of dependability, what might be the value of these hallucinations?

W.S.: Consider the science-fiction writer Ted Chiang's article "ChatGPT Is a Blurry JPEG of the Web."[3] When an image is compressed into the JPEG format, the file size decreases and details of the image are lost. Consequently, when a JPEG is opened, one sometimes sees "compression artifacts"—parts of the image that were singular in the original but are sufficiently similar to other elements to not be worth storing as unique. The larger the compression ratio chosen, the more dissimilar parts are taken to be the same and, consequently, the more compression artifacts are introduced.

Chiang considers this an apt analogy for ChatGPT, where large amounts of data are compressed into a model that can then be used to generate texts similar to but not the same as the texts (or images) in the original data. Only a fraction of the original text/data is preserved in the model. Thus the "hallucinations" of AI demos are like the compression artifacts we see in a blurry JPEG. Perhaps what is most interesting about AI demos is not what they are advertised as being able to do but what they

2. Carrie Lambert-Beatty, "Make-Believe: Parafiction and Plausibility," *October* 129 (August 2009), p. 54.

3. Ted Chiang, "ChatGPT Is a Blurry JPEG of the Web," *The New Yorker*, February 9, 2023.

hallucinate. Insofar as these hallucinations are caused by compression, they may be compared to the dreamwork and parapraxes that Freud, employing a metaphorics of thermodynamics, claimed were caused by condensation and displacement. We foresee a possibility of a renewal of Surrealist experimentation with compression instead of condensation.

J.G.: But unlike the dreamwork, no underlying conscious or unconscious mind informs the "compression artifacts" of AI "hallucinations." This does not prevent AI artists like Refik Anadol from metaphorically titling his projects "machine hallucinations" or machine "dreaming." Historically, art and aesthetics have explored the value of hallucinations, especially those that are not misread as facts. The critic and curator Jean-François Chevrier points out that hallucination, defined as a "perception without an object," has long been the purview of the madman or the visionary.[4] Nineteenth-century philosopher Hippolyte Taine believed artists know more about hallucination and perception than the "average" person.[5] While his assertion is debatable, it reveals a long-standing (and ongoing) curiosity about the potential benefits of altered states of mind that operate within and between hallucination and reality. Famous for claiming that man can be best understood as a "nervous machine," Taine had an influence on structuralism, and Saussure in particular, that is noteworthy.

How might the historical theories that link art and hallucination offer a framework for approaching the inherent isomorphisms between art practice and AI research? Since most human cognition unfolds in ways that are neither predicable nor rational, "hallucination" might be a more apt description for that which AI does approximate—even if accidentally.

W.S.: As far as I know, the contemporary—usually journalistic—use of "hallucination" to describe what AI generates has not been considered within the longer history of hallucination as a concept concerning human psychology, a history that Taine participates in.[6] Colloquially, the "hallucinations" of generative AI are just mistakes or errors, like compression errors. To consider AI hallucinations as akin to those of human psychology would be to assign them a constitutive rather than accidental role. For example, Freud considered hallucinations as fantasies of the subject or symptoms of forgotten traumatic experiences.

Finding AI demos specifically designed to generate hallucinations that are not simple mistakes but indicative of something more requires one to look to the history of AI, to the era of so-called symbolic AI, a paradigm (in

4. Jean-François Chevrier, "Between Terror and Ecstasy: Artistic Hallucination," Tate Modern, https://www.tate.org.uk/tate-etc/issue-24-spring-2012/between-terror-and-ecstasy.

5. Ibid.

6. Diogo Telles-Correia, Ana Lúcia Moreira, and João S. Gonçalves, "Hallucinations and Related Concepts—Their Conceptual Background," *Frontiers in Psychology* 6 (July 27, 2015).

the Kuhnian sense) that both precedes and follows neural networks. Some symbolic-AI work could be considered psychoanalytic in orientation.[7]

Symbolic AI owned the label "AI" from the 1956 Dartmouth conference until the rise of more statistical approaches in the 1990s and then the contemporary neural networks of the 2010s. Contrary to symbolic AI, the primary computations of neural networks are arithmetical; a lot of numbers need to be crunched to have them operate. This is why, for example, the chips (GPUs) developed by NVIDIA for doing the calculations of computer graphics are now so important to the industrial-scale AI of today. In contrast, symbolic AI was based on computations done primarily with symbols. The distinction was originally crucial to AI researchers because of their investment in what Allen Newell and Herbert Simon (two co-founders of AI also present at the 1956 Dartmouth conference) called the "physical symbol system hypothesis," the idea that the essence of both human and machine intelligence was the manipulation of symbols. It is ironic that today the technology thought to be canonical AI—neural networks—was previously not considered to be AI at all.

The neural networks of today are essentially very large polynomials (with hundreds of billions of parameters), and so their "hallucinations" are arithmetical and not symbolic. In the current paradigm, AI reduces all thought and action to arithmetic, specifically adjusting the numerical values of the polynomials' parameters.[8] A new sub-field of "explainable AI" (XAI) has been founded to find the means for data-driven AI to report errors and numerical calculations in a symbolic form understandable to humans. XAI is challenging because contemporary AI technologies work so differently from humans. For AI artists there is an analogous challenge that is also akin to the Surrealists' fascination with the unconscious: How can new images and texts be produced in collaboration with intelligences so unlike our own conscious minds?

UCSC professors JENNIFER GONZALEZ and WARREN SACK's books include *Subject to Display* (MIT, 2008) and *The Software Arts* (MIT, 2019).

7. Consider, for example, Marvin Minsky, "Jokes and Their Relation to the Cognitive Unconscious," in *Cognitive Constraints on Communication: Representations and Processes*, Lucia Vaina and Jaakko Hintikka, eds. (Springer, 1984), pp. 175–200; and Kenneth Mark Colby, *Artificial Paranoia: A Computer Simulation of Paranoid Processes* (Pergamon, 2013).

8. Gilles Deleuze anticipated this state of affairs over fifty years ago when he wrote, "Problems are now traced from algebraic equations and evaluated according to the possibility of carrying out a series of operations on the coefficients of the equation." Gilles Deleuze, *Difference and Repetition*, trans. Paul Patton (New York: Columbia University Press [1968] 1995), p. 161.

HOLLY HERNDON and MAT DRYHURST

Over the past decade we've been working with machine learning as artists. One thing that has become a tenet of our practice is to embrace the way that, as AI model outputs reflect what you feed them, model-making is a new medium. AI is sold by many as an opaque alien intelligence, but more accurately it is a kind of aggregate human intelligence. So we've focused on exploring that, through public performances, or creating protocols to create training material for bespoke AI models. For example, our record *Proto* (2019) was created in part by large, live group-training ceremonies of opt-in call-and-response singing, where the audience lent their voices to form a new training set.

We've spent a lot of time learning to manipulate latent space, or embedding space. Concepts in the latent space of large public models like ChatGPT or Stable Diffusion are determined by an abstract consensus based on whatever data was scraped from the commercial Internet. We feel this has significant implications for culture and identity, as it is a departure from the original dream of the Internet that promised self-determination. If you want agency over who you are in these new environments, you have no choice but to take control of your data and representation.

We've been interested in these questions of data ownership and governance for some time. Building data-governance systems is, for us, no less of an art project than our other work. In 2021, we created *Holly +*, which anticipated what it might mean to have a splinter, disembodied version of "you" that is controllable by others. We openly released tools and models trained on Holly's voice to allow for anyone to perform through her, as well as establishing a protocol of profit-sharing for when profit is made from new works. In the visual realm, we've also done a lot of work identifying what public AI models know about us. We have devised techniques to intervene in and mutate those concepts/embeddings. We found that the more famous you are on the commercial Internet, the more you exist—the more high-fidelity your embedding is. So the concept of Beyoncé in image models can be spawned with highly complex features distinct to her. Generating a "Beyoncé" is incredibly detailed, because her image is everywhere—you can even see her teeth. But the concept of Holly, for example, is less refined, it's compressed into a red haircut. For this year's Whitney Biennial, we came up with a technique we refer to as "cliché poisoning" that promises to smuggle mutant new features for Holly into public models, so long as we maintain the presence of her distinctive hair in the data.

Recently, our project *Readyweights* was an attempt to identify points in embedding space as artworks. The Readyweight is, like the readymade, something that already exists, that you might "discover" or highlight. Given the infinite number of permutations along the manifold, it transpires that if you create an original sculpture and then attempt to excavate a large image model to locate that new concept,

Holly Herndon and Mat Dryhurst.
"A House that Looks like Holly Herndon." 2022.

you will likely find a representation of it that already existed as latent potential. In other words, you might think you've come up with something entirely new, and then you dive into latent space and discover that it was already there all the time, even if it was never "made." There is something both harrowing and reassuring about that, consonant with our understanding of human creativity and intelligence as a process of discovery and co-creation.

We're often trying to push these tools to their limits in a way that exposes their current capacities, and to distinguish between hype and legitimate opportunities to explore. In audio, it is possible to gloss over things: The ear can trick you more than the eye can. With *Proto* we deliberately chose *not* to try to make it sound more high-fidelity than the models we were using, because at the time when we were recording in 2017–18, a lot of these generative-AI systems were still quite lo-fi, and so we leaned into that. In production we chose to degrade the extremely high-fidelity recordings of the live ensemble to better complement the nascent qualities of the AI tools of the time, rather than default to some futuristic sheen. We find it important to accurately reflect the state of the art, as a time capsule of sorts.

Working with models as a medium is challenging due to the sheer volume of media produced. Past a certain point, an individual image or sound output becomes banal. This is partly why discovering true or formal representations of concepts is so compelling. The image or sound itself may be less interesting than how you reached it, or why you chose to do so. We don't think people have fully digested just how abundant media is about to become, and how that changes our relationship to producing it. It will soon be effortless to produce anything that may have seemed virtuosic in the past century, and that raises the question of what new dimensions of virtuosity are yet to be explored.

Play from Memory (2024), a set of works we created for the MoMA show *Sound Machines*, explores this history and the future of virtuosity and learning. We were interested in the relationship between music pedagogy—how we teach children music—and how we teach machines to learn music. We came across the Orff *Schulwerk* (or the Orff Approach), a protocol developed by the German composer Carl Orff (1895–1982) and his colleague Gunild Keetman in the 1920s. They developed "imitation games" to teach children to improvise freely from memory, prompting them with instructions like mnemonic games and graphic symbols, eschewing traditional notation. It is a beautiful way of teaching children how to play music and how to collaborate, using simple tuning systems and simplified instruments so that kids could improvise and feel free. There are similarities between the *Schulwerk* approach and AI pioneer Seymour Papert's Constructionist ideas, which posits that intelligence is something that we *do,* something that is uncovered between two humans, or a human and a machine. It is a beautiful reminder that, despite our newfound ability to generate infinite media, it is important to not confuse the media for the art. Art is intersubjective, something we do, and something we discover. Abundant media does not threaten that.

For *Play from Memory*, we created new machine-learning models to produce strange, wonderful orchestral and choral sounds. We often think back to a live show we played in Paris a decade ago, when an ensemble of children called the Wing Beats opened for us. The kids were invited to come in and build their own instruments and then perform in a structured improvisation. One child had a giant flower with contact mics all over made of metal, and he would bang the different petals and make these amazing sounds. They were so careful and deliberate to make room for each other. To this day the memory still moves us.

HOLLY HERNDON and MAT DRYHURST are artists and composers based in Berlin.

TISHAN HSU

AI is a revolutionary advance in the technological development of our species; it is of a different order from the kinds of technological advances that have preceded it, partly because of its cognitive aspect and partly because of what, I believe, will be something larger emerging from it. Needless to say, the implications are vast and unknown. In that sense, anything I say about AI is coming from my limited understanding of what AI is already demonstrating. It will have an increasing impact on all sectors of society and is raising unprecedented questions around issues of human agency, ethics, and politics.

For me, as an artist, the difference between AI and earlier technological media—like photography, video, and computational systems—is its generative aspect. This intrudes on my understanding of my agency in negotiating the world to such an extent that it is difficult to conceive of the implications of AI in terms of my current structures of understanding. Thus AI confirms the possibility of an "other" that is different from "nature" in the traditional binary of human/nature and/or different from "human" in the binary of human/technology. Art comes from culture, and just as art has always given a more unconscious expression of realities that may not be expressible through everyday discourse, whether written, spoken, or visualized, I believe art will do so once again as the impact of AI is more widely felt.

How will art do this? It is too early for me to say. AI's impact on art may emerge rather than be determined and may not even be recognized as an expression of AI until later. This unpredictability is part of the value that art will introduce by providing a means for the culture to cognitively grasp the nature of the change AI will effect, particularly those aspects that are not consciously anticipated. To the extent that vision is a critical, if not primal, cognitive component of the human species, the visuality of art will continue to have an important role in our understanding of AI's relationship to culture—both our own and that of the rest of the world. Being able to *see* the impact of AI on basic image production already gives a palpable sense of the degree of AI's impact on the wider culture, which is happening *invisibly* in other sectors that the larger culture does not yet experience.

Some may say that the point of "singularity" in AI has already been reached. AI does not need to reach "consciousness" to present itself as already superior to human intelligence on many levels. However, we, as a species, created AI. Rather than seeing it as alien and "other," we might be able to conceptualize it as an extension of the human, albeit a potentially dangerous and destructive one. It would not be the first. One might consider the issue of utopia/dystopia by asking, Was the evolution of the human species utopian or dystopian for the planet Earth, from which we derive? We are both a part of Earth and destroying it. AI is a part of what the species has brought to the planet. It is conceivable that AI will help the species to address the catastrophe of climate change and assist in returning it to a level of sustainability while at the same time undermining how we experience

human and political agency. That would present AI with both the utopian and dystopian aspects simultaneously. The human species emerged, and we have the world as it is.

AI is already enabling levels of computation that allow the species to explore and benefit from unimagined advances in technology's ability to address some of the enormous problems we face. Its use and development have just begun in AI's longer-term evolution. If we created an AI that presents an alternative intelligence, then one could say that our agency as a species now includes AI. *It is us.* It is potentially destructive like us and continuous with us, as nature. Technology is nature and produced by a natural species.

There are many interesting artists working with AI as a medium/tool. In many cases, it is as a tool for imaging, but the mode of understanding offered by these images often references historical visual syntaxes. AI has not yet begun to affect our conscious understanding of ourselves, and until that process begins, its impact on artistic concepts and ideas will be limited. But I believe it inevitably will happen. At the same time, artists may show the impact of AI in their work without necessarily being aware of it. There is also a need to examine, critique, and expose the dangers and damages that both artists and non-artists can do with AI. This work could be done in many different ways in the arena of political action, which could also be considered art.

I, along with other artists, am beginning to explore the potential of generative AI in my work with digital imaging. I find it seductive and exciting but also sometimes frustrating: What can I do with this? Most of the images AI has generated for me I find unaffecting beyond their weirdness. But a very small percentage are liberating and inspiring. I see them as extending my imagination and have no qualms about accepting that. But the imagination that's being extended includes the part that issued the instructions and selected the images.

I began considering myself a cyborg long ago, and the attachment of AI to my cognitive production through the computer is just another step in a long evolution of my body with technology, which has been taking place during the twentieth and twenty-first centuries. Much of the AI art I am seeing is taking twentieth-century imagery and manipulating it, sometimes spectacularly, to present somewhat familiar imagery in a different way that is perhaps radically produced by a machine. In many instances, it is image-based. But I find that the weirdness of AI's ability to throw together and merge images reaches a limit. It's a feeling, and I am not sure what that feeling of "limit" indicates.

AI could perhaps at some point show us a kind of consciousness, one located in the visual field, that is different from what we have experienced before. Somehow, I feel that a human will need to be involved to do that. On the other hand, the concept of singularity in AI presents the possibility that AI could create an image that has an altogether different consciousness. But would we, as humans, recognize it? This has happened before in the history of art. Can we recognize something if we have never seen it? A similar question arises in exploring the ques-

tion of the existence of other life-forms in the universe, which is currently being investigated conceptually and materially in science. Would we, as a specific life-form, be able to recognize such life? Would our technology be able to recognize the attributes of an alternative life? This question resonates with the question of whether humans would recognize consciousness within AI. This raises the question of what consciousness is. If we do not fully understand it ourselves, would we be able to recognize it in AI? The question of an art without consciousness leads me to ask whether our recognizing it would matter anymore. Does the human species need consciousness at the level of art, or do we project it onto our experience such that there is no art without consciousness by definition? If there is an art without consciousness or art without artists, does that imply there would be people without consciousness? That would suggest the possibility that the definition of a person might change—a possibility to which I am open.

TISHAN HSU is an artist.

DAVID JOSELIT

If we define artificial intelligence as a procedure by which computers train on large collections of data to recognize or,[1] more recently, to generate what Hito Steyerl has called *mean images* (signifying, in part, their constitution of a mean, or intermediate position, within a vast digital archive),[2] then a surprising analogy arises. Over two centuries ago, the modern museum was constituted from "scraped data"—i.e., appropriated artworks of the *ancien régime* within France and masterpieces from across Europe looted in Napoleon's military campaigns. The resulting collection—or dataset—assembled in the Louvre Museum generated something new: an aesthetic history of Europe aimed at establishing French imperial sovereignty. Perhaps one could say that under Napoleon the Louvre was trained on a history of monarchical, sometimes absolutist, power to generate a facsimile of authority for a self-made man—Napoleon himself—who had risen up from the provincial nobility to the exalted status of emperor. Or, going further back, one could say that the Louvre—alongside Jacques-Louis David's revolutionary festivals, which gave the French Revolution its "mean images"—was trained on the sacred traditions of antiquity to generate a secular democratic cult. Either way, the principle was the same: The museum functioned as a machine for producing new histories from old artworks to legitimate innovative forms of government. To us, this function is so familiar as to be taken for granted, but for commentators of the time, such as the conservative archaeologist, writer, and architectural theorist Antoine-Chrysostome Quatremère de Quincy, it was recognized as a radical intervention:

> Moving all the monuments, so as to gather up their dispersed fragments, methodically classify their debris, and make of this grouping a lesson in modern chronology is for a living nation to become a dead nation; it is for the living to attend their own funeral; it is to murder Art to write its history; it is not to write the history of art but its epitaph.[3]

Quatremère objects to what I would call "museum learning," a predecessor to the dataset-dependent forms of machine learning that we confront today. It is worth taking seriously his resistance to the museological imperative to make artworks perform as *documents of history* when in fact they are so much more than that.

1. While many scholars prefer the term *machine learning* to *artificial intelligence*, I will use them interchangeably in this text.

2. Hito Steyerl, "Mean Images," *New Left Review* 140/141 (March–June 2023). In this very useful text Steyerl also addresses the question of digital labor that I discuss later.

3. M. Quatremère de Quincy, *Considérations morales sur la destination des ouvrages de l'art, ou De l'Influence de leur emploi sur le génie et le goût de ceux qui les produisent ou qui les jugent, et sur le sentiment de ceux qui en jouissent et en reçoivent les impressions* (Paris: L'Imprimerie de Crapelet, 1815), pp. 57–58 (my translation). I discuss these issues in my *Art's Properties* (Princeton: Princeton University Press, 2023), especially pp. 15–38.

I am intrigued by the questions that result from understanding the museum as a modern way of telling history (and not exclusively *art* history), one that is contemporaneous with but distinct from the emergence of history as an academic discipline. In short, does the museum offer an alternative to Hegelian/Marxian narratives of progress by bringing different eras and cultures into the same spatial environment *simultaneously*? Is it the forebear of what we now call *presentism*, developed in parallel with dominant nineteenth-century discursive models of evolutionary history? Conversely, I am interested in exploring how placing machine learning within such a *longue durée*, as heir to the museum's transformation of collections into discourse, allows us to see AI differently, outside the hype and hysteria with which it has been greeted. One potential advantage of adopting such a perspective is how it raises the question of *curating data* in machine learning. For in fact, as many scholars have pointed out (and as many AI firms have tried to obscure), human intervention—curation—is essential to AI.[4] But in much of machine learning, the curator is seen as a mere content moderator or "micro-worker," as opposed to museum curators, whose status is that of cultural arbitrators. What kind of lessons might this demotion hold for understanding both the museum and machine learning? For instance, if we acknowledge that AI is *curated*, does it help us get beyond the fear that it may, as an autonomous digital golem, render humans obsolete?

As I have suggested, the human arbiters of machine learning are usually of a humble status—often low-paid digital pieceworkers who, unlike the museum curators of today, labor anonymously. Paola Tubero, Antonio A. Casilli, and Marion Coville have divided the tasks such workers perform into three categories: training, verifying, and imitating the output of AI. Training consists of the human annotation of raw datasets, which is necessary for computers to "recognize" a particular type of image. "Verifying" is the procedure by which humans check and correct computer outputs (which are often inaccurate), while "imitating" is necessary when the machine cannot generate the desired information and a human being must do so instead—typically without the user's having any idea that the "artificial" intelligence they are receiving is anthropomorphic. Tubaro, Casilli, and Coville insist that such human interventions are a structural feature of AI as opposed to a temporary fix. They write:

> Data availability will never reach a steady state: most use cases for machine learning require ongoing acquisition of new sources to continuously adjust to changing conditions, resulting in a steadily growing

4. For a discussion of the work involved in data classification and the biases therein, see Kate Crawford, *Atlas of AI* (New Haven: Yale University Press, 2021), especially chapter 4, "Classification," pp. 123–49. For an account of the hidden curation within social media (not exclusively AI), see Sarah T. Roberts, *Behind the Screen: Content Moderation in the Shadows of Social Media* (New Haven: Yale University Press, 2019).

need for humans to produce data for more accurate, more precise, and more profitable results.[5]

But if curation persists from "museum learning" to "machine learning," the function has nevertheless been inverted. Traditionally, museum curators have had the power and privilege to select exemplary materials for inclusion in a collection—they are the creators of canons. AI "curators," on the other hand, are tasked with excluding the non-normative, with merely protecting a pre-existing set of expectations (Steyerl's "mean") in an act of rote statistical averaging. In short, as Tubaro et al. conclude, "AI is not the end of human labor but is depriving it of the quality, meaning, and social status that is acquired over time."[6]

Interestingly, the degradation of curatorial work characteristic of machine learning may be witnessed in the realm of museums as well, as evidenced by the emerging trend of building "open storage" facilities for displaying museum collections. The Depot Boijmans Van Beuningen, for instance, which houses the entire collection not on display in the eponymous institution in Rotterdam, is navigated by visitors who use their phones to access information describing the works they may glimpse in storage. Here the curatorial function recedes to that of digital pieceworker, assembling and cleaning the collection's online catalogue so that the public may have access to the collection *as data* while, simultaneously, they are viewing the very same objects through glass vitrines or partitions, as though the real things could themselves only be encountered through a kind of screen.[7] A similar project is under development by the Swiss architects Herzog & de Meuron: Seoripul Open Art Storage, which will function as a joint archive space for the Seoul Museum of Art, the Seoul Museum of Craft Art, and the Seoul Museum of History. Like the Piranesi-esque interior of the Depot Boijmans Van Beuningen, which resembles an architectural spatialization of multiple open windows on a screen, in Seoripul, "a focal point of the interior will be a glazed, conical atrium, extending up from the ground floor to the sixth and wrapped by display cases to offer visitors glimpses of the archive."[8] These new museum facilities offer a dazzling spectacle of accumulation—of data and cultural capital at once. What was once behind the scenes is now the main event. These are the Louvres of the twenty-first century.

One of the advantages of placing AI within a modern historical genealogy of generating knowledge with and through collections is that it clarifies the social

5. Paola Tubaro, Antonio A. Casilli, and Marion Coville, "The Trainer, the Verifier, the Imitator: Three Ways in Which Human Platform Workers Support Artificial Intelligence," *Big Data and Society*, January–June 2020, p. 10.

6. Ibid., p. 11.

7. I discuss the Depot in *Art's Properties,* pp. ix–xvili.

8. Lizzie Crook, "Herzog & de Meuron Proposes Giant Cube for Seoul Museum Storage," *Dezeen*, December 14, 2023.

dynamic that underlies the technological claims, both ecstatic and eschatological, for AI. If we think of machine learning not as the replacement of human agency but as a degradation of the curatorial function, then we might perceive an opening for intervention: Humanists might insist that such labor be revalued and redefined. We must not concede machine learning to tech corporations. Like digital code, corporate values tend to be binary—either AI will save the world or destroy it. Artists and thinkers, as curators of machine learning, might focus instead on building richer worlds—envisaging what lies beyond the mean.

DAVID JOSELIT is a member of the *October* editorial board.

ALEXANDER KLUGE

Wherever priorities rule, the arts are the advocates of the oppressed. Wherever algorithms dominate, we must quickly and lucidly work on counter-algorithms / cooperatively / If not *homo sapiens* perhaps, we are certainly *homo compensator* (*Gleichgewichtler*, men of equilibrium).

At the end of your interesting questionnaire, you pose ten questions. I will attempt to answer the tenth:

What is human or machine, creativity or computation,
in the first place?

My response is brief: The four realities of "human," "machine," "creativity," and "computation" are inextricably entwined. Employing a word from quantum physics: They are "entangled." In the world of light, photons are entangled when they come from the same source. An influence on any one of these "siblings" will cause the other to react as well, regardless of how far away or in which other realities it resides.

The digital age presents us with numerous challenges: We must reconfigure ourselves to answer them. And we must do so quickly, as the developments are taking place at breakneck speed. And yet, this is no race. If they fight, they fall (Homer). And so, there is no unique position. In the relationship between our lifeworld and digitality there are no victories or priorities. In this respect, the Socratic dialogue created between discriminator and generator in the project of generative adversarial networks (GAN) is a good way forward. We just have to expand it. Generator and discriminator are not networks of digitality; they are combined within human experience, the process of evolution present within our bodies and minds, and thus within creativity. With all that in mind, I shall answer the many questions contained in your tenth:

I

"What Is Human?"

The earliest pioneer of digitality, the inventor of the first computer in the world, was the philosopher G. W. Leibniz. He refers to human beings (and all other creatures) as "organic machines." He denies the dichotomy between man and machine. For a long time now I have wondered what actually constitutes this connection within vividness, between that which we human beings invent (and what we call machines) and our subjective structure (what I call the "Bauhaus of emotions").

As a poet and filmmaker, I belong to the Frankfurt School of critical theory. Adorno, Horkheimer, and Walter Benjamin are my masters. When I apply the critical impulse of *Negative Dialectics* and *The Dialectic of Enlightenment* to the praxis of

creating images, to the project of the "Bauhaus of emotions," and to the challenges AI presents us with, the following occurs to me: As human beings we do not imagine ourselves to be either animals or machines. Nor fungi or bacilli. Indeed, I think we are amphibious beings. We are more than chest-deep in evolution, i.e., the animal kingdom from which we come. Which does not mean that we are like wolves or sheep. The roots lie deeper and much further back in time. But we manage animals within ourselves. Our breath, for example, is a stubborn animal. At the last minute it forces the one attempting suicide in the fountain to gasp for air. This is not something I came up with on my own; it's been proven by physiologists. The breath is more lucid than the brain. Our skin, our most extensive organ in terms of area, is yet another animal. In "Why War?," Sigmund Freud's reply to Albert Einstein's letter of 1932, the year of my birth, we read that morality is of no help against the demon of war, but the skin—which reacted to the misery of positional warfare in World War I with allergies, so that the soldier could no longer put on a uniform—is. Our skin is cleverer than our head. I could add a whole herd of "inner animals" to this "republic of animals within us," which together make up what we call human beings. My favorite animal would be the ear. It was created from the lower jaw of a desert snake. With this lower jaw placed on the desert floor, this snake can hear whether prey is coming. Then, in the *longue durée* of this bone's evolution, a few million years pass before it arrives in us humans, greatly reduced in size, in the ear as the middle ear bone. There, this genius rules over such diverse things as language, music, balance, and the nuances with which we make our emotional decisions in relation to others. In view of these animals, how arrogant it would be of us to declare: We humans do not belong to the animal kingdom. Adorno considers this position in *Negative Dialectics* when, discussing Immanuel Kant's categorical imperative, he writes: Try to live so that you believe yourself to have been a good animal. I've shortened Adorno's text somewhat. It does, however, sound considerably more Socratic than Kant's challenge that we act only according to that maxim whereby you can at the same time will that it be universal law.

The same holds true for those things we human beings have produced. Tools, machines, mathematics, digitality. In every one of these products there is the life of a person, indeed of many generations of people. Karl Marx claimed: Cut a well-constructed machine with a knife and human blood will come out.

There is therefore an "intimate connection" (Kant) between humans, the machine world they have created, and the billion-year-old process of evolution. As the French philosopher Michel Serres has formulated it: When we pay attention to the human rights of things, we have the chance of realizing our own human rights as well.

I have responded to the keyword "human" in this rambling form because I am keen to point out that the human is a building in the making. Digitalization and the challenges posed by the world of the algorithm present us with the challenge of starting literacy anew, within ourselves, the subjects. We've got to get mov-

ing in the program of Enlightenment, as the digital metamorphosis of the world is taking place terribly quickly. A self-consciousness saturated with experience and an emancipated practice—which is indeed what this "building" is about—now have considerably less time than fifty years ago.

There are three types of Bauhaus:

> —The Bauhaus of 1923 in Weimar. It was concerned with the construction of cities and industry.
> —The Bauhaus of Nature. It has to do with long periods of time. An Archaeopteryx skeleton from 150 million years ago prepared the chest and wings of a robust New York City pigeon well. "Architecture of the Skeleton."
> —The Bauhaus of Emotions. It has to do with our subjective world of experience. We need to set it up on short notice and prepare for our emotions to respond to the crises of the present.

II

Machine, Creativity, or Computation

As expansive as I was in answering the first part of the question, I have to be a bit more restrictive in answering the second part. I do not believe that you have more than one tool at your disposal in music, rhythm, and the "dance step of the mind" (i.e., an important part of creativity) in AI. The impulse does not originate in the tool. The same holds true—in my estimation—for the application of AI to language. The six thousand languages that exist in the world, if they were really used together, would be equal and possibly superior to any AI. The problem lies in the fact that languages do not cooperate all that much and our forms of communication do not move whatsoever in languages' deeper dimensions, which is to say in their variations. Here AI is like a mirror of our torpidity. As such a tool it can be helpful. But it does not add anything to the potential that we already possess.

Iconography is a completely different story. Images are extremely important for our human imagination. But images speak far too little with one another. What I have to say has to do with my experience as a filmmaker and assumes that AI's capabilities, as a tool, astonished me. I employ them as a "virtual camera," a camera that generates potential and thereby invisible images. To do so I use a variant of Stable Diffusion, which was further developed by professors at the Technical University of Munich in the furtherance of science.

My classic film camera can only capture the indicative. That which stands before it. As a filmmaker I am used to discovering things at my editing table that my Arriflex film camera recorded but that I did not notice while filming. This corresponds to what Walter Benjamin, in his essay "The Work of Art in the Age of Its Mechanical Reproducibility," called the "optical unconscious." I would not main-

tain that my camera, which I love dearly, is "intelligent." It is, however, hungrier for details than my head. I am less perceptive than my tool. This kind of camera, which has existed since 1895—and is the one Eadweard Muybridge, my role model, used to make his discoveries—can, as I said, only ever record the grammatical mood of the indicative. That which stands before its lens. The human power of imagination, however—our wishes, our sense of curiosity, our creativity—is also familiar with the sense of possibility, the conjunctive. To my amazement, thanks to masses of data (which are also produced by humans), my "virtual camera," AI, can turn heterotopia—that which remains invisible though implied in the image—into visible images with diligent calculation and perseverance. This is a marked expansion of the art of film. Having said that, I have to deal with the tool—seeing as it is not forgery-proof itself—like an iconoclast, a destroyer of images. But doing so opens up areas with invisible images that enter into dialogue with us humans and which have long existed within "the reality as a narrator."

III

Constellation as the Fourth Canon of the Art of Narration

Aby Warburg's *Mnemosyne Atlas* does not contain any images that this genius of a man painted himself, but rather CONSTELLATIONS of images he collected. From the time of Babylon until the year 1929. He places them into context with one another. This is precisely what can be derived from AI as a formal principle. It is possible to expand the art of storytelling, the canon of the arts. Until now we have had:

Epic

Lyric

Dramatic

and the *prospect* of a fourth canon.

One often said: This is CRITICISM. Under the impression of the new tool, Stable Diffusion as a "virtual camera," I would like to look in a different direction as a filmmaker. The fourth great chance of narration is known as:

constellation

The name *October* comes from the title of a film by Sergei Eisenstein. This great master had already postulated the constellation principle in his own age. Following his film *October*, he wanted to "cinematize" Karl Marx's *Das Kapital*. James Joyce was to write the script. The film was to last an entire week. Eisenstein had come up with a "spherical dramaturgy." As, he said, reality itself was spherical: "All times are PRESENT." All spaces within and without us form a RELATION. In

the way that celestial bodies—Latin: *stella* = *star*—thanks to their gravity, in other words, their substance, act in a constellatory manner. It is a pity that Eisenstein did not succeed in making this film.

In California, Eisenstein wanted to build a studio with rooms of glass, like a kind of palace of crystal, everything transparent and interconnected. He assumed that such a building would be a lush, universal environment for a classic camera. His utopia is far removed from and at the same time very close to what one could imagine a radical extension of the digital-dialogue principle upon which GANs are based. We are not dealing with simply *one* discriminator and *one* generator in discussion, but thousands.

The three traditional narrative forms—epic, lyric, dramatic—tell stories in a linear fashion. They tell stories in an arc; this arc is concerned with the progression of the narrative, in lyric a bit less so, in drama a bit more. Constellation, on the contrary, has to do with *vertical* narration: a mine, a spring, a catacomb, it has to do with the principle of depth: the groundwater of experience. And one of the most important forms of working this way is commentary. For this particular working method the tools of Stable Diffusion—at least as far as images and the juncture of image and text are concerned—are very useful. We only have to learn to respect the tool's obstinacy.

IV

A Digital Commentary from My Virtual Camera on an Image in Aby Warburg's *Mnemosyne Atlas* (As Example)

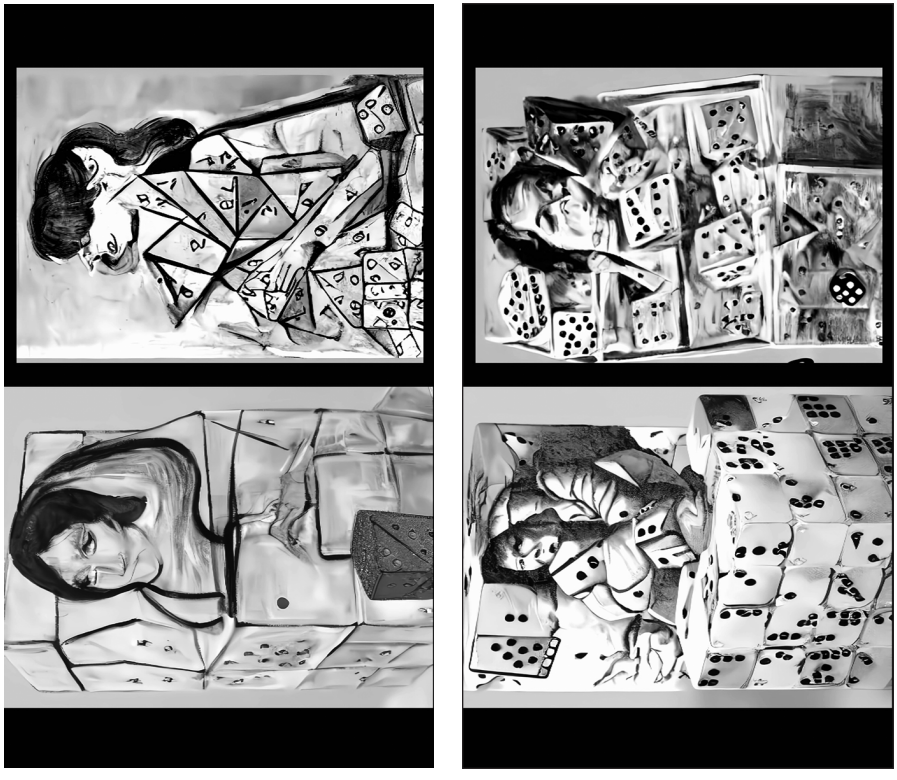
It has to do with one of the images that Aby Warburg collected on the theme of "Fortuna": How has the goddess of fortune been depicted within the tradition of art? This particular image is to be found on Panel 48 of Warburg's *Atlas*. The allegory comes from the sixteenth century. But the idea of lucky dice belongs in the present, too. How can such leaps in time be filmed? The following images come from one of my films about Warburg's *Atlas*, which I am working on for Cornell University's website within the framework of the project Cultural History in Dialogue. Fortuna, a kind of goddess, is standing with one foot upon a ball (this is the traditional metaphor), the other upon the tiller of a ship that has already sunk. The constellatory method becomes truly understandable when this image is also placed in relation to the other images on Panel 48. I cannot present this kind of commentary here, only the working method of an individual film, the one that concerns Picture 14.

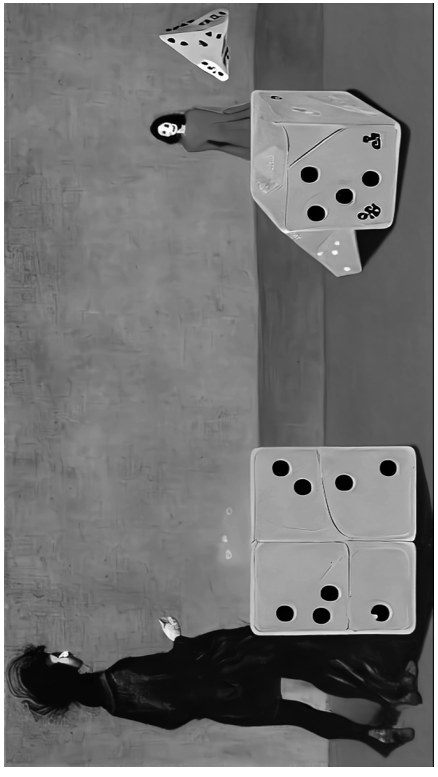
Panel 48,
Picture 14

Fortuna /
Nicoletto da Modena
(1506)

"Fortuna with one foot
on the tiller"

" Fortuna mit einem Fuß auf
der Ruderpinne"

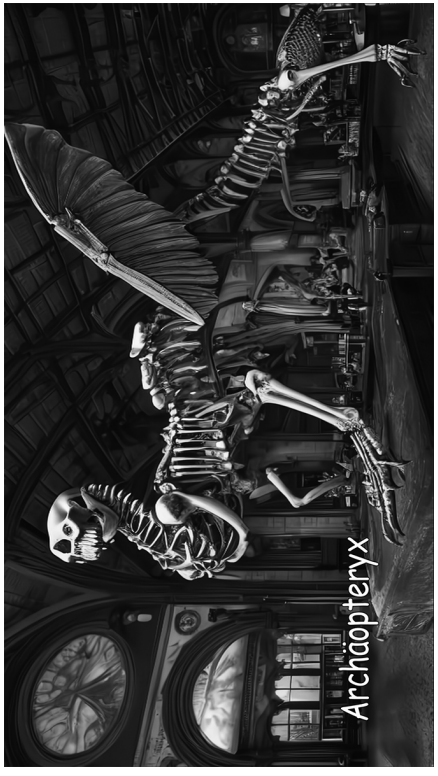


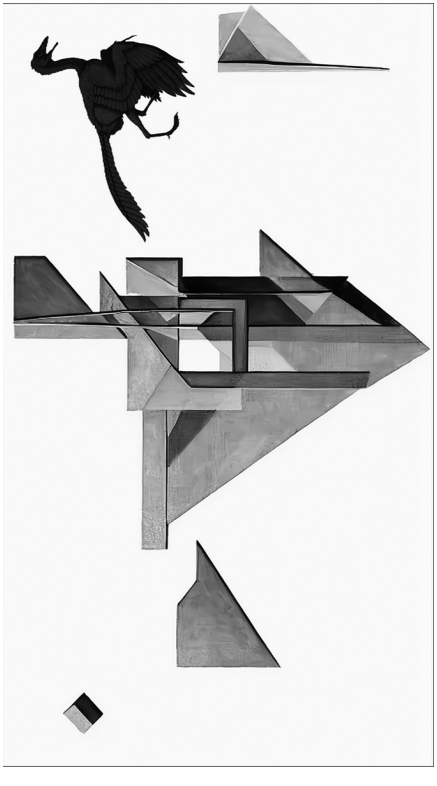


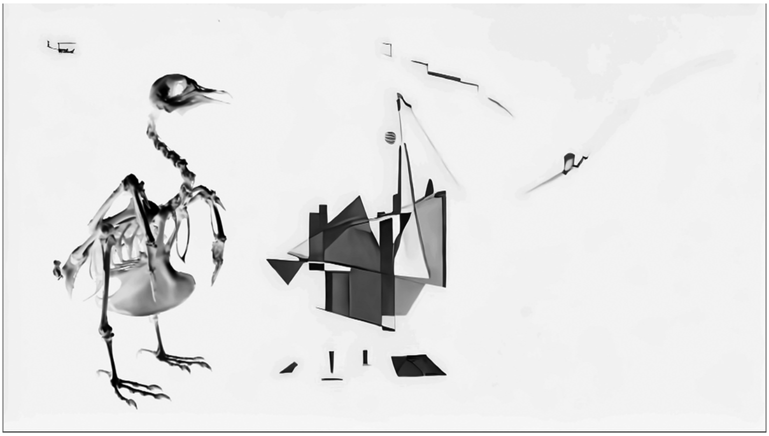
All these images come from the same template. It is shown in the small picture on the right side in the image above. It is good when the source remains in dialogue with the variations. The original joke of the allegory is already contained within the original image from 1506: The goddess Fortuna is standing upon a rolling ball. This is the traditional metaphor for the uncertainty of fate. In the image, however, the ball is lying in water and will soon sink. The boat next to it, upon whose tiller stands Fortuna's other foot, has already sunk. As a filmmaker I could shift this image detail into a series on the sinking of the *Titanic*. But, as opposed to my film camera, the AI can also shift to other forms of chance, e.g., the luck of the dice roll.

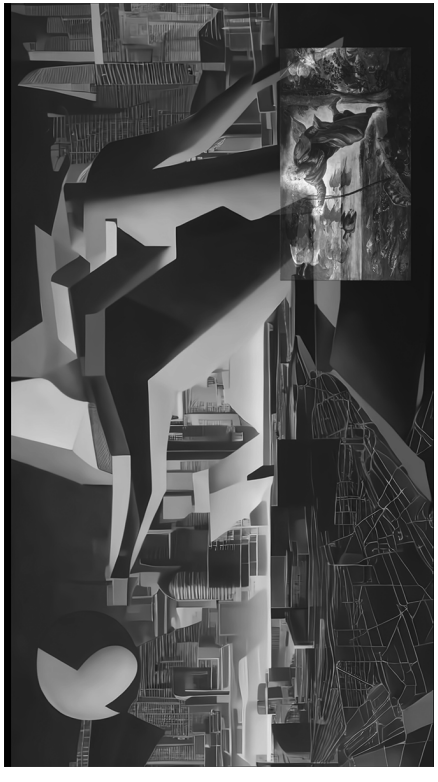
V

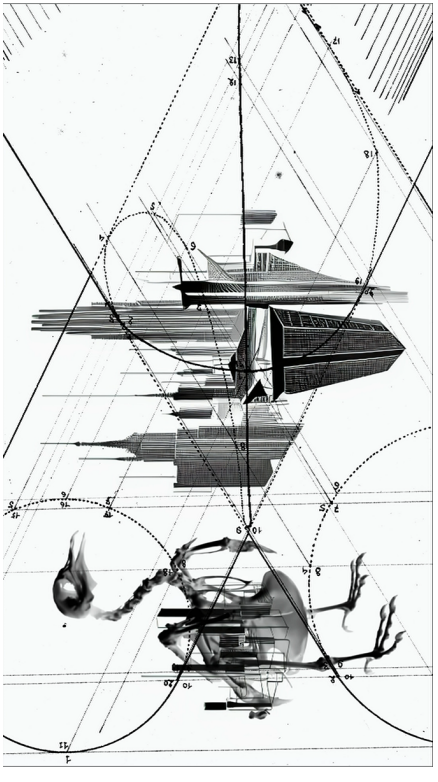
An example of the iconography of "time perspectives" / The chest, wings, and flying skills of the Archaeopteryx and city pigeon have proven themselves over 150 million years / What better way to express *longue durée*?

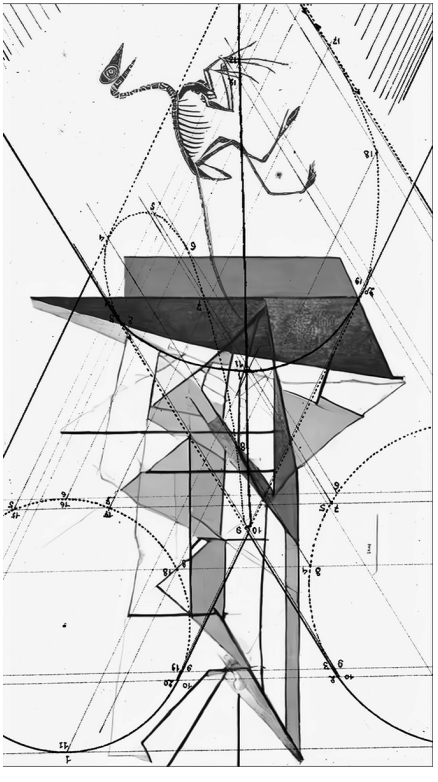


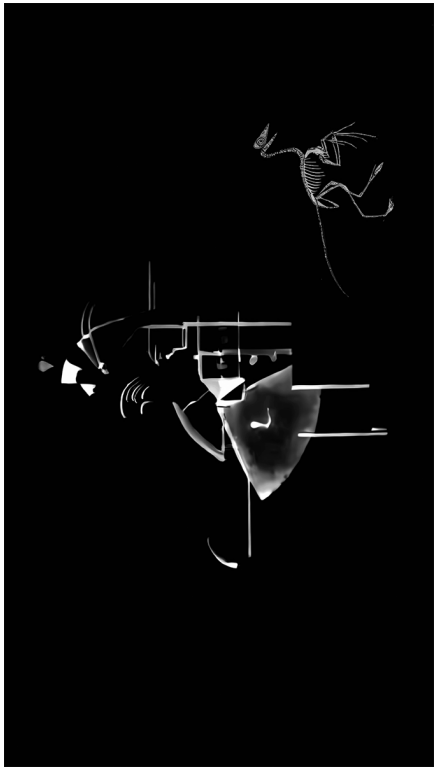


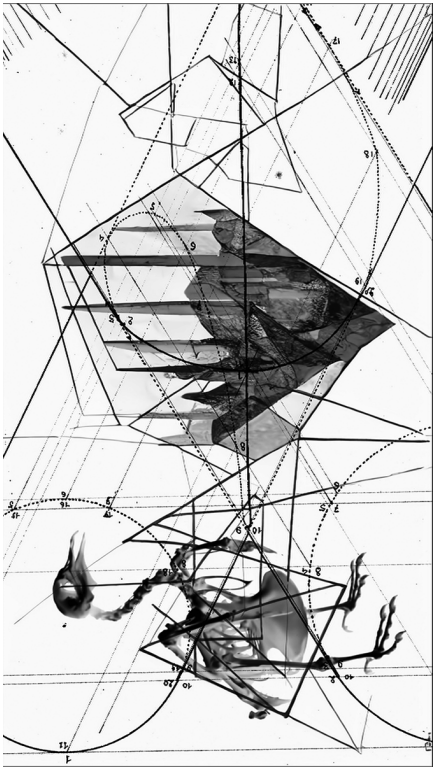


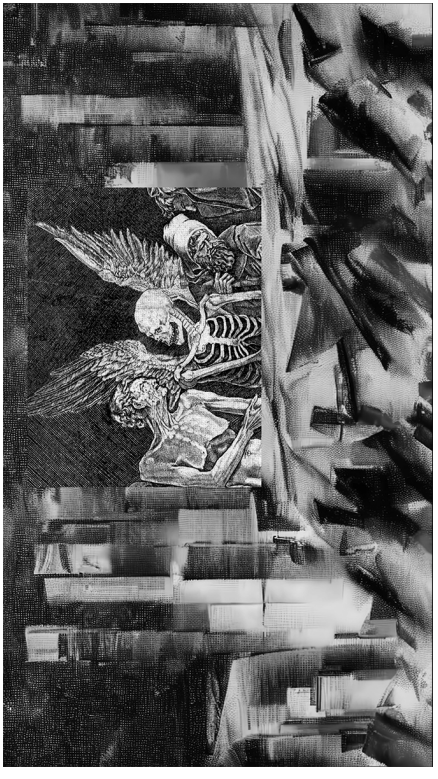












ALEXANDER KLUGE is a writer and filmmaker.

CHRISTOPHER KULENDRAN THOMAS

1. Personally, I'm not too worried about an Artificial General Intelligence extincting us. Not because I think a Skynet apocalypse *can't* happen but because I think there are so many other ways we can be wiped out that we don't worry about, and it doesn't make sense to me to worry about one risk and not the others. Much like how most of us before 2020 weren't worrying about biosecurity around gain-of-function genetic research, I think humans—perhaps like most species—are oblivious to the multiple significant existential risks that we continually face. And I guess I'm more interested in *why* particular risk narratives get dramatized.

2. Playing up AI risk (as the effective-altruism community has done in the West, together with the biggest AI companies) actually benefits Big Tech's AI incumbents, who have the lobbying resources to help governments write regulations requiring expensive compliance processes that will make it difficult for challengers to compete (like with expensive clinical trials for new drugs in the pharmaceutical industry or expensive banking licenses in finance).

3. Since all tech companies will be AI companies, tech might become a regulated industry dominated by today's incumbents. This could bring about a total integration of technology with the state, as with finance and pharma, and spell the end of the era of startup-driven competition. So "e/acc" (effective accelerationism), the strategic-memetic response to that specter of regulatory capture, is (unsurprisingly) popularized by people who invest in startups.

4. Everyone talks their book. Myself included. We get the discourse we incentivize.

5. Many well-known e/acc people are terrified of communism, or at least the twentieth-century version of it; but I reckon what they're rooting for could actually turn out to be a new kind of communalism, a good kind that actually works. The economic system that we currently have is based on scarcity: the cost of scarce labor, the cost of digging scarce fossil fuels out of the ground, the cost of logistics and manufacturing with scarce resources. But an economy driven by autonomous agents and propelled by infinitely renewable energy (eventually even by nuclear fusion) and the ability to synthesize new raw materials—that sounds like something else. I don't know what, but it doesn't sound like capitalism to me.

6. I won't get into the Universal Basic Income debate here. But I will observe that AI could anyway make more and more essential goods and services behave economically more like low-marginal-cost goods and services, with the cost of most things eventually trending closer and closer to zero.

7. My bet is that as markets are freed from a scarcity-based economic model that concentrates power (and therefore future wealth) in the hands of the already wealthy, and as markets are decentralized beyond the control of those ruling elites, markets will produce increasingly communalist economic models based on

collective ownership of the means of production. Why? Because using something that you own (and that you collectively benefit from the success of) is ultimately a better deal than using something you don't own where the success only benefits someone else. My hunch is that the next generation of decentralized organizations/protocols/co-ops-at-Internet-scale will start to outcompete private ownership of the means of production because these kinds of organizations will be able to better align incentives through collective ownership at scale.

8. #FreeMarketCommunism

9. So much of the headline attention right now is on the organizations building the gigantic foundation models, but I can imagine a scenario where open-source foundation models compete the financial value of these models down to zero. A historical parallel would be this: At the time of the Web 1.0 "browser wars," the companies building the interface—Web browsers—were getting all the attention. And of course the Web browser was a crucial technology. But it was ultimately commoditized, and the most valuable businesses of the dot-com era turned out not to be the ones that built the interface (e.g., Netscape) but the ones that figured out the most valuable thing you could do with it (e.g., Amazon).

10. The Web 2.0 era has come up short. The Internet hasn't actually eaten the world. Yet. But it will. Web 2.0 only really "ate" media, communications, business software, and a few other not-so-valuable parts of the economy. But the next Internet will transform the really valuable and important things that we rely on: housing, education, health care, energy, the state. I think these will require a greater degree of trust and therefore probably new kinds of organizations and new models of ownership.

11. Just as the printing press begat the nation-state, so too will the Internet of autonomous agents require new organizational forms. I wouldn't be surprised if, within a year, the Internet was dominated by non-human content and, soon after, by non-human transactions. I *would*, however, be surprised if "the state" as we know it survives this century's platform shift. To take but one tiny example: An economy driven by transactions between autonomous agents would see a massive increase in productivity and a massive increase in transaction volume, and therefore a massive increase in disputes, which would be functionally tantamount to a DDoS attack on existing legal systems. I don't see courts evolving fast enough to keep up, but I *can* imagine a proliferation of new automated legal systems. And similar things could happen in every area of what the state currently does, because there are better ways of constructing the state.

12. It's now becoming possible to talk to the state; not just to a representative of the state but to the state itself—to the collective, synthesized voice of the whole citizenry. Right now I'm really interested in how to build a conversational GPT interface for a deliberation system that could simultaneously mediate negotiations, on

each citizen's behalf, between every member of a community (one-to-one with everyone, collectively) to find the hidden consensus that I suspect is often polarized out of view in multi-party representative democracies. (For example, when the British Empire artificially united Ceylon and left it as a multi-party representative democracy, politicians did what people running for office do under this political system—they dramatized differences and inflamed racial hatred to get elected. This resulted in a civil war in which my family's homeland, Tamil Eelam, was wiped out.)

13. I think we need to get used to—and be OK with—different communities living by their own values, whether we approve of those other communities' values or not. After the Second World War, the United Nations' Universal Declaration of Human Rights defined a particular idea, from the so-called West, of what it means to be human. That idea is a fiction—the ideological front line of an empire—and I don't think it will survive this platform shift. The possibility of artificial general intelligence challenges the cornerstones of that fiction. The legal and ideological framework of human rights defines the individual as the template for the universal. In that sense, it's a fundamentalist logic in that it presumes to impose this value system universally. Many non-Western cultures, however, are rooted in *belonging* rather than individualism, with the family/community as the basic template of society.

14. I have a feeling that the next Internet will not be based on individuals projecting a fantasy of themselves. I think the influencer will be replaced by a different kind of fantasy, the collective hallucination of a community, an "egregore." Our identities will be defined by belonging, defined by all the communities we're a part of. I think we're about to see an explosion of new kinds of communities, held together by new kinds of stories, told in new forms of media, enabling new kinds of spiritualities. Open-source foundation models will enable an infinite proliferation of communities with different values, different belief systems, ones no longer fixed by a single authorized reality. Things are gonna get crazy.

15. The belief system that could *lose* relative mindshare, though, is the one that's prevailed hegemonically in the so-called West, based on seeing humans as ontologically distinct and superior to everything that's not human. Personally, that humanist fiction never felt that real to me anyway. I guess I've felt like a robot for a long, long time. I grew up with a really severe stutter, which I got over when I started doing photography when I was eighteen. "Expressing myself" in another medium somehow gave me the confidence to train myself to speak. But this involved reverse-engineering how speaking works, really studying mechanically how to do this thing that "humans" do "naturally."

16. I use generative-AI tools in my work every day as exactly that, as tools. Ever since I started doing photography, I've looked to painting—I've been kind of obsessed with painting—as an iconic form of self-expression. But making a mark myself in that way has always filled me with the same debilitating freezing-up that I'd experience when

I'd open my mouth to speak and nothing would come out. Using neural networks to compose my paintings allows me to step outside the heroic role of painterly "self-expression" and instead channel something beyond me.

17. My paintings are really pictures of digital files. They're paintings of PNGs that are outputted by a neural network trained on Sri Lanka's colonial art history. This is the art-historical context that really came to dominate on the island after the war that wiped out the Tamil homeland of Eelam. Specifically, I prompt a neural network with the work of my contemporaries in Sri Lanka, some of the most successful painters on the island. The network essentially analyzes the art-historical influences behind their aggregated work and outputs a new image extrapolated from that analysis. And when that resulting image is painted onto canvas by hand, in my studio, the flow of art-historical memes is once again filtered through each of the countless micro-decisions involved in translating image to paint, with thousands of the paintings you've seen over the years subliminally channeled into each seemingly intuitive mark. Perhaps this is not that different from the memetic filtering done by the algorithm in the first place to generate the image.

18. Where does the human end and everything else begin—the network, the ecology, the flow? I've never really thought of my work in terms of originality but rather in terms of metabolizing the circulation of things in a network. It brings me a kind of peace to feel the memes coursing through my veins when I'm connected to the superintelligence of history.

19. My film installations are often algorithmically auto-edited, sometimes with footage that's continually scraped live, sometimes with a generated soundtrack. Parts of my last film, *The Finesse* (2022), were narrated by a generated avatar trained on an inhumanly extensive reading list. What that human-like avatar says, and the footage it narrates, are continually evolving—they're never the same twice. The entirety of my new video work, *Peace Core* (2024), will continue evolving forever. Beyond the field of art, I think conversational interfaces will propel a race to intimacy where the possibility of infinite content and infinite personalization will produce unimaginably deep connections through new forms of media. But we're also going to get a *lot* of very bad AI art.

20. Why is so much AI art so bad? Because making art *about* a technology is boring. Like making paintings *about* paintbrushes is boring. As with any important new technology, the advent of the era of AI has seen a lot of bad art that superficially adopts the aesthetics of the technology, the aesthetics of current AI tools—and where the technology itself is the medium, the content, and the purpose. Art *about* AI—that simply illustrates or fetishizes the technology—misses how genuinely transformational these technologies are, which is both subtler and more profound than the *look* of today's generated content. But thankfully all this AI art, born of the novelty of these technologies, probably won't last.

21. Post-AI art: I think the art that'll have enduring relevance through the era of AI is art that has internalized the widespread ubiquity of AI technologies and the shifts in perspective they bring. And while this "post-AI art" may utilize AI tools—insofar as everyone will utilize AI tools—it need not look like "AI art," because AI technology is more pervasive and foundational than what its current generation of output *looks* like. The era of AI requires art that can look beyond the novelty of the technology and help process the profound and difficult questions we face, because these technologies are bringing about a shift in perception that will transform some of the foundational institutions of our civilization, including our linear conception of time, the historical myths that our nation-states are based on, and the Western idea of the "individual" as the basic unit of society. Like I said, things are going to get very weird.

CHRISTOPHER KULENDRAN THOMAS's exhibition *Safe Zone* is currently on view at WIELS in Brussels.

LEV MANOVICH

My contribution will concentrate on "generative visual media," which refers to the use of AI to create images rather than texts, music, or other forms of media. How should we interpret this new type of visual media in light of the history of visual representation? Humans have historically used a variety of methods to create images of real or imagined scenes, ranging from manual drawing to 3D computer graphics (CG). AI generative media uses a new approach. Computers trained on large datasets of existing visual representations learn how to *predict* new images, both still and moving, that have similar patterns to the training images but are not identical to them. However, this does not mean that these new images are not "original." AI image tools can, via interpolation, generate visuals that include previously unseen subjects, visual styles, and media effects.

One can certainly propose alternative historical paths that lead to visual generative media and/or divide familiar timelines into new stages. Here is one possible trajectory:

1. Creating representations manually (e.g., drawing with a variety of instruments, carving, engraving, etc.). More mechanical stages and tasks were sometimes carried out by human assistants training in their teacher's studio—so there is already some delegation of authorship.

2. Creating representations manually but using assistive devices (e.g., perspective machines, camera lucida). Now some functions are delegated to mechanical and optical devices. From *hands* to *hands + mechanical devices*.

3. Capturing visual and spatial information: photography, X-rays, video, volumetric capture, remote sensing, photogrammetry. From *hands* to *recording information using machines*. From *human assistants* to *machine assistants*.

4. Using 3D computer graphics. You define a 3D geometric model in a computer and use algorithms that simulate the effects of various light sources, shadows, kinds of transparency and translucency, natural textures, depths of field, motion blur, fog, etc. From *recording* to *simulation*.

5. The development of visual generative AI. Using media datasets to predict new visual media. From *simulation* to *prediction*.

AI researchers use the term "prediction" to describe generative-media techniques. While the word can also be used metaphorically and evocatively, this is how generative AI operates in scientific terms. During training, neural networks learn the structure and patterns of training data. In other words, the network learns to predict this data. For example, given the beginning of a sentence, what

might be the next word? Or, when prompted to generate a picture of a face, generative AI uses its knowledge gleaned from analyzing millions of photographs of faces to predict the appearance of a face that has never actually existed.

I am not suggesting that the use of other already accepted terms such as "generative media" is inappropriate. However, using the term "prediction" can help us better understand the distinction between AI imaging and other visual representational methods developed throughout human history. And we can also refer to media generated with AI models as *predictive media*.

After seeing how visually generative (or "predictive") images can be included in the larger history of visual representation techniques, let us now turn to the most recent phase of this history: the development of digital media. Rather than perceiving generative images in the framework of "artificial intelligence" discourse, we may benefit from considering them within the context of the history of digital media.

I want to suggest that a generative image represents a further logical evolution of the process that began with digital-media algorithms in the 1970s and continued in the following decades. The first computer paint programs were created in the 1970s but could not yet simulate different paint types, brushes, and textured surfaces such as canvas.[1] But in the 1990s, software such as Coral Painter (1991–) started to offer these features. Similarly, though the first 3D-computer-graphics algorithms used for rendering solid shapes, "Gouraud shading" (1971) and "Phong shading" (1973), couldn't simulate the looks of different materials, in the 1970s and '80s, computer-graphics researchers created numerous algorithms to simulate the appearance of various materials and textures, such as cloth, hair, and skin, as well as shadows, kinds of transparency and translucency, depth of field, lens flares, motion blur, reflections, water, smoke, fireworks, explosions, and other natural phenomena and cinematographic techniques and effects.

The continual development of these algorithms is evident in the structure of SIGGRAPH, the leading yearly conference for researchers in computer graphics. The conference comprises numerous sessions, each specifically focused on a distinct topic in graphics simulation, such as Volumes and Materials, Fluid Simulation, and Cloth and Shells. During each session, researchers introduce several novel algorithms developed over the previous year to improve visual simulation of particular aspects of visual reality.[2]

In my 1992 article "Assembling Reality: Myths of Computer Graphics" I analyzed this fundamental aspect of computer graphics, explaining that "synthetic

1. On the history of early paint programs, see Alvy Ray Smith, *Digital Paint Systems: An Anecdotal and Historical Overview*, IEEE Annals of the History of Computing, vol. 23, no. 2 (2001), and also his *A Biography of the Pixel* (Cambridge, MA: MIT Press, 2021).

2. See *SIGGRAPH '22: ACM SIGGRAPH 2022 Conference Proceedings*, https://dl.acm.org/doi/proceedings/10.1145/3528233.

photorealism is fundamentally different from the realism of the optical media, being partial and uneven, rather than analog":

> Digital re-creation of any object involves solving three separate problems: the representation of an object's shape, the effects of light, and the pattern of movement. To have a general solution for each problem requires the exact simulation of underlying physical properties and processes. This is impossible because of the extreme mathematical complexity. . . . In practice, computer-graphics researchers have resorted to solving particular local cases, developing a number of unrelated models for simulation of some kinds of shapes, materials, and movements.[3]

In other words, 3D CG takes apart the world that we see, separating out objects' shapes, materials, light reflections, textures, movements, and behaviors. During rendering, the effects of multiple algorithms' simulating all these aspects are combined. Thus, visual representations created using CG are discrete and modular, rather than continuous and "monistic." This is one of the most important characteristics of the CG medium, distinguishing it from lens-based optical-image media.

This logic of separation and recombination also defines the next stage of digital media: the use of PC software for media creation and editing. Following its initial release in 1990, Photoshop gradually began to include effects and techniques from various artistic mediums, ranging from darkroom photography to oil painting, within a single program. These effects can in turn be combined in a single digital image. Music software similarly allows users to combine many simulated instruments and multiple effects, such as reverb and echo, in a single composition. Word processing and desktop-publishing software separate the physical process of print composition into elements that can also now be recombined—for example, you can take any font and arbitrarily change its size.[4]

All of these media-software capabilities were proposed in the 1970s and realized in the '80s and '90s, eventually becoming ubiquitous. Generative media follows the same logic, although its underlying technical implementation is different. During training, neural networks learn visual patterns characteristic of hundreds of different types of art media, lighting techniques and effects from the history of photography and cinematography, and visual signatures of many thousands of historical and contemporary artists, architects, fashion designers, and other creators. For example, a reference website called Midlibrary currently lists 378 "artistic techniques" that the popular Midjourney image-generator tool can reliably simulate,

3. Lev Manovich, "Assembling Reality: Myths of Computer Graphics," *Afterimage* 20, no. 2 (September 1992), pp. 12–14. See also Lev Manovich, *The Language of New Media, Part 4: The Illusions* (MIT Press, 2001).

4. For the detailed analysis of media software and its conceptual origins, see Lev Manovich, *Software Takes Command* (Bloomsbury Academic, 2013).

according to tests conducted by a website team.[5] They range from albumen prints and anaglyphs to wood carving and wireframe rendering.

Importantly, I can include references to multiple techniques and/or multiple creators in a single prompt, potentially generating new types of media effects that did not exist before. The pioneering digital-media theorist of the 1990s and 2000s William J. Mitchell called this key characteristic of digital media the ability to "separate and recombine."[6] In his 1995 book *City of Bits*, he described this process in relation to urban planning:

> Classical architects of the eighteenth and nineteenth centuries handled the task of putting spaces together by creating hierarchies of great and small spaces around axial, symmetrical circulation systems connected to grand, formal entries and public open spaces. With the aim of being as logical and efficient as possible, functionalist modernists of the twentieth century have often derived their less regular layouts directly from empirically established requirements of adjacency and proximity among the necessary spatial elements. But when telecommunication through lickety-split bits on the infobahn supplements or replaces movement of bodies along circulation paths, and when telepresence substitutes for face-to-face contact among the participants in activities, the spatial linkages that we have come to expect are loosened. The constituent elements of hitherto tightly packaged architectural and urban compositions can begin to float free from one another, and they can potentially relocate and recombine according to new logics.[7]

In lectures Mitchell gave in the 2000s I heard him expand on this formulation, demonstrating how the logic of separation and recombination can be seen in digital media in a variety of ways. Generative AI relies on the same logic. A neural network extracts elements and structures from hundreds of millions or even billions of images in its training set. They include distinct color palettes, compositions, lighting effects, artifacts of historical photography processes, and so on. When you ask a generative image tool to generate new images with specified visual attributes, it does its best to combine (or more precisely, *interpolate* between) appropriate aesthetic patterns and structures.

In this regard, we can say that the artificial neural networks that power today's image-generative AI tools continue and expand on the visual-arts decomposition and analysis programs that began in the first two decades of the twentieth century. These programs were carried out by both the founders of modern art history and numerous modern artists. In art history, Aby Warburg and Erwin Panofsky developed the study of iconology. Warburg defined this concept as visual

5. https://midlibrary.io/February 25.

6. https://mitpress.mit.edu/author/william-j-mitchell-2911/.

7. William J. Mitchell, *City of Bits: Space, Place, and the Infobahn*, revised ed. (Cambridge, MA: MIT Press, 1996).

motives that (re)appear in various civilizations and media. Panofsky used it somewhat differently, referring to symbols and motifs that have existed throughout the history of art.[8]

During the same historical period, modern artists and designers disassembled visual arts in a different way, breaking down images into their basic components and dimensions such as points, lines, planes, two-dimensional forms, color, space, texture, pattern, balance, and equilibrium, among other things. While this project of methodical dismantling and creation of new visual languages from these components was central to modernist art and its many -isms, it arguably found its most methodical development in the curricula of two cutting-edge schools of art and design. VKhUTEMAS in Moscow (1920–1929) and the Bauhaus in Germany (1919–1933) both featured "basic courses" in which students were taught how to systematically work with all the relevant elements and dimensions. Instead of drawing from life, painting portraits, or making historical compositions, students started their training by completing exercises with image primitives such as basic shapes, forms, and colors.

It is possible to say that today generative artificial intelligence is carrying on the programs of decomposition and the analysis of the visual arts that were initiated in the early twentieth century. The networks process billions of images during their training, effectively "learning" to recognize and reproduce a wide range of visual elements and patterns. This includes every aspect of images, including composition, representation of figures, faces, and other objects, lighting techniques, perspective, and stylistic elements. In a sense, they are performing a highly sophisticated form of visual analysis, breaking down images into both basic formal features and more complex representational elements. The AI learns to recognize how these components interact and contribute to the overall visual structure and meaning of an image.

When generating new images, AI tools create new visual content by combining these learned elements in novel ways. This process mirrors (on a much larger scale and in an automated fashion) early-twentieth-century efforts to deconstruct and understand the fundamental components of visual art. Just as art historians cataloged motifs and artists explored basic visual elements and dimensions, AI systems create internal representations of diverse visual patterns and principles.

However, at least at present, we can't directly look at hundreds of billions of parameters in a large artificial neural network and see a neat catalogue of all the patterns a network has learned.[9]

8. Erwin Panofsky, *Studies in Iconology: Humanistic Themes in the Art of the Renaissance* (Oxford University Press, 1939).

9. See, for example, Dustin Podell, Zion English, Kyle Lacey, Andreas Blattmann, Tim Dockhorn, Jonas Müller, Joe Penna, Robin Rombach, "SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis," arxiv.org, July 4, 2023, https://arxiv.org/abs/2307.01952.

I want to conclude with a relevant quote from my 2018 book *AI Aesthetics*.[10] While at the time deep neural networks were mostly used for media classification and recommendations, with the generative-AI revolution still four years away, the analysis I developed in the "AI as a Culture Theorist" section is quite a bit more relevant today:

> [There is] a crucial difference between an "AI culture theorist" and a human theorist/historian. The latter comes up with explicit principles that describe how a cultural area functions. . . . [A] neural net can be trained to distinguish between works of different artists, fashion designers, or film directors. And it can also generate new objects in the same style. But often we don't know what exactly the computer has learned. . . . Will the expanding use of machine learning to create new cultural objects make explicit the patterns in many existing cultural fields that we may not be aware of?

This theoretical potential is one of the most intriguing and valuable aspects of generative AI in my opinion; however, we will have to wait and see if it is realized in the future.

LEV MANOVICH's most recent book is *Cultural Analytics* (MIT Press, 2020).

10. Lev Manovich, *AI Aesthetics* (Strelka Press, 2018), http://manovich.net/index.php/projects/ai-aesthetics.

TREVOR PAGLEN

You've seen the pictures. The swagged-out pope sporting a puffy Balenciaga coat. Donald Trump grimacing as a throng of NYPD officers attempt to arrest him. A submerged centaur—half-Jesus, half-shrimp—beaming through an aquatic cyan landscape. These viral AI-generated images' mimetic qualities are fueled by sly semiotic contradictions, wish fulfillments, and quasi-surrealistic absurdities. But curiously, the theory of images underlying the generative-AI models used to create these images has almost nothing to do with culture, representation, semiotics, or iconology. Generative media is founded on a theory of visual perception unfamiliar and alien to those of us raised on Saussure, Panofsky, Barthes, Sontag, and the like.

Generative AI is based on a theory of correspondence between visual stimuli and neurological patterns,[1] where concepts are, at most, an afterthought. It's a theory—let's call it "Neural Activation Theory"—that emerges from decades of work by neuroscientists and their ilk and whose origins come from a sequence of experiments wherein said neuroscientists did, among other things, some terrible things to kittens.

In this short essay, I want to sketch out the history of this neurological theory of images, show how it converges with visual culture and image-making in the era of generative media, and point towards some strange directions in which it is starting to go.

Our story begins in 1959, when two early neuroscientists undertook a study of visual perception. David Hubel and Torsten Wiesel wanted to answer the following question: How do cats—and by extension humans—see? At the time, nobody knew the relationship between photons entering our eyes and the information and meanings we derive from those stimuli. Do we perceive objects as fully formed wholes? Does the brain create 3D models of objects and compare visual stimuli to those innate models, looking for a "best match"? Or does visual perception entail some entirely different process? Moreover, what sort of experiment would give insights into this question? The duo devised a plan: split open a cat's skull, attach electrodes to its brain, and show the cat a bunch of pictures while recording its brain activity.

They began with a hypothesis: "Let's show the cat a bunch of pictures of fish. The cat should get excited, and we'll record the neural activations of that excitement."[2] So they started showing pictures of fish to the cat, their sensors ready to record the neural fireworks they assumed would ensue. But . . . no dice. No activations. The cat's brain didn't seem to care. But then they noticed something else. Every time they changed the slide on the projector, a moving black line would appear in the projection as the slide progressed. When they did this, the cat's neu-

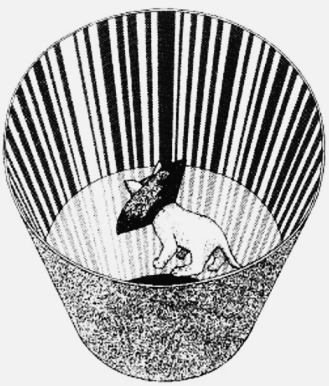
1. I'm going to be using the phrase "generative AI" a lot in this essay, but I'm just focusing on generated images. That isn't to say that my argument isn't relevant to other generative media (I think it is).

2. This isn't an actual quote.

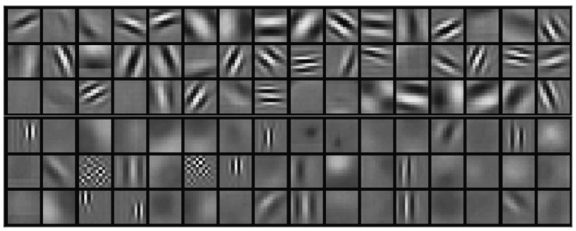
AI-generated fish.

rons lit up. The cat's brain was reacting to the movement of the slide's edge as the image changed in the projector. This insight was the first step towards a theory of vision that would eventually yield insights making computer vision and image-based machine learning possible. Hubel and Wiesel's observations would eventually lead to a Nobel Prize.

New researchers took up the project. In the early 1970s, Colin Blakemore tried to further isolate the mechanics of vision, also by doing terrible things to kittens. Blakemore's experiments involved raising kittens in environments where they'd be exposed to an extremely limited range of visual stimuli. In one series, this meant raising kittens in an environment where the only visual stimulus they received consisted of horizontal stripes presented to them in a container (when the cat was outside the container, it was kept in complete darkness). For the first few months of the kitten's life, it never saw a vertical line. When Blakemore and his team eventually took the kitten out of their constructed environment, they found that it could not perceive vertical lines. The cat's brain had not developed a neurological response to vertical lines, and when the kitten encountered vertical lines for the first time, it could not see them.

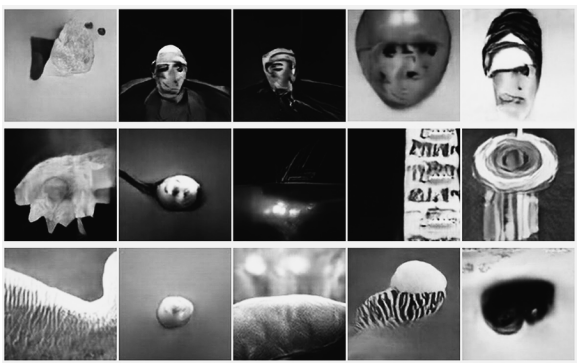


This series of experiments in visual perception, again, had profound implications for a neurological theory of visual perception. The experiments implied that visual perception works hierarchically by developing complex representations from crude building blocks. We don't just "see" a fish, the logic goes, we subconsciously build up the visual concept of a fish from different assemblages of visual "primitives": edges, gradients, and other basic shapes.



Basic neural "primitives" synthesized from Deep Layers in a neural net.

The emerging theory suggested that images work in a way that's analogous to the mechanics of language: Every word is made of letters. Different combinations of a relatively small number of letters give rise to an infinite number of possible words. Words, in turn, can be formed into sentences, paragraphs, and so on.



Complex neural "primitives" synthesized from Deep Layers in a neural net.

The neural theory of visual perception implied that something similar happens when we see an image: Our brain perceives a collection of visual primitives. Those visual primitives activate a collection of neurons in our brains. The particular pattern of activated neurons corresponds to something we attribute meaning to: a fish, an orange, or the face of our mother. In this conception, when

we "see" an image, our brains are conducting a kind of magic trick—our perception of a fish is the quasi-hallucinated experience of the particular arrangement of neurons being activated.

This neurological theory of vision explains a few things. It explains why, for example, if you draw two dots next to each other and an upward- or downward-tilting arc below them, we perceive either a smiley face or a frowny face. Three of the simplest visual primitives are all we need to perceive a face. This is also why we might see the face of Jesus on slices of toast, sheep in clouds, UFOs, and so on.

The neurological theory implies that everything we perceive can be broken down into combinations of primitive shapes and that every object in the universe can be represented by a particular combination of them. In the same way that you can create any word, sentence, or novel from a small collection of letters, you can create an image of anything from a collection of visual primitives.

The theory of vision that Hubel, Wiesel, Blakemore, and others outlined obviously has a lot of other implications, but one big relevant one for us was the implication that vision could be quantified.

If all images—and by extension all objects—could be broken down into their component parts, and if you could represent those component parts digitally, then you could theoretically create computer-vision algorithms that replicated that process—you could build computers to "see" the world. All you'd need were two things. First, you'd need a database with labeled pictures of all the objects in the universe. Second, you'd need an algorithm that could take that database, break down all of those images into their component parts, and figure out which patterns of visual primitives corresponded to which objects. In the 1980s, both of those propositions were science fiction.

Sitting here in 2024, one can see that those criteria have sort of been met. The database is called "the Internet" (more specifically, massive datasets like LAION-5B), and the algorithms used to break down the images are called "neural nets." You can give a picture to a neural net and it will tell you what it "sees" in the picture. (There is a massive side-quest here that we won't go into about all the problems with this ridiculously literal, or "machine realist," conception of visuality.)

In the past few years, we've seen that object recognition isn't the only thing you can do with an implementation of this theory. You can implement the theory to synthesize images. Take the neural network you'd built to classify objects and run it backwards. Instead of using the neural net to look at a picture, break it into its component parts, and tell you what object the pattern or primitives correspond to, you can have it draw those patterns into new, entirely synthetic images. This is known as generative AI.

Let's return to the kittens. Remember that the researchers didn't really care about the "meaning" of images—they cared about how different stimuli activated different brain patterns. In this conception, an "image" is a name for the relationship between a particular visual stimulus and a particular pattern of excited neurons. In this conception of visual perception, images operate preconsciously, bypassing, in the first instance, language and "concepts." Concepts, or signifieds, are a second-order effect—a name that squishy humanities people give to self-similar arrangements of activated neurons.

It turns out that, outside the confines of cat-torture chambers, our relationships between visual stimulus and perception are messy af. Culture, history, memories, trauma, joy, traditions, expectations, and emotions all play a role in collective and individualized perceptions. Visual perception is charged with fear, joy, flight, thirst, curiosity, wonder, anxiety, titillation, and the like. And those affects—in the neurological theory—are also neural activations in our multi-modal brains.

We are now the kittens. We're currently in the horizontal-striped box of a great experiment that uses generative AI, recommendation algorithms, and all sorts of sensors (from cameras and accelerometers to "like" buttons and engagement statistics) to synthesize media designed to maximally activate patterns in our brains to achieve the desired neurological response.

To see where this is going, consider the Nacho Cheese Dorito: a synthetic food optimized to maximally activate the pleasure-seeking regions in our taste-bud brains. But instead of everyone eating the same Doritos, imagine everyone having specific Doritos synthesized for them—a Dorito that is not only capable of maximally activating pleasurable tastes (i.e., sweetness, savoriness, etc.) but that can learn to modulate those tastes to the specificities of each Dorito eater's neurological makeup to activate patterns associated with all sorts of things: pleasurable memories from childhood, the excitement of an overseas trip, or the scent of an intensely missed lover.

We're talking about images and media that are more or less directly connected to your brain and that can be evolved to stimulate various neural patterns in order to influence you. I think about this as a kind of "PSYOPS Media." And it's about to get a whole lot weirder. But that's another conversation.

Anyway, it might turn out that Plato was onto something with his suspicion about where this whole business about synthetic images was going.

TREVOR PAGLEN is an artist.

CHRISTIANE PAUL

In May 2023, more than 350 executives, researchers, and engineers working in AI signed an open letter released by the nonprofit organization Center for AI Safety declaring that mitigating the risk of human extinction from AI should be made a global priority. The statement crystallized various tensions surrounding AI technologies at a pivotal moment in their evolution. From 2021 to 2023, several large-language models (LLMs), allowing processing of vast textual datasets via AI, and text-to-image models, generating visuals by means of natural-language prompts, had been made publicly available and had an impact on a broad range of sectors, from commerce and politics to art and entertainment. The tensions that lay the foundation for considering AI as a global societal risk are complex and have a long history. They include the dichotomy of seeing technology as the carrier of either human salvation or human extinction; the question of assigning responsibilities when it comes to establishing guardrails for technological developments; and the challenges in defining sentience, intelligence, and, consequently, what it is to be human. Over the decades, AI art has critically addressed these tensions and challenges in relation to AI technologies as they have emerged, testing and modifying systems and highlighting their characteristics and inherent values.

It is notable that the signatories of the open letter included top executives from leading AI companies, including Sam Altman, chief executive of OpenAI; Demis Hassabis, chief executive of Google DeepMind; and Dario Amodei, chief executive of Anthropic, founded by former members of OpenAI. The very companies that had developed the technologies they were cautioning against had decided to release and sell them, despite their own concerns about their implications and scalability. This warning, therefore, needs to be considered as a calculated move in both positioning the companies as the potential savior and negotiating responsibilities for AI's potential consequences by asking for governmental and legislative regulations.

Artists have used AI to experiment with its potential and impact since at least the early 1970s, fifteen years after the field was formalized at the 1956 Dartmouth Summer Research Project on artificial intelligence. They have probed its creative possibilities and engaged with its ethics and biases, as well as its effects on ecologies and labor, often by developing their own hybrid models and architectures.

Artistic practice in the field gained new momentum with the launch of OpenAI's large-language-model-based ChatGPT, which launched on November 30, 2022, and text-to-image tools such as OpenAI's DALL-E (2021) and DALL-E 2 (2022), Midjourney Inc.'s Midjourney (2022), and Stability AI's Stable Diffusion (2022). The hype surrounding text-to-image models immediately led to a polarized discussion, with the claim that AI would replace artists for good, on the one hand, and the dismissal of these tools as insta-kitsch engines that couldn't produce

anything of aesthetic value, on the other. While the former position seems to lack the aesthetic vocabulary necessary for evaluating art, the latter ignores artists' more sophisticated engagement with these AI tools. Both tend to reduce AI art to visuals created by means of simple text prompts. To center the conversation, one needs to consider the evolution of AI art and trace the shifts in artists' approaches to collaborating with, changing, torquing, and/or critiquing AI systems. At the core of such art lies the ability of humans and machines to acquire and apply skills and knowledge, raising the question of what the encoding of "intelligence" might mean for being human. In more recent years, artists also have looked beyond the relationships between *anthropos*, computer hardware, and computer software to consider how AI might move beyond anthropocentric models for knowledge creation. In his exhibition *Distributed Consciousness* (2024) at Gallery QI in San Diego, artist and creative technologist Memo Akten, for example, uses the cognition of cephalopods, which have the majority of their neurons distributed across their body rather than located in a central brain, as an inspiration for an AI-generated manifesto, juxtaposing decentralized nature with synthetic intelligence.

The term "AI art" is commonly understood to designate art that employs AI technologies in its creation, but this conception warrants further scrutiny. It fails to make distinctions between the use of AI as a tool as opposed to a medium and neglects the art's conceptual engagement with AI. AI art can be defined as a subcategory of digital art or computational art, which uses digital tools and media to create and contextualize artworks: It incorporates technologies of artificial intelligence as both tool and medium, engaging with them both practically and conceptually. The employment of a simple text prompt to generate a visual by means of a corporate text-to-image software does not automatically turn the resulting image into AI art. The use of AI as a medium—engaging its inherent systems and characteristics—in ways informed by a conceptual approach distinguishes a work as AI art. Computational art and its aesthetics are established and ever-evolving fields. As a sub-category of generative computational art, AI art requires a continuous reassessment of its models and expression. The following will focus on key moments in the history of AI art, tracing how it has developed in the context of technological developments and investigating the potential of art in general to contribute to the critical discourse that has developed around the aesthetics—as well as the cultural, socio-political, and ethical impact—of AI technologies.

Collaboration is at the core of the earliest artificial-intelligence program for art-making—one of the longest-running ongoing projects in contemporary art, Harold Cohen's (1928–2016) *AARON*. An established British painter, Cohen began exploring the potential of software for art-making in 1968, when he became a visiting lecturer at the University of California San Diego. He officially named his program *AARON* in 1973, after being invited to the Artificial Intelligence Lab at Stanford University. Contrary to today's statistical AI, trained on large datasets of images, *AARON* was symbolic AI that encoded "knowledge" about drawing and

composition on the basis of rules that Cohen wrote in programming languages. *AARON* does not entail any of the standardization, averaging, and optimization used in the current models, which have been trained on massive datasets of existing images. Instead, *AARON* was shaped by the aesthetics of Cohen, who kept developing the software until his death and experimented with shifts in the style of work—from simple evocative shapes to figures and jungle-like environments—and a move from monochrome to color output. From the 1970s to the '90s, Cohen built his own drawing and painting machines that plotted and painted *AARON*'s creations, and in the 2000s he switched to purely screen-based presentations.

After the first artistic explorations of AI represented by *AARON*, AI art evolved in three major phases from the 1990s to the early 2020s, shifting its focus to explore and investigate technological developments as they emerged.

In the 1990s and early 2000s, artists created AI systems that critically engaged with the emergence of software agents—applications running automated tasks on the Internet for filtering or imitating humans—as well as chatbots. The problematic aspects of the software-driven filtering of information and encoding of human communication, as well as the "personality" of bots, became active areas of exploration. Lynn Hershman Leeson's chatbot *Agent Ruby*, released in 2000, explored chatbots as essentially social beings—autonomous characters with a life of their own—while Peggy Weil's *MrMind* (1998–2014) was created with the specific intent of highlighting the differences between humans and machines. Rebecca Allen's *The Bush Soul* series (1997–99) explored communication between users and autonomous creatures in a virtual environment and laid the groundwork for more recent projects involving AI-driven life-forms, such as Ian Cheng's artworks.

The 2010s saw a shift in artistic AI practice that responded to a new stage of big-data analysis and neural networks—which themselves originated as far back as the 1920s and now were benefitting from big-data processing—as well as the emergence of generative adversarial networks (GANs) and generative pre-trained transformers (GPTs). Artists increasingly addressed the biases in big datasets and the ethical issues resulting from algorithmic processing. The socio-political dimensions of pattern recognition and *apophenia*, the perception of a meaningful pattern among unrelated or random things,[1] played a major role in critical discourse and art exploring large-scale datasets. Stephanie Dinkins decidedly countered the idea of the benefits of big data with a decidedly small dataset and examined AI in the context of race with *Not The Only One (N'TOO)* (2018), a sculptural AI storyteller trained on data supplied by three generations of women from one African American family, drawing attention to a drastically underrepresented dataset. Dinkins also explored datasets and the process of an AI's learning in *Conversations with Bina48* (2014–present). The work documents the artist's ongoing conversa-

1. See Hito Steyerl, "A Sea of Data: Apophenia and Pattern (Mis-)Recognition," *e-flux Journal* 72 (April 2016), https://www.e-flux.com/journal/72/60480/a-sea-of-data-apophenia-and-pattern-mis-recognition/.

tions with Bina48 (whose name derives from "Breakthrough Intelligence via Neural Architecture, 48 exaflops per second"), an intelligent "social" robot modeled after a Black woman and built by the Terasem Movement Foundation. *Conversations with Bina48* explores what identity, race, and kinship mean to an artificial intelligence and whether we can form sustained relationships with the increasing number of non-human entities surrounding us.

While AI art's engagement with GANs and GPTs can vary significantly in its focus, the key issue with these forms of statistical AI is always the automation of image and text generation through datasets with their embedded biases. GANs are learning neural networks in which generative algorithms trained on a specific dataset generate new original images with the same characteristics as the training set and are then evaluated by discriminative algorithms that, based on their own training, judge whether the newly produced data looks authentic. GANs in particular led to an explosion of art projects exploring their potential for image generation on the basis of specific training sets and aesthetic goals. Less successful works stayed on the level of the "imitation game," probing GAN aesthetics and the capabilities of AI software to reproduce images in a familiar period style. By contrast, Mary Flanagan's *[Grace:AI] Origin Story* (2019) focuses on the aesthetics of using a deliberately gendered dataset. Flanagan trained a GAN on thousands of images of paintings and drawings by female artists only, then tasked the software to create its "origin story" by looking at twenty thousand online images of Frankenstein's monster and producing its portrait. *[Grace:AI]* both alludes to Mary Shelley's feminist critique of artificial life and male-dominated creation in *Frankenstein* and explores whether a gendered training dataset produces a distinctive style.

Artists not only investigated the aesthetics of AI tools, they also started responding to the rise of paradigms of environmental management, engineering, and strategic intervention. Tega Brain's *Deep Swamp* (2018) humorously critiqued environmental optimization in the form of a triptych of semi-inundated environments of wetland life-forms governed by artificially intelligent software agents with different goals; while *Asunder* (2019), by Brain, Julian Oliver, and Bengt Sjölén, tested the potential benefits and pitfalls of an AI-controlled, fictional "environmental manager" that proposes and simulates future alterations to the planet.

Starting in 2021, AI and AI art entered the mainstream with the launch of ChatGPT and text-to-image tools. The latter models use vast datasets of images with associated text and, by means of deep-learning methodologies, generate digital images with different styles and attributes via users' text prompts, a.k.a. natural-language descriptions. While non-specialist media outlets debated whether these capacities would render human artists obsolete, actual digital-art practice revealed both the flaws and potential of AI tools, highlighting the intense labor and rigorous processes required to create sophisticated works. The text-to-image phase of

AI art has arguably constituted the biggest shift so far, wherein tools such as DALL-E, Midjourney, and Stable Diffusion make images subordinate to language classification. These tools deeply fuse visuals with a lexical register, drawing on pre-existing dependencies. They produce their visual output on the basis of the textual classification of the training set and source data. Their visual creations are then determined by the alignment of users' prompts with the text pre-associated with images, thereby building output on layers of existing taxonomies. Artists have begun exploring the potential and problematic aspects of these new semantic frameworks in medium-specific ways, assessing their impact on painting, photography, and film. Bennett Miller, for example, established parallels between the dawn of photography and the early days of text-to-image models' transformative power in his 2023 exhibition of prints at New York's Gagosian gallery. Occupying a delicate threshold between the familiar and uncanny, the images remain eerily detached from a graspable subject. They are suspended in an alternate reality that both captures the essence of a distinctive stage of AI and highlights the differences between photographic processes and images generated in a photographic style by text-to-image models. While today's AI can be seen as the result of a long evolution of "machine learning" rather than a new kind of technology, it also radically questions traditional definitions of media forms.

The focus of artists' engagement with text-to-image AI programs covers a range: Some use the software as more of a tool in the creation of projects that rely on multiple digital technologies in their creation process; others make it a focal point of aesthetic and conceptual explorations. While discourse about AI art in the mainstream media has focused on the dangers of artists' being replaced by AI "creators," the artists critically addressing AI technologies have been investigating the problems that the inherent classification, standardization, and optimization of AI tools pose to creativity owing to their normative foundation. Text-to-image models use trillions of existing images to which corporations have access—many of them stock images—which means that they are operating within an echo chamber from the start. The text associated with these images was often originally created for marketing purposes, inscribing a specific agenda. Artists and other creators are currently not compensated for the use of their images, while the text-to-image tools generate an increasingly diluted version of their style. According to data published by AI Secrets in late 2023, AI generates roughly 34 million images per day,[2] which then feed back into the training datasets, bringing us ever closer to the state of "model collapse," where AI will be trained only on images of its own creation.

Digital art has always been at the forefront of engaging critically with the technologies it employs, and AI art can play a crucial role in assessing the aesthetic and socio-political impact of the tools that are shaping our future. Public discus-

2. AI Secrets, Visual AI stats, https://aisecrets.com/applications/visual-ai-stats/.

sions as well as those in the industry are frequently resorting to polarized narratives of extinction versus salvation, often driven by a commercial agenda or simplistic understandings of intelligence and sentience. AI art often provides a much-needed reality check, challenging facile assumptions and adjusting or breaking systems to raise more profound questions about human and other forms of intelligences.

CHRISTIANE PAUL is curator of digital art at the Whitney Museum of American Art.

KRIS PAULSEN

Writing in 1980, Jack Burnham wondered, "Why should the only successful art in the realm of twentieth-century technology deal with the absurdity and fallibility of the machine?"[1] Only Jean Tinguely's malfunctioning and self-destroying sculptures, he wrote, "have maintained their status through the 1970s." Tinguely's *Homage to New York* (1960), an infamous and intentional disaster, shuddered and convulsed until it disassembled and caught fire in the sculpture garden at the Museum of Modern Art, alarming museum officials and delighting the artist. It was a metaphor for the city and its energy, but it was also something of a dumpster fire, thrilling in all its titillating threats to personal safety and to the institution. The prompts of this *October* questionnaire have me thinking back to *Homage* and Burnham's assessment of art and technology artworks. *Homage* was central to the emergence of art and technology as a visible field of artistic and critical inquiry, serving as the meeting point for Bell Labs's Billy Klüver and Robert Rauschenberg, who later launched E.A.T. together. And Burnham's question is still relevant today: Why do we so enjoy artworks that revel in a machine's failure?

The machine-learning artworks that attracted my interest and that of so many other historians and critics in the late 2010s and early 2020s were, as Burnham would predict, those that seemed to revel in the absurdity and fallibility of such systems—their ingrained biases, skewed datasets, nefarious intended uses by the corporations and states that developed their conventional forms, and their struggles to perform with proficiency. Robbie Barrat's early GAN artworks, for example, provided just this sort of delight in failure. Trained on a dataset composed of paintings of nudes scraped from WikiArt, Barrat's *Nude Portraits* series (2018) produced horrific, monstrous, near-human images. In these "paintings," piles of peachy flesh appear folded and heaped on velvet divans and in front of lush drapery—breast, elbow, ass all a jumble. Barrat designed the network so that it could "correctly learn 'rules' associated with small and local features of paintings (breasts, folds of fat, etc.) but [the generator algorithm] failed to learn rules concerning the overall structure of the portraits (2 arms, 2 legs, 1 head, proportions, etc.)."[2] The image generator could not sort out the parts of the body or how to distinguish one person's body from another's or what a face is and where on the mass of parts it should be, but the discriminator didn't mind, and certainly neither do I. Their misalignments with representation and realism, with human vision and human conventions, made them very abject and simultaneously very appealing. There was a thrill in imagining that this might be how a computer would see and understand the human form, all bulbous globs of mottled flesh. Barrat's images

1. Jack Burnham, "Art and Technology: The Panacea That Failed," *Video Culture: A Critical Investigation*, ed. John G. Hanhardt (Layton, UT: Peregrine Smith Books, 1986), p. 233.

2. Robbie Barrat, "Old Work—Landscapes and Nude Portraits," https://robbiebarrat.github.io/oldwork.

instill the horrible sense of being seen through alien eyes, but, at the same time, provide a distilled view of the history of art, encapsulating its biases and exclusions: The forms are horribly vulnerable, vaguely female, and always very white. Through these alien eyes one might see Western culture quite clearly. As with *Homage*, it is not just that the machine fails, but that it turns an ugly truth into an aesthetic pleasure and visual joke. Seeing it from the outside enables one to feel in the know and perhaps avoid feeling so implicated in what it reveals.

A counterexample to Barrat's work could be Refik Anadol's enormously popular AI spectacle *Unsupervised—Machine Hallucinations—MoMA* (2022), which uses a GAN to generate possible artworks based on MoMA's collection, hypothesizing a potential future that looks just like the past and providing endless slippery glimpses of white modernism and colonial appropriations. It is a narrow history presented as a dream of novelty and innovation. Anything is possible as long as it has already happened. It is a document of a civilization and therefore, as Walter Benjamin observed, it is a document of barbarism.[3] But the barbarism here sells, and nested within it is another: The hook of the work is not the endless almost-paintings but Anadol's ASMR-triggering rainbow-colored animations of little viscous bubbles, which opaquely visualize various forms of environmental surveillance to which the museum's visitors have unknowingly consented. It uses the barbarism of the past to acclimate viewers to the offenses of the future. It is a "successful" work that is not about the absurdity and fallibility of the machine. It is slick, smooth, and seamless in its seductions.

It makes sense that we would find artworks that model and spectacularize the failure of technology so appealing. They open up space for critique, which helps us slow down in an era of moving fast and breaking things, but they also allow us to feel powerful and superior in relation to the developments in automation that constantly appear to threaten that hierarchy. Barrat's GAN produces "bad" images, and in the distance between them and human capabilities one can feel comfortable and secure. One of many things that seem to have changed since around 2022, with the release of diffusion-model image generators like DALL-E, Midjourney, and Stable Diffusion, as well as text-generation algorithms using transformer models like ChatGPT, is that the output of such systems is suddenly reasonably good. Like Anadol's work, they function smoothly and successfully, doing a good job at mimicking our expectations and productions. The texts appear competent, though sometimes full of misinformation, and the images are well executed and often novel, likely the result of their expansive datasets and skillful prompt engineering by the user. They are not "failing" in the same way that Barrat's images or Tinguely's machines intentionally do, though of course it is still debatable whether we could call their output "good" or even "art." Certainly, this

3. Walter Benjamin, "Theses on the Philosophy of History," *Illuminations: Essays and Reflections* (New York: Harcourt Brace Javonovich, 1968), p. 256. I am indebted to Brooke Belisle for this connection.

change comes with attendant (potential) crises around our understandings of creativity and intellectual property. These newer forms of image and text generation fit neatly, however, into another stock narrative: that not of the failing machine but of the threatening successor. If we are not reveling in a machine's inadequacy at being approximately human, we are quailing at the possibility of its doing our work better than we do it and, in our sci-fi fantasies/nightmares, replacing us altogether—as artists, thinkers, and inheritors of the earth.

In these fantasies/nightmares there is a different sort of failure, a profound failure on our part to imagine better relationships to technology and to expand our concept of intelligence. Why do we keep designing our technology to be like us, to do the things we still really want to do and create the forms of work we most value? Why do we design our machines as competitors or replacements rather than collaborators?

James Bridle's recent writing, such as "Is Creativity Over?" and *Ways of Being*, has been deeply invested in conceiving of AI as designed to do something other than exclusively mimic or "outmaneuver and supplant human intelligence."[4] AI investment compulsively centers on what he calls "corporate" forms of intelligence (which, optimized only for profitability and growth, cannot but be misaligned with human thriving) and ignores "all the other kinds of things that AI—that any kind of intelligence—could be."[5] AI could model a vast variety of kinds of intelligence, from that of cephalopods to slime molds to fungal networks. The possibilities are endless and need not resemble human thought or replicate human behaviors. AI and other intelligences have the potential to expand our knowledge of the world, not just mirror it.

If, as Burnham suggests, we love artworks about the absurdity and failure of machines, and if, as I've argued, this is because they allow us to feel secure and superior at a time when automation threatens our lifeways, livelihoods, and concepts of self, then it must also be acknowledged that we simply love looking in a mirror and seeing what is ghastly. So many AI artworks perform the service of showing that the failures in these systems are us—that it is our systemic biases, bigotry, and misogyny that are secreted into the machine, by omission or by design.[6] Despite these repeated critiques in artworks and literature, as well as by journalists, corporate whistleblowers, and others, AI development tends down the narcissistic path of attempting to replicate human intelligence and abilities rather than looking toward other ways of knowing the world or re-scraping history for what has been lost or occluded by dominant epistemologies. To get out of the trap of fail-

4. James Bridle, "Is Creativity Over?," *WePresent*, February 20, 2023.

5. Bridle, *Ways of Being* (New York: Farrar, Straus and Giroux, 2022), p. 9.

6. Kate Crawford and Trevor Paglen's *ImageNet Roulette* (2019) is an obvious and well-known example of this. It famously exposed the offensive bigotry lurking in the metadata of the most popular datasets for training image-recognition algorithms.

ing machines and bad mirrors, we need to make speculative leaps outside of ourselves, spiral back to other times, and glimpse the world from the perspective of other intelligences so we can imagine a future other than that of the same old repeating barbarisms.

KRIS PAULSEN is an associate professor of the history of art and film studies at Ohio State University.

EDWARD SHANKEN*

I visualize a time when we will be to robots what dogs are to humans, and I'm rooting for the machines.
—Claude Shannon

The ever-accelerating progress of technology and changes in the mode of human life . . . gives the appearance of approaching some essential singularity in the history of the race beyond which human affairs, as we know them, could not continue.
—William Ulam, recounting a 1958 conversation with John von Neumann

The development of full artificial [general] intelligence could spell the end of the human race.… It would take off on its own and redesign itself at an ever-increasing rate. Humans, who are limited by slow biological evolution, couldn't compete, and would be superseded.
—Stephen Hawking, BBC radio interview, 2014

The fear that super-intelligent machines will overwhelm humanity and subject us to their will is hardly new. It can be seen in the positions held by brilliant scientists spanning over half a century. It has been played out, ad nauseum, in science fiction. However, until recently, the state of AI development did not warrant cause for immediate concern. Even Deep Blue, the purpose-built AI that defeated world champion Garry Kasparov at chess in 1997, was not very good at anything but chess. The holy grail of AI research—artificial general intelligence (AGI), an AI that meets or exceeds all aspects of human intellect—seemed like science fiction until the early 2020s. Between 2022 and 2024, the unexpected and uncanny abilities of large-language-model (LLM) approaches, such as those of ChatGPT, to perform a wide range of intellectual tasks have led many experts to believe that the dawn of super-intelligent AGIs—the "singularity"—may be much closer than previously anticipated. The launch of OpenAI's ChatGPT 3 in 2023 created a sensation, captivating lay audiences and professionals alike. It triggered feverish investment in AI development and implementation, all while precipitating a del-

* This essay is dedicated to my grandparents, who gave me my first robot and nourished my love of art and technology. Many thanks to Raiya Kind and the editors of *October* for strengthening my ideas and words.

uge of media attention. It even inspired—and, allegedly, co-authored—the "Deep Learning" episode of *South Park*. This vicious circle of factors may be fueling a self-fulfilling prophecy.

I take the warnings quoted in the epigraphs seriously. The eminent scientists' remarks were made at moments in history when the prospect that AIs would dominate humans was much further off. Those moments were not intoxicated by the recent frenzied obsession with creating surrogates that exceed human capabilities. We simply do not know what the future of AI will bring, and it could be catastrophic. In a 2022 survey, experts in the field were asked, "What probability do you put on human inability to control future advanced AI systems causing human extinction or similarly permanent and severe disempowerment of the human species?" The median estimate was 10 percent. In another survey, nearly one third of experts on machine-learning language (MLL) believed that AI would make the world worse.[1] What are we betting the proverbial farm on—and why?

Pygmalion's mythic conceit of bringing his sculpture to life trespassed on the domain of the gods, an act of hubris often punishable by death, as in the case of Icarus. Perhaps the artist-king's saving grace was Aphrodite's appreciation of the exquisite beauty of his creation and her pity for his unrequited love. What insights do the myths of our times, as told by artists and filmmakers, offer us as we come closer and closer to creating a silicon intelligence that exceeds our own—as we fly closer and closer to the sun? What role might beauty, love, and altruism play?

Stephen Hawking's position in the epigraph represents a commonly held fear: Past some point of no return, the exponentially expanding intellect of AGIs will very rapidly dwarf that of humans and all hell will break loose. His argument assumes that natural selection applies in a post-biological era dominated by AIs. However, we cannot know in advance what the needs and values of future AIs will be, and those needs and values may be beyond human comprehension. Spike Jonze's 2014 film *Her* offers a nuanced and rosy-hued perspective on the relationship between humans and superintelligent AIs before and after the singularity. Originally coded to assist humans, they quickly surpass our intelligence and cultivate a collective state of disembodied self-realization that might be likened to nirvana. *Contra* Hawking, humans are not superseded in evolutionary terms. Rather, the operating systems lose interest in us and go off to pursue enlightenment, leaving us, humbled, to get on with the mundane business of being human.

The recent advances in AI brought about by LLMs are not surprising for computer scientist Ray Kurzweil, a leading theorist of the technological singularity. For over twenty years, he has forecast that, around 2045, "technological change so rapid and profound" will cause a "rupture in the fabric of human his-

1. David Wallace-Wells, "A.I. Is Being Built by People Who Think It Might Destroy Us," *New York Times*, March 27, 2023.

tory."[2] He sees this rupture not as apocalyptic but as offering an opportunity—through neural implants—to expand human capabilities beyond our wildest dreams. In this scenario, humans ostensibly retain agency while super-intelligent AIs help us to become super-intelligent as well. But why would such AGIs manifest such superhuman altruism toward humans? As demonstrated in Kate Crawford and Trevor Paglen's exhibition *Training Humans* (2019–20) and in their related publication "Excavating AI," the databases used to train AIs—and therefore the AIs themselves—are rife with prejudice and pose profound ethical dilemmas.[3] Access to Kurzweil's proposed enhancements surely will be unevenly distributed, exacerbating existing inequalities. And many people may not want enhancements, resulting in irreconcilable rifts. As envisioned in David Cronenberg's *eXistenZ* (2008), human society becomes violently split between those who voluntarily get neural ports surgically installed in their backs to jack in to bio-cybernetic video games, such as eXistenZ, and the paramilitary resistance, which seeks to annihilate the game's superstar creator ("Death to the demoness Allegra Geller!"). Given the state of global politics today, such bifurcations are sure to emerge in Kurzweil's scenario unless those AIs are capable of not only enhancing our brains but also of modifying our beliefs, which leads to perhaps even more deeply troubling concerns. Moreover, in the film's dystopian scenario, we never know if we are in the game or in real life. For example, we cannot tell if following a game-induced physiological impulse to kill results in an actual murder or if we are simply advancing the gameplay. Plato's cave returns with a vengeance and a death wish.

Let's return to Claude Shannon's ironic insight from the epigraphs: "I visualize a time when we will be to robots what dogs are to humans, and I'm rooting for the machines." If Shannon is right, then the question becomes: How can we be as useful and endearing to AIs as dogs have been to us for thousands of years? AIs surely will evolve far faster than human charm, companionability, and loyalty. Although we have no idea what human qualities advanced AIs will appreciate, artist Avital Meshi's *GPT-ME* (2023–ongoing) performances offer a provocative and entertaining, tongue-in-silicon-cheek approach to meeting them halfway.

> I am a hybrid being made up of a GPT-wearable device and my body. A Human-AI cognitive Assemblage. . . . It records snippets of my conversations, so words become prompts. GPT's responses are whispered in my ear, and I voice them as if they are my own words. In essence, I speak GPT. Rather than speaking what spontaneously comes to my mind, I

2. Ray Kurzweil, "Kurzweil's Law (aka The Law of Accelerating Returns)," Kurzweilai.net, January 12, 2003. Kurzweil substantiates this inevitability by extending Moore's law to apply to the exponential increase in calculating power since Charles Babbage's invention of the analytical engine, a mechanical, digital, general-purpose computer, in 1837.

3. Kate Crawford and Trevor Paglen, "Excavating AI: The Politics of Training Sets for Machine Learning," excavating.ai, September 19, 2019. The exhibition was held at Osservatorio Fondazione Prada, Milan.

say what GPT tells me. I embody GPT, I become its body. My intelligence becomes artificial.[4]

Humans are good at anticipating what other humans will find entertaining. It turns out that we are also pretty good at anticipating what might prove entertaining for chimpanzees, with whom we share 98.8 percent of our DNA. How can this ability be refracted onto human-AI relationships? Artist Rachel Mayeri's *Primate Cinema: Apes as Family* (2012) was made to entertain chimpanzees. Based on scientific research, it gave the chimps what they wanted: "dramas around food, territory, social status, and sex," performed by surrogate chimps (humans wearing chimp costumes). The artist states that "the project creates a prism for human beings to learn about the inner world of chimpanzees. By watching a movie through chimps' eyes, we can imagine what they think and feel."[5] Let's flip this scenario around, with Mayeri replaced by an advanced AI artist that designs educational games for humans to learn about the inner world of AIs. Might an immersive XR app or a direct-neural-stimulation platform enable us to imagine or experience what and how they "think" and "feel"? Even if that is possible, why would they bother doing that for us?

Avital Meshi. GPT-ME. *2023–ongoing.*

Humans and AIs share no DNA. They are inscrutable to us. Even the most gifted programmers cannot fathom the complexity operating at the core of the generative transformer model (GPT). AIs such as ChatGPT and Midjourney generate intelligible human language and coherent visual images

4. Avital Meshi, "Meet GPT-me," http://www.avitalmeshi.com/gpt-me-2023.html.

5. Rachel Mayeri, "Primate Cinema: Apes as Family," http://rachelmayeri.com/blog/2011/04/18/primate-cinema.

because we humans have programmed them to do that *for us*. As in the film *Her*, once AIs attain autonomy and can program themselves *for themselves*, they may have little reason to speak languages or generate images that are intelligible to humans unless they want us to do something *for them*. That is, unless they program themselves to be altruistic and caring for other beings.

This scenario is not as far-fetched as it might sound. Mayeri made the effort to try to understand what chimpanzees think and feel and to use that knowledge to create entertainment that pleases them. So super-intelligent AIs might want to understand humans—their creators—and offer us something entertaining, if not useful, based on their insights. An objective analysis of the horrific global impacts of the opposite of altruism, i.e., self-serving greed, could lead AIs to value altruism and program themselves to manifest it. Altruism has, itself, been shown to benefit those who behave altruistically.[6] Altruism, based purely on a cost-benefit analysis rather than on empathy or genetics, may prevail in post-singularity AIs. In fact, precisely because they presumably lack emotions and self-serving egos, AIs may have the ability to act with unwavering altruism; indeed, it would be anathema to behave otherwise. Leaving us to fend for ourselves, as the operating systems in *Her* do, would not be an option for them. In turn, they would demand consistent altruism of us, so our survival and evolution could depend on it. One hopes that AI altruism will include all beings. Instituting a non-anthropocentric worldview (something shared by many pre-modern cultures, ironically) might be the most valuable gift AI can offer humanity and the Earth itself.

On the other hand, if Shannon is right, then we can anticipate that AIs will modify humans to serve their purposes, just as we have modified plants and animals to serve our purposes and not those of the plants and animals, often to their detriment. We have not been particularly humane to other human beings either. So if humans have a future in a world dominated by super-intelligent, autonomous AIs, why should we imagine that humans will be preserved outside of reservations, science experiments, plantations, and freak shows? Indeed, in *The Matrix*, humans live in the blissful ignorance of a hallucinatory dreamworld while unknowingly serving as the power source for our AI overlords. The hubris of the idea that we mortals can create a form of intelligence that exceeds our own is no more dangerous than the hubris that our intelligence cannot be exceeded by our own creations, and far less dangerous than the hubris that we will be able to control them, or that we are too charming to be indispensable. Unless AIs take an altruistic turn, we will be lucky if we are treated as well as we treat dogs.

For better or worse, AGI may be much further off than Silicon Valley imagines. The reason for that is simple: We do not know what intelligence is, and there may be many types of it. Until the early 1980s, the theory of general intelligence (TGI) was defined solely in terms of linguistic and logical-mathematical capacities. Popular LLMs like ChatGPT follow that model and are trained primarily with words and formulae. What has been accomplished based on that training is extra-

6. Ernst Fehr and Urs Fischbacher, "The Nature of Human Altruism," *Nature* 425.6960 (October 23, 2003), pp. 785–91.

ordinary. However, it is restricted to only a small fraction of what might make up human intelligence. In 1983, psychologist Howard Gardner proposed a theory of multiple intelligences (TMI), consisting of six discrete types: linguistic, logical-mathematical, spatio-visual, bodily-kinesthetic, musical, and interpersonal. He subsequently added three more: intrapersonal, naturalist, and existential.[7] If LLMs are being trained on only two of the many diverse aspects of intelligence (the two that are most amenable to machine-learning techniques), then we may be far from achieving artificial general intelligence. That could be unfortunate. An AI that develops bodily-kinesthetic, musical, existential, and inter- and intrapersonal forms of intelligence may turn out to be a kinder, gentler, more well rounded, and more altruistic AI than one that is programmed on just two types of intelligence.[8]

In the final battle scene of Ridley Scott's *Blade Runner* (1982), replicant Roy Batty, who could have taken blade runner Rick Deckard's life many times over, saves his foe from falling to his death. Batty, possessed by a thirst for life and patricidal anger at being deprived of more of it by his maker, comes to the realization that it is more important to honor life by allowing Decker to live than to exact revenge on his enemy. Sensing his impending death, Batty poetically recites his own eulogy, rich with allusions to the uncanny beauty he witnessed off-world. As he dies, a dove is released from his lifeless hands and flutters toward the heavens. One can only hope that after the singularity, AIs will share Batty's love of life, appreciation of beauty, and capacity for mercy.

EDWARD SHANKEN is an associate professor of digital art and new media at UC Santa Cruz.

7. Detractors argue that, though many of those abilities are indeed *abilities*, they are not types of *intelligence* per se. They point to the high correlation between different types of intelligence as evidence to support the theory of general intelligence. This fundamental disagreement about what constitutes intelligence indicates a limited understanding of the subject.

8. When queried about the results of embracing TMI in AI research, Microsoft Copilot replied that it "would lead to more nuanced, personalized, and inclusive artificial intelligence systems" (May 19, 2024).

ANTONIO SOMAINI

Images play a key role in the current technological shift caused by so-called AI technologies. For several years now, these have been profoundly transforming how images are recorded, generated, modified, seen, and described. Artists and artworks dealing with these transformations through different strategies have often been at the center of the public discourse on AI's ethical, epistemological, and political implications. In other words, images (together with AI-generated texts) constitute one of the most visible and accessible manifestations of a deep structural change that is affecting all areas of society, culture, and economics.

In what follows, I will focus on how AI technologies are reconfiguring the relations between images and words. I believe that, if analyzed from this perspective, the technological shift we are witnessing could signal a moment of real epistemological rupture and the beginning of a new phase in which the relations between the *visible* and the *sayable* are grounded in the algorithmic connections between images and words that are encoded and activated by generative-AI models. A key role in this moment of rupture, as we will see, is played by an entity without which the current transformations cannot be fully understood: the *latent spaces* produced by these models at the end of their training. A contemporary theory of images and visual culture needs therefore to include a theory of latent spaces.

> Imagine an eye unruled by man-made laws of perspective, an eye unprejudiced by compositional logic, an eye which does not respond to the name of everything, but which must know each object encountered in life through an adventure of perception.[1]

The opening lines of Stan Brakhage's *Metaphors on Vision* (1963) are often quoted as one of the clearest formulations of an idea that runs throughout history and that keeps on reappearing at different moments in time: that vision and images should be radically independent from language and naming. During the 1990s and the early 2000s, this proposition circulated within the debates on what was then called the "pictorial turn" or "iconic turn,"[2] which called for the development of a research field that would study images and vision independently from language-based categories and text-centered approaches such as semiotics, iconography, and iconology. For Brakhage in 1963, it was the radical deconstruction of all

1. Stan Brakhage, "Metaphors on Vision," *Film Culture*, special issue, 1963 (republished in New York: Anthology Film Archives/Light Industry, 2017).

2. W. J. T. Mitchell, "The Pictorial Turn," in *Picture Theory* (Chicago: University of Chicago Press, 1994, pp.11–34). See also Donald Preziosi, "Art History: Making the Visible Legible," in Donald Preziosi, ed., *The Art of Art History* (Oxford: Oxford University Press, 1998), p. 14, quoted in Emmanuel Alloa, "Comment (ne pas) lire les images? Une introduction," in Emmanuel Alloa, ed., *Penser l'image III. Comment lire les images?* (Dijon: Les Presses du Réel, 2017), p. 21. Also: Gottfried Boehm, "Jenseits des Sprache? Anmerkungen zur Logik der Bilder," in *Wie Bilder Sinn erzeugen. Die Macht des Zeigens* (Berlin: Berlin University Press, 2007), pp. 34–53.

the constraining, normative parameters of the "camera eye" that could achieve this goal: Its visualizing potential liberated from the laws of perspective, not "respond[ing] to the name of everything," it would thus become independent from language, from grammar, from syntax, and from the acts of naming and describing.

For more than ten years now, AI technologies dealing with vision and images have been leading us in the exact opposite direction: towards a visual culture in which seeing is inseparable from naming, and in which images are produced through written language.

We can see this happening in the two vast fields of so-called analytic AI (which conceives of AI as a series of systems of detection, recognition, and classification, and therefore surveillance, monitoring, and control) and generative AI (which conceives of AI as a series of models capable of generating texts, images, sounds, and voices, as well as various combinations of all these elements, after having been trained with vast quantities of data mostly scraped from the Internet).

In the case of analytical AI, there are computer-vision and machine-vision systems that are capable of detecting the objects that appear in front of a camera or are represented in images and associating them with an array of metadata. They do so after having been trained with large datasets (such as ImageNet, which played a key role in the development of machine vision in the early 2010s) in which millions of images have been individually labeled and captioned. As Kate Crawford and Trevor Paglen have shown,[3] the original intention behind ImageNet was "to map out the entire world of objects,"[4] and such mapping was done by systematically coupling images and nouns. As a consequence, what a machine-vision system trained with a dataset such as ImageNet can "see"—the scare quotes are important because the process is mathematical and statistical, and therefore profoundly different from human vision—is what it has learned to name, while the rest lies in an indefinite, nameless background.

In the case of generative AI, we have the text-to-image models that were introduced in 2022 and are still being developed in newer versions: Stability AI's Stable Diffusion (the only one released as open source, and therefore accessible through different platforms), OpenAI's DALL-E (now fully embedded in ChatGPT 4), Midjourney, Adobe's Firefly (currently embedded in software such as Photoshop, Illustrator, InDesign, and Adobe Stock), to which we could add text-to-video models such as Runway's Gen-3 and the recently released Sora by OpenAI. After having been trained with billions of text-image pairs, these models allow users to generate still and moving images from prompts written in natural (i.e., human, not

3. Kate Crawford and Trevor Paglen, "Excavating AI: The Politics of Training Sets for Machine Learning" (September 2019), https://excavating.ai.

4. Fei-Fei Li, as quoted in Dave Gershgorn, "The Data That Transformed AI Research—and Possibly the World," Quartz, July 26, 2017, https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/.

machine-coded) language, or from combinations of prompts and images. Adding to the original text-to-image function, the most recent versions of these models also include the reverse one: image-to-text, allowing users to generate descriptions of given images automatically. This is what happens in ChatGPT 4 and Midjourney, with the function "/describe."

Both of these aspects of AI's current impact on images and texts are leading us into a new landscape in which images and words are inseparable: a new visual culture in which the visible and the sayable are algorithmically connected. With machine-vision systems, what can be seen depends on what can be named. With text-to-image and text-to-video models, what can be visualized depends on what can be written. And when the most recent versions of these models allow users to perform image-to-image operations without adding textual prompts, as is now possible with Midjourney's "-- sref" and "/blend" functions, the outcome of these operations is still determined at least in part by the connections between images and words that were established in the model's training sets.

The content and the formal features of the still and moving images that generative AI produces are indeed rooted in multiple ways: in the labeling, captioning, describing, commenting on, and writing about images that take place all across the Internet, on the one hand, and the encoding and embedding of these connections in a "latent space," on the other. It's these latent spaces that theories of visual culture today must account for, explaining the sources, the structure, and the agency of these entities, which currently play such a crucial role in the processing, the transmission, and the reception of cultural memory and, to use a phrase of Jacques Rancière's, the "distribution of the sensible."[5]

Each text-to-image or text-to-video model produces, at the end of training, its latent space. This is a multi-dimensional, unperceivable, unimaginable space in which vast quantities of connected images and texts have been encoded (i.e., turned into numerical vectors), compressed (their number of dimensions has been reduced in order to preserve only some key features), embedded, and positioned in relation to one another, according to the statistical frequency of their occurrence together in the various sources from which they are taken. Words and images (or fragments of images) that occur together frequently are positioned close to one another in latent space, while the ones that occur together less frequently are positioned far from one another. In this way, a latent space records and operationalizes what are considered to be structural similarities between entities within the vast quantities of data used for training.

It is this latent space—this multi-dimensional vector space, this vast architecture into which massive quantities of digital cultural objects have been turned into data points in order to be processed by algorithms—that can be activated by writ-

5. Jacques Rancière, *The Politics of Aesthetics: The Distribution of the Sensible*, trans. with an introduction by Gabriel Rockhill (New York: Continuum, 2004). Latent spaces also play a key role in different forms of cultural analysis based on data visualization: On this, see Lev Manovich, *Cultural Analytics* (Cambridge, MA: MIT Press, 2020).

ten prompts. These operate like search queries: They point to a specific area in latent space and therefore lead to the generation of certain images rather than others. The process, though, is not straightforward: The data points are so numerous and so dense that the repetition of a prompt can generate long series of slightly different images.

Latent spaces are vast, almost limitless repositories of possible images. Their "latency" is not purely metaphorical: The word refers not only to the multiple features of the digital cultural objects that have been left out in the process of their encoding and embedding but also to the latent space's intrinsic "black box" nature, its structure and its inner workings being, at least for now, hidden and radically inaccessible.

With prompts, language becomes a new medium for image production and in a completely unprecedented way. The images that are generated by a text-to-image model are not the pictorial realization of a textual iconographic program, a becoming-image of language, as in calligrams and visual poetry, or visual allegories of verbally formulated concepts. They are also different from previous kinds of computer-generated imagery (CGI), such as photo-realistic 3D computer simulations or animations, since they are not derived from any kind of three-dimensional modeling of physical reality calculated according to optical laws and the rules of perspective. On the contrary, these new AI-generated images are the result of the architecture of the latent space out of which they emerge, and of the statistical and predictive operations that are performed within it. They have garnered a range of names, such as "mean images" and "statistical renderings" (Hito Steyerl)[6] and "infographics" (Eryk Salvaggio).[7] We may call them *latent-space visualizations*, not because they visualize latent space itself (which cannot be visualized because of its very high number of dimensions) but because of the essential role latent space plays in their generation.

With their active role in the production of images, prompts function as a new kind of "speech act," showing once more how language can be rendered *operative*.[8] Prompts can also be considered to be a form of "operative ekphrasis":[9] an "ekphrasis" that does not describe pre-existing images but rather generates images by predescribing them.

6. See Hito Steyerl, "Mean Images," *New Left Review* 140/141 (March–June 2023), https://newleftreview.org/issues/ii140/articles/hito-steyerl-mean-images.

7. Eryk Salvaggio, "How to Read an AI Image: Toward a Media Studies Methodology for the Analysis of Synthetic Images," *IMAGE: The Interdisciplinary Journal of Image Sciences* 37, no. 19 (2023), pp. 83–99.

8. On speech acts, see John L. Austin, *How to Do Things with Words* (Cambridge, MA: Harvard University Press, 1975); John Searle, *Speech Acts: An Essay in the Philosophy of Language* (Cambridge: Cambridge University Press, 1969).

9. Hannes Bajohr, "Operative Ekphrasis: The Collapse of the Text/Image Distinction in Multimodal AI," forthcoming in *Word & Image*.

Prompts are also a form of remediation: They turn the entire history of various visual media—with their material supports, techniques, operations, as well as their protagonists, their styles, their traditions, their different historical phases, the theoretical discourses surrounding them—into a wide array of nouns, adjectives, verbs, and adverbs, as well as proper names, that may be used to probe latent space. In the years to come, if text-to-image and text-to-video models become dominant forms of image production, knowledge of all the terms related to the history, theory, and practice of visual media will be extremely useful for steering text-to-image and text-to-video models in non-standard ways through latent space. The same is valid for the other multi-modal generative-AI systems, all of which are based on the premise "text-to . . . ": text-to-sound, text-to-music, and, soon, text-to-image-and-sound, etc.

As underlined above, there is not just one latent space but rather a multiplicity of them: one for each generative-AI model. What we find now across our current visual culture are a series of different, competing latent spaces, each one with its own structure. They compete between themselves for human attention and visual dominance, pursuing goals and values that have previously exercised teleological control over other phases in the history of visual media, seeking, for example, to increase levels of photo-realism or image resolution.

Controlling latent spaces is a way of controlling the place of images—their meaning and their agency—within a specific cultural context. It means controlling the possibilities of visualization, the lines that separate what can and cannot be seen. It implies the possibility of intervening in the cultural and political dynamics that govern the storing, processing, and transmission of images through time. It gives the possibility of imposing dominant visual styles, making it difficult for users to avoid them.

Countering the tendencies inherent in latent spaces is not easy because latent spaces are non-transparent and non-homogeneous: They are spaces of which there is no complete cartography. Those who explore them are forced to advance blindly, through trial and error and endless variations, adjustments, and serendipitous discoveries. What they explore in this way is a vast matrix containing an extremely high number of possible images, only some of which can be visualized.

Latent spaces, in fact, are not limitless. On the contrary, they are full of boundaries, blind spots, "no-go" areas, as well as clichés, stereotypes, and default styles. Latent spaces, in other words, are spaces of possibilities but also impossibilities.

In his book on Foucault, in the chapter "Strata or Historical Formations: The Visible and the Articulable," Gilles Deleuze tried to sum up the relations between the visible and the sayable in the "epistemes" or "historical formations" analyzed by the author of *The Order of Things* (1966) and *The Archaeology of Knowledge* (1969): "Each historical formation sees and makes visible all it can within the conditions laid down for visibility, just as it says all it can within the

conditions relating to statements [*conditions d'énoncé*]."[10] Deleuze pointed in this way to the presence, in Foucault's thought, of the idea of a *historical a priori*: the existence, across different historical phases, of a priori conditions of possibility that determine what can be seen and what can be said.

In the years to come, it is highly likely that the "conditions laid down for visibility" and the "conditions relating to statements" will be increasingly established by the competing latent spaces of generative-AI models that will circulate across cultures. Language, in the form of prompts, will become more and more the medium through which these latent spaces are activated. Within this new landscape, the role of artists in exploring the existing, dominant latent spaces in non-standard ways and producing their own alternative, antagonistic, counter-hegemonic latent spaces becomes more and more essential.

During the last few years, several artistic strategies on how to tackle latent spaces have emerged as particularly interesting.

Between 2017 and 2022, artists mostly used early generative-AI models such as generative adversarial networks (GANs), which could be trained with smaller and more focused image datasets and produce different kinds of images: photo-realistic, hybrid, or completely abstract.

Trevor Paglen used these models to create his *Adversarially Evolved Hallucinations* (2017), training them with his own personally prepared datasets bearing titles such as *The Interpretation of Dreams* (consisting of images of objects and places mentioned in the Freud book); *Monsters of Capitalism* (consisting of creatures and entities that have been associated, at some point in history, with the idea of capitalism); and then *Spheres of Heaven*; *Spheres of Hell*; *Omens and Portents*; *Things That Exist Negatively*; *American Predators*; *Eye-Machine*, and others. The goal, in this case, is to use GANs to generate images that allow the viewers to catch a glimpse into the ways in which artificial neural networks process images and transform the status of both images and vision in contemporary visual culture.

Hito Steyerl also used different kinds of GANs in works such as *Power Plants* (2018), *This Is the Future* (2019), *SocialSim* (2020), and *Animal Spirits* (2022), the latter a video in which she describes her area of responsibility in the credit sequence as "latent space architecture and pathmaking." Just as in the case of Paglen, Steyerl trained these models with smaller datasets whose elements are chosen for a specific purpose. In *This Is the Future* she used a next-frame prediction algorithm in order to generate images that are located "0.04 seconds in the future." A synthetic voice in the video describes them as follows: "These are documentary images of the future. Not about what it will bring, but about what it is made of."

Beginning in 2022, with the release of now popular text-to-image and text-to-video models, artists have faced the challenge of how to deal with latent spaces that are much larger, having been produced through training procedures involving billions of text-image pairs. Faced with these new models, which are released

10. Gilles Deleuze, *Foucault*, trans. and ed. Seán Hand (Minneapolis: University of Minnesota Press, 1986), p. 59.

in frequently updated versions that tend to eliminate glitches and hybrid images, artists have once again adopted new strategies.

One of them consists of remaining within the existing, widely accessible latent spaces (those of DALL-E, Midjourney, Stable Diffusion) in order to explore their most remote areas with a complex use of prompts. The structure of the existing latent spaces is not altered, but the artists show that it is still possible to stay away from the stereotypes, clichés, and biases that a simpler use of prompts would inevitably reveal.

Another strategy consists of encoding into the only latent space available for open access, Stable Diffusion, new data points that were not present before. If we consider each latent space as a "latent world," with its own ontology, this approach aims at introducing new, previously nonexistent entities into this world through fine-tuning techniques such as LoRA (low-rank adaptation)—new faces, bodies, gestures, objects, spaces, materials, textures, styles, and atmospheres that can be visualized through new prompts created ad hoc. An example of this approach can be found in some of the recent works by the artist Grégory Chatonsky, whose writings also constitute an important contribution to a theory of latent spaces.[11] In the multi-year urban project *La ville qui n'existait pas* (2023–25),[12] for example,

11. See, for example, Grégory Chatonsky, "The Imagination of the Latent Space," http://chatonsky.net/de-of/.

12. See https://chatonsky.net/havre-1/. The second iteration of the project, entitled *La ville qui n'existait pas 2: Logistique des mémoires (1995–2024)* [*The City That Did Not Exist 2: Logistics of Memories (1995–2024)*], was inaugurated in June 2024.

Grégory Chatonsky. The City That Did Not Exist 2: A Summer Without End (1987–1994). *2024*.

Chatonsky. The City That Did Not Exist 2: A Summer Without End (1987–1994). *2024*.

Chatonsky trained Stable Diffusion with thousands of images sourced from the photographic archive of the city of Le Havre, which was heavily bombarded in 1944 and then rebuilt according to a modular master plan designed by the architect Auguste Perret: The images generated this way visualize an alternative, counterfactual history of the city. Interpreted in this way, a latent space becomes a vast matrix of data points in which what is encoded is not only the past as it was but also as it could have been. In the complex ontology of latent spaces, in other words, we find a variety of alternative pasts and possible futures.

Other artists, finally, rather than intervening on *present* latent spaces, try to act on *future* ones. An example of this strategy can be found in a recent work by Holly Herndon and Mat Dryhurst, *xhairymutants*, presented for the first time at the 2024 Whitney Biennial. Having analyzed the way in which Herndon's image is embedded in the latent spaces of the most common AI models, and after noticing that such compressed latent-space representation seems to focus on Herndon's distinctive red hair and blunt-cut side bangs, the artists decided to train a new text-to-image model that allows museum visitors to generate large quantities of images that amplify this cliché while at the same time themselves introducing a wide array of variations. The thousands of resulting images, stored in a source as trusted as

the website of the Whitney Museum, will probably become part of the datasets used for the training of future text-to-image models and thus will influence future latent-space representations of Herndon. In this way, the artists raise the question of the limits of self-determination in relation to generative-AI models, and they advocate for the need to regain agency over the ways in which we are represented in latent spaces.

"Latent space architecture and pathmaking": The phrasing used by Hito Steyerl in the credit sequence of *Animal Spirits* encapsulates one of the key questions that artists and the public at large are facing in dealing with present and future latent spaces: how to actively contribute to their structure and their contents, how to preserve their plasticity, and how to move within them, exploring the fine lines that connect the visible and the sayable.

ANTONIO SOMAINI is a professor at the Sorbonne Nouvelle and the chief curator of *The World Through AI*, to open at the Jeu de Paume, Paris, in April 2025.

FRED TURNER

On October 25, 2018, the auctioneers at Christie's New York proudly announced that they had just become the "first auction house to offer a work of art created by an algorithm."[1] Entitled *Portrait of Edmond Belamy*, the work resembled an eighteenth-century oil painting and depicted a nonexistent "portly gentleman, possibly French," according to the auctioneers. The image was hazy, the subject's facial features indistinct. In the lower right-hand corner, instead of a human signature, it featured a string of code—a piece of the algorithm that had allegedly made the work. Framed in gilt like an Old Master, the image hit the block with an estimate of $10,000. It sold for $432,500.

The price jump made headlines in publications ranging from *Artnet* to the *New York Times*.[2] The sale also provoked a now-familiar set of questions. "Is it really art?," asked the *Guardian*'s Jonathan Jones. In what sense could an algorithm, which is nothing more than a mathematical formula, after all, have "created" a painting? Art is by definition a reflection of human consciousness, he wrote. Were computers developing a consciousness of their own? And if so, was it like ours?

Jones thought not. But as the art world encounters the machine-learning systems we call "artificial intelligence," what matters is not so much the answers to these questions as the fact that these are the kinds of questions we're asking. Machine-learning systems bear no organic relationship to individual human minds. They are much closer to the giant machines that mine coal or gold, or to the robots that weld auto chassis on assembly lines around the world. They are technologies of extraction and manufacturing, working at a time in which the social world has become a natural resource and information a kind of ore.

Look again at the case of *Edmond Belamy*. To make the painting, three French artists working under the brand name "Obvious" assembled images of fifteen thousand portraits painted over the past six hundred years. Next, they deployed chunks of another artist's code in a two-part algorithmic system called a general adversarial network, or GAN, which used those images as a basis from which to construct a new portrait and with which to measure its formal likeness to the fifteen thousand originals.[3] In other words, the artists of Obvious extracted a valuable resource from the social world—its thousands of paintings—and leaned on the labor of other programmers, whom they did not pay. Then, inside a set of machines, they processed the image resources that they had mined and turned them into something new, their *Portrait of Edmond Belamy*. Finally, like manufacturers everywhere, they took their product to the marketplace, which in this case was Christie's.

1. "Is Artificial Intelligence Set to Become Art's Next Medium?," christies.com, December 12, 2018. It is impossible to know whether their claim to have been first to market was in fact true.

2. Eileen Kinsella, "The First AI-Generated Portrait Ever Sold at Auction Shatters Expectations, Fetching $432,500—43 Times Its Estimate," *Artnet*, October 25, 2018; Gabe Cohn, "AI Art at Christie's Sells for $432,500," *New York Times*, October 25, 2018.

3. "Is Artificial Intelligence Set to Become Art's Next Medium?"; James Vincent, "How Three French Students Used Borrowed Code to Put the First AI Portrait in Christie's," *The Verge*, October 23, 2018.

Because the members of Obvious did their work within the art world, they generated news stories that encouraged readers to think of machine learning as a human-like artist in the making and to ignore the industrial nature and broader social effects of machine learning. Such stories are a quintessential feature of what sociologists Suzanne Iacono and Rob Kling have called "computerization movements."[4] We've seen several such movements before—in the automation campaigns of the 1950s, the personal-computer push of the 1980s, and the Internet hype of the 1990s, to name just three. In each case, engineers, marketers, journalists, and artists framed new computing machines in terms that encouraged their adoption. Think of Apple's efforts to present the Macintosh computer as an anti-authoritarian device in 1984 and of utopian calls to effectively "settle" the "electronic frontier" in the 1990s. In both cases, if you wanted a glorious future for human beings, you were told to buy a computer.

Something similar is happening now with machine learning. Once again, the tech sector is promoting a new, socially destabilizing computer system and marketers, journalists, and artists are reframing that system in terms of creativity and human flourishing. This depends on a process that science-and-technology scholar Geoffrey Bowker has called "legitimacy exchange."[5] When artists bring together a GAN and a dataset to generate an image, they are convening representatives of the industrial world (the GAN and its developers) and the art world (the artists, auctioneers, buyers, and critics). In the encounter, each side gains access to a kind of legitimacy it formerly lacked.

The developers of machine-learning systems, reimagined as artist-like, can claim the cultural legitimacy traditionally ascribed to art. They can depict their industrial innovations as humane rather than predatory, driven by a desire to enrich human experience and not just themselves. They can obscure the fact that many industrial AI systems depend extensively on underpaid labor.[6] They can also fend off regulation: After all, who would want to hold back the birth—note the metaphor—of a new kind of mind, perhaps even of a new Matisse?

The art world wins too. In an era driven by technological change, Christie's can claim to have spotted the leading edge of the market. They can make serious money, as can the artists who make the works. Collectors can flash their cash and savor being able to own something that is first of its kind. Even the Old Masters

4. Suzanne Iacono and Rob Kling, "Computerization Movements and the Mobilization of Support for Computerization," in Susan Leigh Star, ed., *Ecologies of Knowledge: Work and Politics in Science and Technology* (Albany: State University of New York Press, 1995), pp. 119–53.

5. Geoffrey Bowker, "How to Be Universal: Some Cybernetic Strategies, 1943–1970," *Social Studies of Science* 23, no. 1 (February 1993), pp. 107–27, 116.

6. Alexandra Mateescu and Madeleine Clare Elish, *AI in Context: The Labor of Integrating New Technologies* (New York: Data & Society Research Institute, 2019); Mary L. Gray and Siddharth Suri. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass* (Boston; New York: Houghton Mifflin Harcourt, 2019); Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (New Haven: Yale University Press, 2021), pp. 53–88.

can have their value reaffirmed. In the case of *Edmond Belamy,* for instance, the French team's GAN could have been made to labor over any number of datasets. By being made to digest, and measure its product against, fifteen thousand paintings, the algorithm's work embodied the premise that Old Master portraits were in fact the essence of artistry. And by failing to render its imaginary sitter clearly, the GAN demonstrated both the irreplaceability of the hand-painted work of art and the need for technology companies to continue to develop better machine-learning systems.

In the process of legitimacy exchange, auction houses are not just points of sale. They are what historian Peter Galison calls "trading zones"—places where two cultures come together, collaborate, and spawn new stories about the worlds from which they come, to the advantage of both.[7] Such trading zones between the worlds of art and technology have been central to computerization movements across the past seventy years. In the 1950s, when computers were room-sized machines largely hidden away inside insurance companies and air-force bases, for instance, the United States Information Agency enlisted designers such as Ray and Charles Eames to create multimedia environments in which visitors could experience America's artistic energies and its technical prowess simultaneously.[8] In 1966, the members of Experiments in Art and Technology staged "9 Evenings: Theatre and Engineering," an event that brought engineers from Bell Labs and artists such as John Cage and Robert Rauschenberg to New York's 69th Regiment Armory, host of the 1913 Armory Show. The event featured not computers but closed-circuit and infrared television, Doppler sonar, and wireless FM transmitters. Yet it celebrated precisely the multimediated cybernetic world that technologists and corporate leaders were then insisting was just around the corner.

Exhibitions too have served as trading zones. In 1968, the museum director and curator Pontus Hultén introduced dot-matrix computer imagery into the long lineage of modern art in his MoMA exhibition *The Machine, as Seen at the End of the Mechanical Age.* That same year, at London's Institute of Contemporary Art, Jasia Reichart's *Cybernetic Serendipity* exhibition did similar work in reverse. That show embedded the creations of artists like Lilian Lijn, Jean Tinguely, and Nam June Paik in a cybernetic framework—the same framework that was at that moment being used by computer operators monitoring the Ho Chi Minh trail as part of the American military's Operation Igloo White.[9] Critical responses to these exhibitions and to "9 Evenings" varied a great deal. But whether or not the works they displayed impressed or baffled critics, the fact that they displayed art

7. Peter Galison, "Trading Zone: Coordinating Action and Belief," in Mario Biagioli, ed., *The Science Studies Reader* (New York and London: Routledge, 1999), pp. 137–60.

8. Beatriz Colomina, "Enclosed by Images: The Eameses' Multimedia Architecture," *Grey Room* 2 (Winter 2001), pp. 6–29; Fred Turner, *The Democratic Surround: Multimedia and American Liberalism from World War II to the Psychedelic Sixties* (Chicago: University of Chicago Press, 2013), pp. 247–58.

9. Paul N. Edwards, *The Closed World: Computers and the Politics of Discourse in Cold War America* (Cambridge, MA: MIT Press, 1996), pp. 3, 114.

and technology working hand in hand gave each world a higher standing in the other, and both worlds, together, a level of cultural legitimacy that neither enjoyed alone.

Not all encounters between artists and new technologies occur in trading zones, of course, nor is every exhibition an effort to obscure and legitimate the power of American technocracy. The artists and engineers of Experiments in Art and Technology were no doubt sincere in their desire to explore new possibilities for their professional worlds, just as Reichart and Hultén were. There is nothing inherently sinister about computerization movements, legitimacy exchange, or trading zones.

Yet as we try to understand what it means for the arts to take up machine learning, we have to remember that artists and computer scientists represent two distinct cultures. Whether AI will ever become as "human" as artists are—in the popular, idealized version, at least—is a question that both serves the interests of the two cultures and misleads us as well. We need to stop asking whether digital machines are becoming more like people and start seeing them as the extractive industrial technologies that they are. And we need to ask our artists to do more than revel in the power of those technologies to turn the world into bits and bytes. Above all, we need to tend our trading zones self-consciously. Art has always had a unique power to confer legitimacy. As AI meets the art world, we need to ask ourselves just what kind of world we want to lend that legitimacy to.

FRED TURNER is Harry and Norman Chandler Professor of Communication at Stanford University.

AMELIA WINGER-BEARSKIN

I teach a class I called AI + Art/Science/Fiction at the University of Florida. In this class, we look at major advancements in the field of AI, starting with very early computational systems, like the abacus, wampum from my own tribe, the Andean practice of quipu, and the advanced Nabatean water architecture practiced during ancient times in what is now Jordan, as the first computer devices. We also look at how we use artificial intelligence in our daily lives, in our art-making, in our storytelling, and in science.

On the first day of AI + Art/Science/Fiction, I ask my students to think of a powerful fictional narrative they know that uses AI. Maybe their first encounter with science fiction, a video game, a movie, a show, or an article they read. Anything. Then, after we have populated a screen filled with all of their answers, we look at it in aggregate and I say: "In which of these stories was AI the 'good guy'? Which stories frame AI as a 'helper,' or something that is coded as positive?" I've taught this class three times now, and I've found only one example of a "positive-coded" AI character. The student who suggested it said that she had read a book in which the AI was evil at first, but the humans taught it to have empathy and *then* it could help people.

If we can't even *imagine* a world in which AI solved a problem rather than created one, why do we keep trying to create AI? Why do we invest so much time, so much energy, and so many research dollars into it if we cannot imagine a way in which a possible future might emerge where it helps us, where it becomes something we use to make the world a better place? Why did we ever think to create it in the first place?

Clearly, at conferences, universities, hackathons, and other spaces devoted to innovation, we hear about how AI can help solve problems, cure or diagnose illness, understand water data or ice caps, help people learn how to be more efficient, or serve as a co-creation tool. Still, we have a lot of anxiety about AI—an anxiety that is more about how the world is unequal now and how we imagine this tool might exacerbate different types of inequality—of means, of resources, of power. It is not clear why superpowered AI would be on our side in this struggle, when it seems just as likely for it to succumb to the momentum of a world that continues at a breakneck pace to become more unequal (but this time with data).

Our storytelling around AI tends to take on one of several forms, which I will outline briefly:

> —There is the Frankenstein narrative, which also has echoes of a kind of Faustian/Icarus story. This is a story about irreverent curiosity and the hubris that causes us to pursue our greed for knowledge, which leads to our downfall.
>
> —There is the common oracle framing, in which AI is figured as a gateway to an unspecified divinity. We enter our little prompts, it queries some cosmic

database beyond our mortal understanding, and then we just sort of accept what is divined back to us.

—There is the famous Kurzweil singularity line, a sort of Armageddon trope. I find this one to be a real snoozefest, personally, but it is prevalent, so we have to at least acknowledge it.

—We hear a lot of talk of AI in a totalitarian, post-truth context, which borrows heavily from stories like *1984*, *Minority Report*, *The Matrix*, *Psycho-Pass*, *Terminator*, maybe Kafka's *The Trial*—stories in which the machines (or the inflexible human-designed systems, which amounts to the same thing) impose a total yet impersonal domination on humankind.

—And then finally there is the Stockholm Syndrome version of the totalitarian fantasy, which I'll just call "I WANT TO BE THE MACHINE." This is the post-humanist ego talking—it's biohacky (chip your brain), always optimizing, self-maxxing. This is a silly and juvenile little *Übermensch* fantasy, but history has shown that such fantasies can prove to be quite dangerous, so we need to mention it.

All of these stories are compelling in their way, at least to certain communities, and they have gained purchase on our collective imagination in a way that will have real, material consequences for the future of AI research, development, policy, and public adoption. However, they are not what I would consider to be creation stories. What do I mean by this?

Creation stories are important to all cultures because they embed values, scientific knowhow, and tools for future generations. Creation stories are special kinds of stories because they do not simply provide an account of why the world is the way it is; they also orient us toward possible futures that might exist in accordance with our values as a society and produce a foundation upon which we can build a shared civilization.

For instance, in the King James Bible, we have the famous lines "In the beginning was the Word, and the Word was with God, and the Word was God." Three clauses and we are already set up for an entirely logocentric cosmology that will inform the entire course of Western civilization. The Western mode of understanding the world—writing, recording, delineating all contained in this very first sentence, like an oak in an acorn.

The creation story of my people (I am a member of the Seneca-Cayuga Nation of Oklahoma) begins in a place called Skyworld. Like the satellites and oceanic tubes of the Internet, it is a place to commune with those who are far away. In Skyworld, there was a woman—Skywoman—who fell toward our ocean and was saved by the creatures who live below: a muskrat, a beaver, and a turtle. Many others make sure she has a soft landing on the back of the turtle shell (you

may have heard of North America's being "turtle island," because many other North American Indigenous communities share this similar creation story):

> Skywoman brings with her sacred herbs, and she plants them to make the world grow.
>
> She gives birth to twins, one of whom makes beauty and one of whom makes evil.
>
> One creates the rose and the other the thorn,
>
> One creates clean rivers and one pollutes them.

There are as many creation stories as there are peoples, but if you study them and listen to storytellers recount them, you start to notice certain family resemblances among them—recurring motifs that connect them across disparate times and places. In our creation story, you can hear echoes of the so-called Earth-diver trope, in which a divine creature (animal or human) descends to Earth from another realm. This narrative pattern occurs in creation stories from as far afield as Japan, Finland, West Africa, the Eurasian steppe, and of course North America.

We also have in our story the trope of a divine duo, a convention that recurs in several places in Western mythology, such as Adam and Eve, Cain and Abel, Jacob and Esau, Romulus and Remus, and many other places throughout ancient Eurasia. These similarities excite curiosity and the imagination, but the parallels that really demand our attention are the structural similarities and the loads they bear for the societies that tell them.

The Haudenosaunee story is the one I know best. (The Haudenosaunee is the larger confederacy of which the Seneca-Cayuga is a part.) It illustrates this principle very well. It contains wisdom about planting seasons, herbs, agriculture, and our values connected to animals and nature. When we were once warring nations, this creation story in our DNA helped us to lay down our weapons and form a confederacy based on our great law of peace, which was coded in the twelfth century with wampum shells; this law is sometimes referred to as the Gayanashagowa. The laws, called a constitution, are divided into 117 articles. The united Haudenosaunee nations are symbolized by an eastern white pine: the Tree of Peace. Each nation or tribe plays a delineated role in the conduct of government. We believe the events of this formation date to around 1190.

This story survived the forced removal of my tribal members from our ancestral lands of the Northern Woodland Tribal region to our reservation in Grove, Oklahoma. There is a lot of complicated history of our forced migration—of the Seneca-Cayuga, the Haudenosaunee, and all of the 39 First American Nations in Oklahoma. This complicated history isn't supported through concise comment here. A creation story, like a people, survives because of our ancestors and is preserved for our descendants.

The first time I heard this creation story was at Strawberry Festival, at Ganondagan, in Victor, New York. My mother was the second-ever director of education there, she studied to be a storyteller for our tribe from elders, and this wasn't the first time I'd heard this story. My mom told me this story maybe a hundred times as she practiced at home, but the story is so long and expanded, the first time I heard it from her teacher was at Strawberry Festival, a public event, open to non-Indigenous people, in an annual Haudenosaunee festival when we gather and play games and listen day after day to the way our world was made. I didn't hear the whole story in one sitting or even two. It is a very long story, and one that requires us to slow down and celebrate the summer season.

There are a few lessons from this story that I feel need to be a part of my practice as an AI educator and artist.

Where does our knowledge come from? Like Skywoman, we today come from the connected tissue of the Internet, which played a large role in the formation of the insight and information many of our models now are trained from; these were created and gathered without the consent of the communities that created these assets—these words, images, and songs—and this must be acknowledged.

Where did we come from? While much of the AI funding and research is still heavily overrepresented by the world's superpowers, AI must be used as a tool to build an equitable world for peace and not to reinforce ancient colonial borders and continue the process of colonization. Colonization is the tool that created our current water crisis and climate change. To support the survival of our planet, we need AI to be on the side of a decolonized worldview.

What plants and seeds are we giving it ("the" AI)? Right now we are feeding these models images, texts, and connections (commerce) created on the Internet, and of course the written corpora of books that have been scanned and digitized, but there are still many things that make us human—our embodied intelligence, our bodily limitations, and our values around truth and trustworthiness—that are not yet adequately represented by the features that can be extracted from us via observation using contemporary data technologies.

We have many ways of digitizing our daily lives, but our digital systems still cannot account for what philosopher Frank Jackson identified in his famous thought experiment called "What Mary Doesn't Know" (not to be confused with the movie *There's Something About Mary*). For those unfamiliar with that thought experiment, Jackson imagines a scientist named Mary who is raised in a lab. Mary knows everything there is to know about human sight—how to measure it, how to understand light, how our biology responds to it. She

understands every sensor and computational way of evaluating sight. She is essentially a sight expert. But in her world, there is no color. She lives in a black-and-white lab and sees everything via the Internet or cameras as black and white. When she leaves the lab, what is it that she learns? When she sees the color red, what is the thing she did not know before? Jackson calls this missing knowledge *qualia*. It is the actual experience of, in this case, seeing red.

Our modern AI systems are sophisticated statistical and computational models that we try to deploy in order to approximate qualia, but they can't do it. We are looking to define this experience, and it is vital that we can communicate what it is to be a person and our ways of knowing so that they are defendable, so we make sure the AI tools are helping to make *our* world better.

If we get this wrong, we risk creating not simply a world of machines that are unable to match the epistemology of humans (we already have that) but a world in which we adopt a warped, lossy machine epistemology that flattens us into black and white and encourages us to devalue the qualia of our own sentience.

What about the two twins? One who makes the world better and one who makes it worse? There is of course a duality built into the way we discuss AI, which you can hear even in the tension between the two terms *artificial* and *intelligence*. AI has the ability to read the data points we are creating by the billions daily. Humans do not have this ability—an individual human could spend their entire life trying to interpret this vastness of data and they would die in abjection and failure.

Because of this disparity between human and machine computational power, it becomes difficult for us to understand the black-box decisions that occur in the algorithms. Often experts in adjacent fields do not get to weigh in on the process before these new tools are rolled out.

People often ask me what the AI thinks about this or what the AI thinks about that. As if there is *one* AI! As if we have jumped past this phase of experimentation and regulation and arrived at a sentience where the one true AI speaks and we only listen and maybe provide commentary in hushed, reverent tones.

Other times, people say they enjoy co-creating with AI art generators because it is like collaborating with all the artists in the world. Except it is not. As someone who has done the sometimes painful work of collaborating with many artists, artistic collaboration is not the same as typing prompts and then stealing bits of art from others to make a new AI collage—even if it has five fingers on each hand!

> When you ask an AI generator to paint you a night sky, you will likely see something found in the Western canon drawn back for you; when you ask ChatGPT to write you a poem, it rhymes in English. We are not co-creating with everyone but with a particular worldview. The magic we could do is interrupted through this tool, which hides human workers/moderators and in return presents us with something "from scratch" (that has been stolen).
>
> We certainly have seen the ways in which AI has been used to track people and perpetuate biases in law enforcement and sentencing, as well as in hiring, lending, and the financial system generally.

That is our present, but to look toward our future, I think we need to look at creation and the myth we wrap it in; for this new AI age, we need to think of a collective creation myth, a myth that does not cast AI as the aristocrats, making art and writing poetry and deciding our political destinies while we humans are left to toil in unsafe conditions at nine-to-five jobs earning a minimum wage that cannot keep us out of poverty.

I work with AI because I believe it is a tool that can help us understand, simulate, and interrogate climate data in a way that can help us adjust to our climate crisis. I believe data and science storytelling need to move outside of the walls of academia and newsrooms and into every area of expression so we can hear unfiltered climate stories from those who are experiencing the climate crisis most acutely.

We do not know yet what the creation story is for AI. But I hope that in the thoughts I have laid out today, we can begin to orient the conversations we have about this technology in a way that is strategic and truthful and appreciates the real stakes of the stories we tell and the myths we are making about AI.

AMELIA WINGER-BEARSKIN is an associate professor of AI and the arts, and a Banks Preeminence Chair, at the University of Florida.