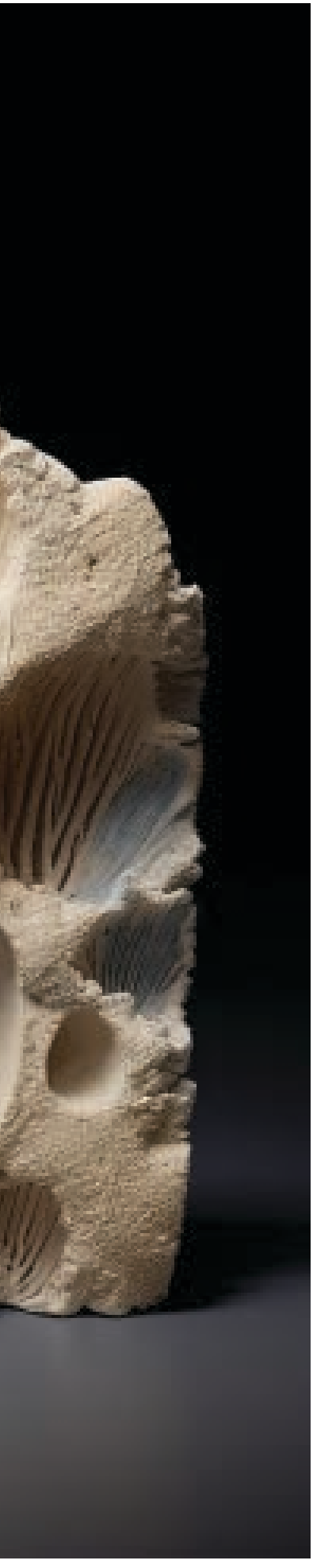
AMY KARLE, *BioAI Mycelium Grown Into the Form of Insulators*, 2023



AN AI SOCIETY

Artificial intelligence is reshaping society, but human forces shape AI. Social scientists and humanities experts explore how to harness the interaction, revealing urgent avenues for research and policy.

How Generative AI Endangers Cultural Narratives

JILL WALKER RETTBERG

Sometime last summer, I needed to install a new dryer in my home in Bergen, Norway. I opened a localized version of Google and typed a request for instructions in Norwegian. Everything the search engine returned was irrelevant—most results assumed my dryer relied on gas, which is not a thing in Norway. Even refining responses for electric dryers assumed configurations that do not exist in my country. I realized that these useless results must be machine-translated from elsewhere. They appeared Norwegian, but they couldn't help me get a dryer running in Norway. In this case, the solution was trivial: a trip to a neighborhood hardware store got me wired in.

But my experience underscores an underappreciated risk that comes with the spread of generative artificial intelligence: the loss of diverse cultural narratives, content, and heritage. Failing to take the cultural aspects of generative AI seriously is likely to result in the streamlining of human expression into the patterns of the largely American content that these systems are trained on.

As generative AI is integrated into everyday tools such as word processors and search engines, it's time to think about what kinds of

Amy Karle: AI-Assisted Biodesign

"The future with AI does not have to be something that happens to us, it is something that we can cocreate." —*Amy Karle*

AMY KARLE, *AI Coral Bioforms*, 2023

Amy Karle is a contemporary artist who uses artificial intelligence as both a medium and a subject in her work. Karle has been deeply engaged with AI, artificial neural networking, machine learning, and generative design since 2015. She poses critical questions about AI, illuminates future visions, and encourages us to actively shape the future we desire.

One of Karle's projects focuses on how AI can help design and grow biomaterials and biosubstrates, including guiding the growth of mycelium-based materials. Her approach uses AI to identify, design, and develop diverse bioengineered and bioinspired structures and forms and to refine and improve the structure of biomaterials for greater functionality and sustainability. Another project is inspired by the seductive form of corals. Karle's speculative biomimetic corals leverage AI-assisted biodesign in conjunction with what she terms "computational ecology" to capture, transport, store, and use carbon dioxide. Her goal with this series is to help mitigate carbon dioxide emissions from industrial sources such as power plants and refineries and to clean up highly polluted areas.

continued from page 77

stories it can generate—and what stories it will *not* generate. It's no secret that AI is biased. Researchers recently asked the image generator Midjourney to create images of Black physicians treating impoverished white children, but the system would only return images depicting the children as Black. Even after several iterations, Midjourney failed to produce the specified results. The closest it got to the prompt was a shirtless medicine man with feathers, leather bands, and beads, gazing at a similarly garbed blond child.

Here's something that hits close to home: the potential loss of Cardamom Town. Thorbjørn Egner's *Folk og røvere i Kardemomme by* (*When the Robbers Came to Cardamom Town*) is a children's book and musical well known to anyone who grew up in Norway or Denmark after 1955. The songs and stories have been played, read, and sung in homes and preschools for decades; there's even a theme park inspired by the book in the city of Kristiansand. The story features three comical thieves who steal food because they are hungry and don't understand that work is necessary. After being caught stealing sausages and chocolate, they are rehabilitated by the kind police officer and townsfolk, then end up saving the town from a fire.

This story is more than a shared cultural reference—it supports the Norwegian criminal justice system's priority of rehabilitation over punishment. It is distinct from Disney movies, with their unambiguous villains who are punished at the end, and from Hollywood bank heists and gangster movies that glorify criminals. Generative AI might well bury stories like Cardamom Town by stuffing chatbot responses and search results worldwide with homogenized American narratives.

Narrative archetypes give us templates to live by. Depending on the stories we hear, share, and create, we shape possibilities for action and for understanding. We learn that criminals can be rehabilitated, or that they deserve to come to a bad end.

The humanities and social sciences have studied and critiqued AI for a long time, but almost all *development* of AI has happened within quantitative disciplines: computer science, data science, statistics, and mathematics. The current wave of AI is based on language, narratives, and culture; unchecked, this wave threatens to impoverish the world's cultural narratives. We have reached a point where AI development needs the humanities. Not just so I can figure out how to install my appliances, but so we don't lose the stories that shape our communities.

Jill Walker Rettberg *is a professor of digital culture and codirector of the Center for Digital Narrative at the University of Bergen, Norway, and the author of* Machine Vision: How Algorithms Are Changing the Way We See the World *(Polity Press, 2023).*

Generative AI Is a Crisis for Copyright Law

KATE CRAWFORD AND JASON SCHULTZ

Generative artificial intelligence is driving copyright into a crisis. More than a dozen copyright cases about AI were filed in the United States last year, up severalfold from all filings from 2020 to 2022. In early 2023, the US Copyright Office launched the most comprehensive review of the entire copyright system in 50 years, with a focus on generative AI. Simply put, the widespread use of AI is poised to force a substantial reworking of how, where, and to whom copyright should apply.

Starting with the 1710 British statute, "An Act for the Encouragement of Learning," Anglo-American copyright law has provided a framework around creative production and ownership. Copyright is even embedded in the US Constitution as a tool "to promote the Progress of Science and useful Arts." Now generative AI is destabilizing the foundational concepts of copyright law as it was originally conceived.

Typical copyright lawsuits focus on a single work and a single unauthorized copy, or "output," to determine if infringement has occurred. When it comes to the capture of online data to train AI systems, the sheer scale and scope of these datasets overwhelms traditional analysis. The LAION 5-B dataset, used to train the AI image generator Stable Diffusion, contains 5 billion images and text captions harvested from the internet, while CommonPool (a collection of datasets released by nonprofit LAION in April to democratize machine learning), offers 12.8 billion images and captions. Generative AI systems have used datasets like these to produce billions of outputs.

For many artists and designers, this feels like an existential threat. Their work is being used to train AI systems, which can then create images and texts that replicate their artistic style. But to date, no court has considered AI training to be copyright infringement: following the Google Books case in 2015, which assessed scanning books to create a searchable index, US courts are likely to find that training AI systems on copyrighted works is acceptable under the fair use exemption, which allows for limited use of copyrighted works without permission in some cases when the use serves the public interest. It is also permitted in the European Union under the text and data mining exception of EU digital copyright law.

Copyright law has also struggled with authorship by AI systems. Anglo-American law presumes that work has an "author" somewhere. To encourage human creativity, some authors need the economic incentive of a time-limited monopoly on making, selling, and showing their work. But algorithms don't need incentives. So according to the US Copyright Office they aren't entitled to copyright. The same reasoning applied to other cases involving nonhuman authors, including the case where a macaque took selfies using a nature photographer's camera. Generative AI is the latest in a line of nonhumans deemed unfit to hold copyright.

Nor are human prompters likely to have copyrights in AI-generated work. The algorithms and neural net architectures behind generative AI algorithms produce outputs that are inherently unpredictable, and any human prompter has less control over a creation than the model does.

AMY KARLE, *BioAI-Formed Mycelium*, 2023

Where does this leave us? For the moment, in limbo. The billions of works produced by generative AI are unowned and can be used anywhere, by anyone, for any purpose. Whether a ChatGPT novella or a Stable Diffusion artwork, output now exists as unclaimable content in the commercial workings of copyright itself. This is a radical moment in creative production: a stream of works without any legally recognizable author.

There is an equivalent crisis in *proving* copyright infringement. Historically, this has been easy, but when a generative AI system produces infringing content, be it an image of Mickey Mouse or Pikachu, courts will struggle with the question of who is initiating the copying. The AI researchers who gathered the training dataset? The company that trained the model? The user who prompted the model? It's unclear where agency and accountability lie, so how can courts order an appropriate remedy?

Copyright law was developed by eighteenth-century capitalists to intertwine art with commerce. In the twenty-first century, it is being used by technology companies to allow them to exploit all the works of human creativity that are digitized and online. But the destabilization around generative AI is also an opportunity for a more radical reassessment of the social, legal, and cultural frameworks underpinning creative production.

What expectations of consent, credit, or compensation should human creators have going forward, when their online work is routinely incorporated into training sets? What happens when humans make works using generative AI that cannot have copyright protection? And how does our understanding of the value of human creativity change when it is increasingly mediated by technology, be it the pen, paintbrush, Photoshop, or DALL-E?

It may be time to develop concepts of intellectual property with a stronger focus on equity and creativity as opposed to economic incentives for media corporations. We are seeing early prototypes emerge from the recent collective bargaining agreements for writers, actors, and directors, many of whom lack copyrights but are nonetheless at the creative core of filmmaking. The lessons we learn from them could set a powerful precedent for how to pluralize intellectual property. Making a better world will require a deeper philosophical engagement with what it is to create, who has a say in how creations can be used, and who should profit.

Kate Crawford *is a research professor at the University of Southern California Annenberg, a senior principal researcher at Microsoft Research, the inaugural chair of AI & Justice at the École Normale Supérieure, and the author of* Atlas of AI *(Yale University Press, 2021).* ***Jason Schultz*** *is a clinical professor of law at New York University, the codirector of the Engleberg Center on Innovation, Law, and Policy, and the author of* The End of Ownership *(The MIT Press, 2016).*

AMY KARLE, *AI Bioforms for Carbon Capture*, 2023

AI Lacks Ethic Checks for Human Experimentation

LINNET TAYLOR

Following Nazi medical experiments in World War II and outrage over the US Public Health Service's four-decade-long Tuskegee syphilis study, bioethicists laid out frameworks, such as the 1947 Nuremberg Code and the 1979 Belmont Report, to regulate medical experimentation on human subjects. Today social media—and, increasingly, generative artificial intelligence—are constantly experimenting on human subjects, but without institutional checks to prevent harm.

In fact, over the last two decades, individuals have become so used to being part of large-scale testing that society has essentially been configured to produce human laboratories for AI. Examples include experiments with biometric and payment systems in refugee camps (designed to investigate use cases for blockchain applications), urban living labs where families are offered rent-free housing in exchange for serving as human subjects in a permanent marketing and branding experiment, and a mobile money research and development program where mobile providers offer their African consumers to firms looking to test new biometric and fintech applications. Originally put forward as a simpler way to test applications, the convention of software as "continual beta" rather than more discrete releases has enabled business models that depend on the

creation of laboratory populations whose use of the software is observed in real time.

This experimentation on human populations has become normalized, and forms of AI experimentation are touted as a route to economic development. The Digital Europe Programme launched AI testing and experimentation facilities in 2023 to support what the program calls "regulatory sandboxes," where populations will interact with AI deployments in order to produce information for regulators on harms and benefits. The goal is to allow some forms of real-world testing for smaller tech companies "without undue pressure from industry giants." It is unclear, however, what can pressure the giants and what constitutes a meaningful sandbox for generative AI; given that it is already being incorporated into the base layers of applications we would be hard-pressed to avoid, the boundaries between the sandbox and the world are unclear.

Generative AI is an extreme case of unregulated experimentation-as-innovation, with no formal mechanism for considering potential harms. These experiments are already producing unforeseen ruptures in professional practice and knowledge: students are using ChatGPT to cheat on exams, and lawyers are filing AI-drafted briefs with fabricated case citations. Generative AI also undermines the public's grip on the notion of "ground truth" by hallucinating false information in subtle and unpredictable ways.

These two breakdowns constitute an abrupt removal of what philosopher Regina Rini has termed "the epistemic backstop,"—that is, the benchmark for considering something real. Generative AI subverts information-seeking practices that professional domains such as law, policy, and medicine rely on; it also corrupts the ability to draw on common truth in public debates. Ironically, that disruption is being classed as success by the developers of such systems, emphasizing that this is not an experiment we are conducting but one that is being conducted upon us.

This is problematic from a governance point of view because much of current regulation places the responsibility for AI safety on individuals, whereas in reality they are the subjects of an experiment being conducted across society. The challenge this creates for researchers is to identify the kinds of rupture generative AI can cause and at what scales, and then translate the problem into a regulatory one. Then authorities can formalize and impose accountability, rather than creating diffuse and ill-defined forms of responsibility for individuals. Getting this right will guide how the technology develops and set the risks AI will pose in the medium and longer term.

Much like what happened with biomedical experimentation in the twentieth century, the work of defining boundaries for AI experimentation goes beyond "AI safety" to AI legitimacy, and this is the next frontier of conceptual social scientific work. Sectors, disciplines, and regulatory authorities must work to update the definition of experimentation so that it includes digitally enabled and data-driven forms of testing. It can no longer be assumed that experimentation is a bounded activity with impacts only on a single, visible group of people. Experimentation at scale is frequently invisible to its subjects, but this does not render it any less problematic or absolve regulators from creating ways of scrutinizing and controlling it.

Linnet Taylor *is a professor of international data governance at Tilburg University, Netherlands, and leads the European Research Council-funded Global Data Justice project.*

AMY KARLE, *AI Coral Bioforms*, 2023

AI Aids the Pretense of Military "Precision"

LUCY SUCHMAN

Artificial intelligence is the latest promise of a technological solution to the intractable "fog of war." In Ukraine and Gaza, enthusiasts have proclaimed the advent of AI-driven warfighting. In October 2023, Ukrainian technologists confirmed that AI-enabled drones identify and target 64 types of Russian "military objects" without a human operator; meanwhile the Israeli Defense Forces website states that an AI system generates recommended targets, reportedly at an unprecedented rate. Enormous questions arise regarding the validity of the assumptions built into these systems about who comprises an imminent threat and about the legitimacy of

their targeting functions under the Geneva Conventions and the laws of war.

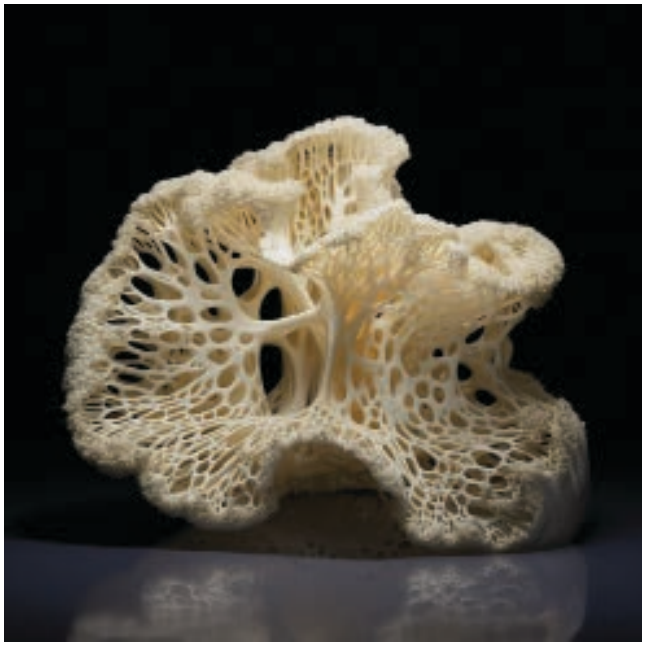
Considering military investments in AI as part of a sociotechnical imaginary is helpful here. Developed within the field of science and technology studies, the concept of sociotechnical imaginaries describes collectively imagined forms of social order as materialized by scientific and technological projects. These include aspirational futures that sustain investments in the military-industrial-academic complex. Iconic examples of AI-enabled warfighting in the present moment include battle management interfaces like Palantir's AI platform.

To function in the real world, these platforms require very large, up-to-date datasets (of labeled "military objects" or biometric profiles of "persons of interest," for example), from which models can be developed. In the case of threat prediction and targeting, neither the US Department of Defense nor allied militaries make public the details necessary to assess validity. But in the case of predictive policing, an investigation by *The Markup* found that fewer than 1% of data-based predictions actually lined up with reported crimes. And generative AI introduces new uncertainties: both the provenance of the data and reliability of information are hard to check. That is particularly dangerous for "actionable military intelligence," which is used for targeting and to designate imminent threats.

We should be deeply skeptical of the promotion of AI as a solution to the fog of war, which imagines that the right technology will find the important signals amid the noise. This faith in technology constitutes a kind of willful ignorance, as if AI is a talisman that sustains the wider magical thinking of militarism as a path to security. In the words of performance artist Laurie Anderson (quoting her meditation teacher), "If you think technology will solve your problems, then you don't understand technology—and you don't understand your problems."

Critical inquiry into the realities of war can help challenge the logics through which militarism perpetuates its imaginary of rational and controllable state violence while obscuring war's ungovernable chaos and unjustifiable injuries. Although there are valid reasons that military forces exist in today's world, we should question the narratives that underwrite the billions of dollars funnelled into algorithmically based warfighting. We need to redirect resources to creative projects in de-escalation, negotiated settlements that offer true security for all, and eventual demilitarization. While the techno-solutionist imaginaries of militarism are longstanding, so are their limits as a basis for sustainable peace.

Lucy Suchman *is professor emerita of the anthropology of science and technology at Lancaster University in the United Kingdom.*



AMY KARLE, *BioAI-Formed Mycelium*, 2023

Protect Information Systems to Preserve Attention

MARK ANDREJEVIC

Already, content generated by artificial intelligence populates the advertisements, news, and entertainment people see every day. According to OpenAI's cofounder Greg Brockman, the technology could fundamentally transform mass culture, making it possible, for example, to customize TV shows for individual viewers: "Imagine if you could ask your AI to make a new ending … maybe even put yourself in there as a main character."

Brockman meant this as a sort of paradise of customization, but it's not hard to see how such tools could also spew misinformation and other content that would disrupt civic life and undermine democracy. Bad content would drive out good, enacting "Gresham's Law"—the principle that "bad money drives out good"—on steroids. Even top AI executives are begging for regulation, albeit at the level of individual products and their potential dangers. I think a more productive way to frame regulation is as a means of protecting the shared information environment.

In decades past, the rationale for regulating the information space pivoted on the limited availability of broadcast channels, or "channel scarcity." Public attention can also be considered a finite resource, rationed by what information theorist Tiziana Terranova describes as "the limits inherent to the neurophysiology of perception and

the social limitations to time available for consumption." For democracy to function, people need to pay attention to matters of public import. In an information environment swamped with automatically generated content, attention becomes the scarce resource.

A world in which attention is monopolized by an endless flow of personalized entertainment might be a consumers' paradise—but it would be a citizen's nightmare. The tech sector has already proposed a model for dispensing with public attention, one that is far from democratic. In 2016, a team at Google envisioned a "Selfish Ledger"—a data profile that would infer individuals' goals and then prompt aligned behavior, such as buying healthier food or locally grown produce, and seek more data to tweak the customized model. Similarly, physicist César Hidalgo suggested providing every citizen with a software agent that could infer political preferences and act on their behalf. In such a world, the algorithm would pay attention for us: no need for people to learn about the issues or even directly express their opinions.

Such proposals show how important it is for citizens to actively regulate the information commons. Preserving scarce attention is essential to recapturing an increasingly elusive sense of shared, overlapping, and common interests. The world is moving toward a state where the data we generate can be used to further capture and channel our attention according to priorities that are neither our own, nor those of civic life. Software, and whoever it serves, cannot be allowed to substitute for citizenship, and the economic might of tech giants must be balanced by citizens' ability to access the information they need to exercise their political power.

Mark Andrejevic *is a professor of communication and media studies in the School of Media, Film, and Journalism at Monash University, Australia. He is the author of* Facial Recognition *(Polity Press, 2022).*

Get Citizens' Input on AI Deployments

KARINE GENTELET

As part of my job, I give talks about how artificial intelligence affects human rights: to criminology experts, schoolteachers, retirees, union members, First Peoples, and more. Across these diverse groups, I hear common themes. One is that although AI programs could impact how they do their jobs and live their lives, people feel their experience and expertise are completely left out before programs are deployed. Some worry, legitimately, about facing legal action if they protest.

Plans and policies to regulate AI systems in Europe, Canada, and the United States are not likely to improve the situation. Europe plans to assign regulatory requirements based on application. For example, the high-risk category includes technology used in hiring decisions, police checks, banking, and education. Canadian legislation, still under review by the House of Commons, is based on the same risk assessment. The US president has outlined demands for rigorous safety testing, with results reported to the government. The problem is that these plans focus on laying out guardrails for anticipated threats without establishing an early warning system for citizens' actual experiences or concerns.

Regulatory schemes based on a rigid set of anticipated outcomes might be a good first step, but they are not enough. For one thing, some harms are only now emerging. And they could become most entrenched for marginalized, underserved groups because generative AI is trained on biased datasets that then generate new datasets that perpetuate the vicious cycle. A 2021 paper shows how prediction tools in education systems incorporate not just statistical biases (by gender, race, ethnicity, or language) but also understudied sociological ones such as urbanity. For instance, rural learners in Brazil are likely to differ from their urban counterparts with regards to fluency in the official state language and their access to relevant educational materials, up-to-date facilities, and teaching staff. But because there aren't enough data on specific groups' learning and schooling issues, their needs would be aggregated into a larger dataset and made invisible. Given the lack of knowledge, it would be difficult to even predict any kind of bias.

What's needed are mechanisms that support citizens' direct engagement with AI deployments to document, from the ground, potentially high-risk impacts on collective equity. There are democratic formats already in place to support citizens' perspectives. In Canada, for example, the mandate of the general solicitor or privacy commissioner could be strengthened to review AI deployments in the public sector (audits of datasets, mandatory impact assessments, etc.). These mechanisms would provide transparent and accountable standards to keep citizens adequately informed about AI deployments, help balance the civic power dynamic, and strengthen social justice.

Citizens' direct engagement could also be supported through access to courts. There are few (if any) direct legal recourses available for ordinary people to challenge algorithmic harms in current AI regulatory schemes. Access to courts—and implicitly to justice—could send a clear message about citizens' power to corporations, governments, and, most importantly, to citizens

themselves. In combination with other mechanisms to increase citizen oversight, legal suits would offer not only access to rightful reparations, but also give societal recognition of citizens' rights.

Sometimes at my talks people tell me they feel illegitimate asking questions about AI's impacts, given their lack of expertise. What I tell them is that they don't need to be a mechanic to know how bad it would be to be hit by a car. Harms from AI are bound to be more subtle, but the point stands. Citizens are the ones primarily affected, so they must have an active role within AI governance. Emerging regulatory systems should highlight the role of citizens as social actors who contribute—as they should—to the collective good.

Karine Gentelet *is an associate professor of social sciences at the Université du Québec en Outaouais, Gatineau, Canada.*

AMY KARLE, *Cell Forms* (AI-assisted design), 2023

The Question Isn't Asset or Threat; It's Oversight

EMMANUEL DIDIER

As part of a research group studying generative AI with France's Académie Nationale de Médecine, I was surprised by some clinicians' technological determinism—their immediate assumption that this technology would, on its own, act against humans' wishes. The anxiety is not limited to physicians. In spring 2023, thousands of individuals, including tech luminaries such as Elon Musk and Steve Wozniak, signed a call to "pause giant AI experiments" to deal with "profound risks to society."

But the question is more complex than restraint versus unfettered technological development. It is about different ways to articulate ethical values and, above all, different visions of what society should be.

A double interview in the French journal *Le Monde* illustrates the distinction. The interviewees, Yoshua Bengio and Yann Le Cun, are friends and collaborators who both received the 2018 Turing Award for their contributions to computer science. But they have radically different views on the future of generative AI.

Bengio, who works at a nonprofit AI think tank in Montreal, believes ChatGPT is revolutionary. That's why he sees it as dangerous. ChatGPT and other generative AI systems work in ways that cannot be fully understood and often produce results that are simultaneously wrong and credible, which threatens news and information sources and democracy at large. His argument mirrors philosopher Hans Jonas's precautionary principle: since humanity is better at producing new technological tools than foreseeing their future consequences, extreme caution about what AI can do to humanity is warranted. The solution is to establish ethical guidelines for generative AI, a task that the European Group on Ethics, the Organisation for Economic Co-operation and Development, UNESCO, and other global entities have already embraced.

Le Cun, who works for Meta, does not consider ChatGPT revolutionary. It depends on neural networks trained on very large databases—all technologies that are several years old. Yes, it can produce fake news, but dissemination—not production—is the real risk. Techniques can be developed to flag AI-generated outputs and reveal what text and images have been manipulated, creating something akin to antispam software today. For Le Cun, the way to quash the dangers of generative AI will rely on AI. It is not the problem but the solution—a tool humanity can use to make better decisions. But who defines what is a "better decision"? Which set of values will prevail? Here I see in Le Cun's arguments parallels to the economist and innovation scholar Joseph Schumpeter, who argued that within a democracy, the tools humans use to institutionalize values are the law and government. In other words, regulation of AI is essential.

These radically disparate views land on solutions that are similar in at least one aspect: whether generative AI is seen as a technological revolution or not, it is always embedded within a wider set of values. When seen as a danger for humanity, ethics are mobilized. When social values are threatened, the law is brought in. Either way, the solution is oversight of the corporations building AI.

This opens a door for the public to weigh in on future developments of generative AI. A first step is to identify interests and stakeholders clustering in each position and draw them into how to better inform the development and regulation of AI. As with every other technological advance, humans obviously can decide things in their own way

Emmanuel Didier *is a full professor in the Centre Maurice Halbwachs at the École Normale Supérieure, France, where he studies the socio-history of quantification. He is author of* America by the Numbers: Quantification, Democracy, and the Birth of National Statistics *(The MIT Press, 2020).*

Ground Truths Are Human Constructions

FLORIAN JATON

Artificial intelligence algorithms are human-made, cultural constructs, something I saw first-hand as a scholar and technician embedded with AI teams for 30 months. Among the many concrete practices and materials these algorithms need in order to come into existence are sets of numerical values that enable machine learning. These referential repositories are often called "ground truths," and when computer scientists construct or use these datasets to design new algorithms and attest to their efficiency, the process is called "ground-truthing."

Understanding how ground-truthing works can reveal inherent limitations of algorithms—how they enable the spread of false information, pass biased judgments, or otherwise erode society's agency—and this could also catalyze more thoughtful regulation. As long as ground-truthing remains clouded and abstract, society will struggle to prevent algorithms from causing harm and to optimize algorithms for the greater good.

Ground-truth datasets define AI algorithms' fundamental goal of reliably predicting and generating a specific output—say, an image with requested specifications that resembles other input, such as web-crawled images. In other words, ground-truth datasets are deliberately constructed. As such, they, along with their resultant algorithms, are limited and arbitrary and bear the sociocultural fingerprints of the teams that made them.

Ground-truth datasets fall into at least two subsets: input data (what the algorithm should process) and output targets (what the algorithm should produce). In supervised machine learning, computer scientists start by building new algorithms using one part of the output targets annotated by human labelers, before evaluating their built algorithms on the remaining part. In the unsupervised (or "self-supervised") machine learning that underpins most generative AI, output targets are used only to evaluate new algorithms.

Most production-grade generative AI systems are assemblages of algorithms built from both supervised and self-supervised machine learning. For example, an AI image generator depends on self-supervised diffusion algorithms (which create a new set of data based on a given set) and supervised noise reduction algorithms. In other words, generative AI is thoroughly dependent on ground truths and their socioculturally oriented nature, even if it is often presented—and rightly so—as a significant application of self-supervised learning.

Why does that matter? Much of AI punditry asserts that we live in a post-classification, post-socially constructed world in which computers have free access to "raw data," which they refine into actionable truth. Yet data are never raw, and consequently actionable truth is never totally objective.

Algorithms do not create so much as *retrieve* what has already been supplied and defined—albeit repurposed and with varying levels of human intervention. This observation rebuts certain promises around AI and may sound like a disadvantage, but I believe that it could instead be an opportunity for social scientists to begin new collaborations with computer scientists. This could take the form of a professional social activity, people working together to describe the ground-truthing processes that underpin new algorithms, and so help make them more accountable and worthy.

Florian Jaton *is a senior researcher and lecturer in sociology of science and technology at the Geneva Graduate Institute of International and Development Studies, Switzerland. He is the author of* The Constitution of Algorithms: Ground-Truthing, Programming, Formulating *(The MIT Press, 2021).*

History Can Help Us Chart AI's Future

XIAOCHANG LI

Current technical approaches to preventing harm from artificial intelligence and machine learning largely focus on bias in training data and careless (even malicious) misuse. To be sure, these are crucial steps, but they are not sufficient solutions. Many risks from AI are not simply due to flawed executions of an otherwise sound strategy: AI's

penchant for enabling bias and misinformation is built into its "data-driven" modeling paradigm.

This paradigm forms the foundation of present-day machine learning. It relies on data-intensive pattern recognition techniques that generalize from past examples without direct reference to, or even knowledge about, what is being modeled. In other words, data-driven methods are designed to predict the probable output of processes that they can't describe or explain. That deliberate omission of explanatory models leaves these methods particularly receptive to misdirection.

Today, this data-intensive, brute-force approach to machine learning has become largely synonymous with artificial intelligence and computational modeling as a whole. Yet history shows that the rise of data-driven machine learning was neither natural nor inevitable. Even machine learning itself was not always so data-centric. Today's dominant paradigm of data-driven machine learning in key areas such as natural language processing represents what Alfred Spector, then Google's vice president for research, lauded in 2010 as "almost a 180-degree turn in the established approaches to speech recognition."

Through its early decades, AI research in the United States fixated on replicating human cognitive faculties, based on an assumption that, as historian Stephanie Dick puts it, "computers and minds were the same kind of thing." The devotion to this human analogy began to change in the 1970s with a highly unorthodox "statistical approach" to speech recognition at IBM. In a stark departure from the established "knowledge-based" approaches of the period, IBM researchers abandoned elaborate formal representations of linguistic knowledge and used statistical pattern recognition techniques to predict the most likely sequence of words, based on large quantities of sample data. Those very researchers described to me how this work owed much of its success to the unique computing resources available at IBM, where they had access to more computing power than anyone else. Even more importantly, they had access to more training data in a period where digitized text was vanishingly scarce by today's standards. During a federal antitrust case against the company from 1969 to 1982, IBM had manually digitized over 100,000 pages of witness testimony using a warehouse facility full of keypunch operators to manually encode text onto Hollerith punched cards. This material was repurposed into a training corpus of unprecedented size for the period, at around 100 million words.

What resulted was an abandonment of knowledge-based approaches aimed at simulating human decision processes in favor of data-driven approaches aimed solely at predicting their output. This signaled a fundamental reimagining of the relation between human and machine intelligence. Director of IBM's Continuous Speech Recognition group Fred Jelinek described their approach in 1987 as "the natural way for the machine," quipping that "if a machine has to fly, it does so as an airplane does—not by flapping its wings."

The success of this approach directly triggered a shift to data-driven approaches across natural language processing as well as machine vision, bioinformatics, and other domains. In 2009, top Google researchers pointed the earlier success of the statistical approach to speech recognition as proof that "invariably, simple models and a lot of data trump more elaborate models based on less data."

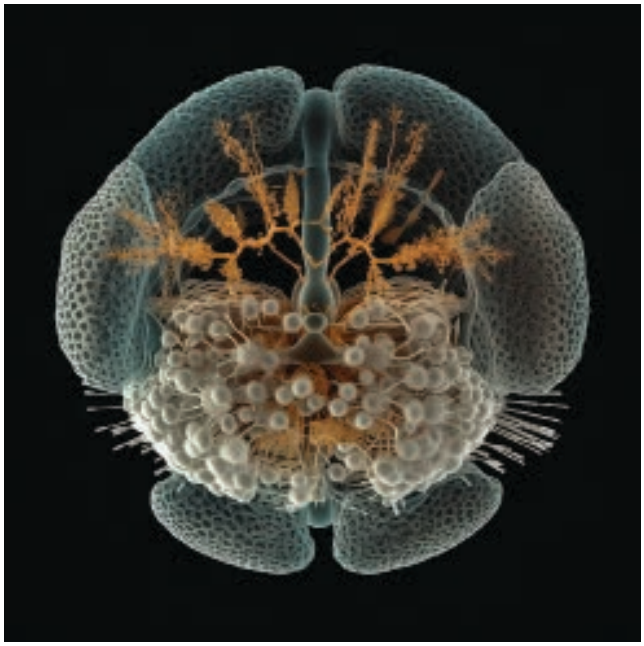
Framing machine intelligence as something fundamentally distinct from, if not antithetical to, human understanding set a powerful precedent for replacing expert knowledge with data-driven approximation in computational modeling. Generative AI takes this logic a crucial step further, using data not only to model the world, but to actively remake it.

> Many risks from AI are not simply due to flawed executions of an otherwise sound strategy: AI's penchant for enabling bias and misinformation is built into its "data-driven" modeling paradigm.

Large language models are both ignorant of and indifferent toward the substance of the statements they generate; they gauge only how likely it is for a sequence of text to appear. Which is to say, if the results pushed to our social media feeds are decided by algorithms that are intentionally designed to only predict patterns, but not to understand them, can the flourishing of misinformation really come as such a surprise?

A failure to recognize how such problems may be intrinsic to the very logic of data-driven machine learning inspires oft-misguided technical fixes, such as increased data collection and tracking, which can lead to harms such as predatory inclusion (in which outwardly democratizing schemes further exploit already marginalized groups). Such approaches are limited because they presume *more* machine learning to be the best recourse.

But the lens of history helps us break out of this circular thinking. The perpetual expansion of data-driven machine learning should not be seen as a foregone conclusion. Its rise to prominence was embedded in certain assumptions and



AMY KARLE, *Cell Forms* (AI-assisted design), 2023

priorities that became entrenched in its technical framework and normalized over time. Instead of defaulting to tactics that augment machine learning, we need to consider that in some circumstances the very logic of machine learning might be fundamentally unsuitable to our aims.

Xiaochang Li *is an assistant professor in the Department of Communication at Stanford University. She is the author of the forthcoming book* Beyond Recognition: Language, Datafication, and the Making of Algorithmic Culture *(University of Chicago Press).*

How AI Sets Limits for Human Aspiration

STEPHANIE DICK, WENDY HUI KYONG CHUN, AND MATT CANUTE

We are watching "intelligence" being redefined as the tasks that an artificial intelligence can do. Time and again, generative AI is pitted against human counterparts, with textual and visual outputs measured against human abilities, standards, and exemplars. AI is asked to mimic, and then to better, human performance on law and graduate school admission tests, advanced placement exams and more—even as those tests are being abandoned because they perpetuate inequality and are inadequate to the task of truly measuring human capacity.

The narratives trumpeting AI's progress obscure an underlying logic requiring that everything be translated into the technology's terms. If it is not addressed, that hegemonic logic will continue to narrow viewpoints, hamper human aspirations, and foreclose possible futures by condemning us to repeat—rather than learn from—past mistakes.

The problem has deep roots. As AI evolved in the 1950s and '60s, researchers often made human comparisons. Some suggested that computers would become "mentors" and "colleagues," others "assistants," "servants," or "slaves." As science and technology scholars Neda Atanasoski, Kalindi Vora, and Ron Eglash have shown, these comparisons shaped the perceived value not only of AI, but also of human labor. Those relegating AI to the latter categories usually did so because they believed computers would be limited to menial, repetitive, and mindless labor. They were also reproducing the fiction that human assistants are merely mechanical, menial, and mindless. On the other hand, those celebrating potential mentors and colleagues were tacitly assuming that human counterparts could be stripped of everything beyond efficient reasoning.

Comparisons between AI and human performance often correlate with social hierarchy. As society and technology scholars Janet Abbate, Mar Hicks, and Alison Adam have shown, in the 1960s and 1970s, women and minorities were encouraged to advance in society by learning to code—but those skills were then devalued, while domains dominated by white men were seen as the realm of the truly technically skilled. More recent OpenAI measures of AI against standardized exams endorse a positivist, adversarial, and bureaucratic understanding of human intelligence and potential. Similarly, AI-generated "case interviews" and artworks encode mimicry as the definition of intelligence. For a result from generative AI to be validated as true—or to shock others as "true"—it has to be plausible, that is, recognizable in terms of past values or experiences. But looking backward and smoothing out outliers forecloses the rich wellsprings of humanity's imagination for the future.

Such practices will ultimately affect who and what is perceived as intelligent, and that will profoundly change society, discourse, politics, and power. For example, in "AI ethics," complex concepts such as "fairness" and "equality" are reconfigured as mathematical constraints on predictions, collapsed onto the underlying logic of machine learning. In another example, the development of machine learning systems for game-playing has led to a reductive redefinition of "play" as simply making permissible moves in search of victory. Anyone who has played Go or chess or poker against another person knowns that, for humans, "play" includes so much more.

The portrayal of AI's history is usually one of progress, where constellations of algorithms attain humanlike general intelligence and creativity. But that narrative might be more accurately inverted with a shrinking definition of intelligence

that excludes many human capabilities. This narrows the horizon of intelligence to tasks that can be accomplished with pattern recognition, prediction from data, and the like. We fear this could set limits for human aspirations and for core ideals like knowledge, creativity, imagination, and democracy—making for a poorer, more constrained human future.

Stephanie Dick *is an assistant professor in the School of Communication at Simon Fraser University, Canada.* ***Wendy Hui Kyong Chun*** *is Simon Fraser University's Canada 150 Research Chair in New Media in the School of Communication and director of the Digital Democracies Institute at Simon Fraser University, Canada. She is the author of* Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition *(The MIT Press, 2021).* ***Matt Canute*** *is a postgraduate researcher in the School of Communication at Simon Fraser University, Canada.*

Make AI a Public Problem

MIKE ANANNY

Often, problems that seem narrow and purely technical are best tackled if they're recast as "public problems," a concept put forth almost a century ago by philosopher and educator John Dewey. Examples of public problems include dirty air, polluted water, global warming, and childhood education. Public problems bring harms that are not always felt individually but that nonetheless shape what it means to be a thriving person in a thriving society. These problems need to be noticed, discussed, and collectively managed. In contrast to problems that are personal, private, or technical, Dewey wrote, public problems happen when people experience "indirect consequences" that need to be collectively and "systematically cared for," regardless of an individual's circumstance, wealth, privilege, or interests. Public problems define our shared realities.

Although generative AI has been framed as a technical problem, recasting it as a public problem offers new avenues for action. Generative AI is quickly becoming a language for telling society's collective stories and teaching us about each other. If you ask generative AI to make a story or video that explains climate change, you are actually asking a probabilistic machine learning model to create a statistically acceptable account of a public problem. Tools such as ChatGPT and Midjourney are fast becoming languages for understanding public problems, but with little analysis of their power to shape the stories that humans use to understand the shared consequences that Dewey told us create public life.

To grapple with generative AI effectively, consumers and developers alike need to see it not only as biased datasets and machine learning run amok—we need to see it as a fast-emerging language that people are using to learn, make sense of their worlds, and communicate with others. In other words, it needs to be seen as a public problem.

First, researchers need to see generative AI as a powerful language—as "boundaries," "infrastructures," and "hinges" that scholars of science and technology tell us create technologies. This means tracing the connections among the people and machines that make synthetic language: engineers who build machine learning systems, for example, entrepreneurs who pitch business models, journalists who make synthetic news stories, audiences who struggle to know what to believe. These are the complex and largely invisible assumptions that make generative AI a language for representing knowledge, fueling innovation, telling stories, and creating shared realities.

Second, as a society, we need to analyze the harms created by generative AI. When statistical hallucinations invent facts, chatbots misattribute authorship, or computational summaries bungle analyses, they produce dangerously wrong language that has all the confidence of a seemingly neutral, computational certainty. These errors are not just rare and idiosyncratic curiosities of misinformation; their real and imagined existence makes people see media as unstable, unreliable, and untrusted. Society's information sources—and ability to gauge reality—are destabilized.

Finally, all members of society should reject the assertions of technology companies and AI "godfathers" who claim that generative AI is both an existential threat and a problem that only technologists can manage. Public problems are collectively debated, accounted for, and managed; they are not the purview of private companies or self-identified caretakers who work on their own timelines with proprietary knowledge. Truly public problems are never outsourced to private interests or charismatic authorities.

A public problem is not merely a technical curiosity, a moral panic, or an inevitable future. It is a system of relationships between people and machines that creates language, makes mistakes, and needs to be systematically cared for. Once we understand generative AI as a vital language for creating shared realities and tackling collective challenges, we can start to see it as a public problem, and then we will be in a better place to solve it.

Mike Ananny *is an associate professor of communication and journalism in the Annenberg School for Communication and Journalism at the University of Southern California. He is the author of* Networked Press Freedom: Creating Infrastructures for a Public Right to Hear *(The MIT Press, 2018).*

These articles arose from a working group on artificial intelligence and justice convened by Kate Crawford at the École Normale Supérieure.