"On Bullshit Engines' Politics"

Dr. Damien P. Williams, PhD[1]

Abstract:

The contemporary world has become increasingly entangled with so-called artificial intelligence— systems that seem to replicate human speech and thought, proclaiming to make our lives easier, more efficient, and more interesting. But what do these tools and systems actually do, and how do they do it? What systems of power, politics, and other values are embedded within these systems, and who put them there? Throughout this chapter, I take inspiration from and deploy Langdon Winner's theories of how politics are embedded within and reproduced by technologies. Specifically, I use Winner's framing to examine values embedded and entangled with "AI" and how those entanglements lead to what has already occurred, and what can potentially go wrong when we integrate "AI" into various aspects of our lives. This discussion and exploration then form the basis for recommendations as to how we might build these systems differently— that is to say, how we might go about building into these systems the values and politics we want to see rather than the ones we allow to make their way in via our uncritical assumptions and unexamined biases. Neither "AI" qua "AI" nor the kyriarchal, hegemonic, environmentally damaging, capitalist, white supremacist, ableist, eugenic, misogynistic, and otherwise prejudicially biased forms these tools and systems currently tend to take are inevitable. Racist facial recognition, sexist and ableist resumé and hiring tools, and other "AI" systems do not care about the harms they perpetuate or the truth content of their determinations— that's what makes them bullshitters. but in order to change these kyriarchal systems' configurations, we must understand what they truly do, why they are the way they are, what values they reproduce and reify, and how they could be otherwise.

[1] University of North Carolina at Charlotte

Introduction

One of the main works we'll use to help us think through "AI"'s power, political, and valuative relationships is Langdon Winner’s seminal 1980 article “Do Artifacts Have Politics?” Winner was one of the first to articulate a theory of politics and power in technoscientific artifacts which helped us to understand that by the politics of artifacts and systems, we don't simply mean the idea that the political space in which a piece of technology finds itself matters— that the values of the culture in which it's built matter. Those things do matter, but for Winner and many thinking with— as in alongside and after— him, when we say "artifacts have politics," we are also trying to get at an understanding that any built thing also has embedded in it— in the very fact and manner of its construction— some determination about what we think is good and bad in the world.

In “…Artifacts…,” Winner primarily showcases artifacts— built things, unique items— but he also discusses systems— interrelated webs of dependencies, networked concepts, and material conditions, including artifacts. Even in the late 20th century, it was increasingly difficult to distinguish artifacts from systems, but it is especially true today. Within the fields of science and technology studies and the philosophy of technology, understanding the interconnected interplay of artifacts that exist and the webs of systems in and of which they are made is a primary purpose. A dry-erase marker means little without something to write on; the whiteboard means nothing without a marker; the whiteboard's existence entails a particular kind of marker; and the dry-erase marker needs a particular kind of surface. I can write with this marker on other surfaces, like my desk or a piece of paper, but it will not write as well, and it will, in fact, eventually damage the marker. I can write on the whiteboard with other things, such as a sharpie, paint, or a permanent marker— but those things will damage the whiteboard. I can clean the whiteboard with other things than a specialized spray and a cloth— I could use hydraulic fluid, or acid, or I could just splash a whole bunch of water all over the place; and all of those things would work— to an extent. But they would not do the job that ***that*** cleaner was meant to do for ***that*** whiteboard written on with ***that*** marker.

Every artifact has its relation, and those relations are both necessitated by the materiality of the artifacts and the world and contingent on the human choices which have brought them into being. Neither “AI” qua “AI” nor the forms these tools and systems currently tend to take are inevitable; but in order to change the so-often kyriarchal, hegemonic, environmentally damaging, capitalist, white supremacist, ableist, misogynistic, and otherwise prejudicially biased configurations, we must understand what these system truly do, why they are the way they are, and how they could be otherwise.

Winner’s Artifacts

When Winner says that artifacts have politics, he is conveying a situation in which a built and physical thing has human values embedded in the very fact of its construction— that in every made thing exists some traces of what its creators think is good and bad in the world. To

discuss this, Winner uses the examples of tomato harvesters, signifying labor automation technology in general, nuclear power, and the bridges and parkways of Robert Moses. Throughout the mid-20th century, Robert Moses designed much of the iconic landscape of New York City as we know it today. In his comprehensive 1976 volume *The Power Broker*, biographer and historian Robert Caro describes Moses as a racist of multiple stripes and someone who also bore distaste for the lower classes.[2] Moses believed that certain types of people should not be allowed in certain parts of New York City, and he was explicit in communications with others about whom he thought of as worthy of using certain parts of "his" city— Moses did not want Blacks and Puerto Ricans near his most prized public projects. He was also extremely adept at wielding power.

Caro describes how Moses built relationships in which he had bureaucratic authority over things like who got what kinds of contracts to build, roads, bridges, and parks, or to lay out new building construction throughout New York. He was so good at understanding the law and crafting policy and legislation to suit his needs that he outmaneuvered Fiorello La Guardia, one of the most powerful and politically deft mayors in New York City history. Under the legislation that Moses put together, everything had to run through him. Robert Moses, very specifically, built many of the parkways and the bridges in NYC, and these parkways were intended to be parks that one could literally drive through. A trip on the parkway would be a scenic wooded drive through nature so that rather than just sitting around in Central Park or only driving through smog-filled highways, you would drive through the park, you would get to have a nice Sunday drive with your family in the beautiful scenery of New York City. To this end, Moses designed the entrances and overpasses of the parkways to be of a very specific size, such that trucks or other large commercial vehicles could not drive on them and spoil the scenery. But if trucks can't drive on the parkway, other truck-sized things can't drive there either. And so city busses could not drive on the parkway.[3]

Many New York residents didn't have personal cars, at this point in time; more specifically, many low-income residents did not have personal cars. Not coincidentally, many of these same low-income residents were of Black and Puerto Rican descent, and they, like most people without a personal conveyance, used the bus to get from place to place. When Robert Moses designed the parkways to be accessible only to cars, only people who could afford to ***own*** cars got to experience the parkway. Correspondingly, when Robert Moses made his bridges and overpasses too low for busses to pass through, and those bridges and overpasses connected the inner city to the beach, the main way to access the beach was by driving there. That, in turn, restricts who the beach is for. If the vast majority of the people using the bus "just so happened" to be the kind of people that Robert Moses deemed undesirable, and it "just so happened" that Moses designed roads and bridges that would manage to exclude just those very people, that,

[2] Caro, Robert. *The Power Broker: Robert Moses and the Fall of New York.* New York: Knopf, 1974.
[3] Winner, Langdon. "Do Artifacts Have Politics?" *Daedalus*, vol. 109, no. 1, 1980.

Winner argues, is not just a coincidence: it is Moses' politics coming through in the artifacts and systems he designed and built.

In the many years since *The Power Broker* and "Do Artifacts Have Politics?," people have looked at Winner's work and writing and reexamined the lenses through which historians have interpreted Robert Moses. Some critics have speculated that, "maybe the situation isn't just a cut and dry matter of saying Robert Moses was a racist who built his bridges so that the people he didn't like couldn't get there" even noting that there were other routes busses could take to get to Jones Beach, and other ways for people in the city to get onto the parkway.[4] Others have argued that Moses can be absolved of having made racist bridges because he likely did not do so intentionally.[5] Now, Winner does not argue that every single thing that is built has only one kind of political value to it; rather, his argument is, when we build things, we build our values into it. Consciously or unconsciously, we build our values, perspectives, beliefs, and politics into the things we create. Conversely, our values get shaped by the things we experience and engage with, by the artifacts and systems we interact with daily.

The internet, as a whole, is a massive series of systems— a system of systems. Everything from algorithms to user experience design elements to specific frameworks of architecture and artistic design all exist on the internet and work together to make our experience of the internet what it is. And some of those elements were created so deep into the past, so far away from our present experience, that we simply do not and cannot really know how they shape what we think and what we feel as a result of the systems with which we interact, let alone their constituent elements. Consider the origin of Facebook: a small, knocked-together project called "Facemash," created as a bespoke clone of "Hot or Not"— an online quiz app to rate pictures of women. It was launched in the Harvard intranet system in 2004, the product of a collaboration between Mark Zuckerberg and the Winklevoss twins to make a fun game where people could rate and rank the dates they had. And people ***loved*** it. Facesmash takes off amongst the Harvard dorms, and people start to share it and talk about it. It expanded, and people now use it for more than just rating each other. In much the same way, once the internet happened, email happened, and shortly after Facemash reached popularity, people wanted a way to message each other. Somewhere in here, Facemash becomes "The Facebook."

Prior to 2004, a "Facebook" was primarily known as a physical artifact. It was a book in which universities would keep their students' images (faces) and contact information. And so, in being rechristened "The Facebook," Zuckerberg and the Winklevosses sought to signify their creation of a digital extension of a known category of physical artifact. When people started messaging each other on the Facebook, they started talking about what they might do later on in the day, how they were going to hang out, dates that they might want to go on, or things that had

[4] Kessler, Glenn. "Robert Moses and the saga of the racist parkway bridges." *The Washington Post*. 2021.
[5] George Stevens. "Robert Caro's Moses: A Historian's Critique," in Joann P Krieg *Robert Moses: Single-Minded Genius.* New York: Heart of the Lakes Publishing, 1989.

nothing to do with dates; sharing thoughts or early memes— things like chain letters of the kind people would share in an email; all of the usual internet ephemera that happens whenever people get together and do Internet things. While it took time for the Facebook to become Meta as we know it today, it still took its start from the objectification and the rating of women— thus making what we now know about the effects of Meta's products not at all surprising.

Using the theoretical frames partially developed by Winner to understand the social impacts of technology, we can easily understand how a social media technology, which started as a means to objectify and rank women, went on to iterate into a version of itself that drastically harms everyone's mental health, and especially that of women and young girls.[6] We now have multiple studies, external ***and*** internal research, and even sworn congressional testimony, all of which attest to the harm done by the algorithmic content delivery mechanisms in systems like Facebook and Instagram, and now Tik Tok.[7] These internet cultures rely primarily on visual presentation, where comparison to and evaluation against others— objectification and ranking; "hot" or "not"— is a fundamental component. The values and politics of their origins are baked into these algorithmic systems, and they have no easily accessible brakes, making it hard to push back against those values and politics. The affordances of these online ecosystems mean that the easiest way to behave in them is to take the path of least resistance.

Broadly, the affordances of a system or an artifact are those design elements that indicate and influence its use. The affordances of a stool make it easier to sit in than use as a mode of conveyance; the affordances of a website's algorithms and interface make it easier to, e.g., share reposts and comments, than original content or links to other websites . However, a thing's affordances do not foreclose all other possible modes of use; instead those other modes may simply be harder. But a tool or system may also be what Robert Rosenberger terms "multistable,"[8] meaning such a technology has more than one valence of meaning and experience. Put simply, if we work against the affordances of systems and their attendant harms, we may, to an extent, push the system in a different direction and mitigate those harms, but if we merely let a system do what its design enables, then what it will do are the things that are easiest for it to do, which are what the values built into it will shape it toward. As digital tools and systems change the processes by which we interact with each other and influence the meanings of those interactions, we are thus faced with implications for the social process that shape the production of technology and, more generally, knowledge. Because the production and certification of knowledge is not value-free, it is a value-laden political process, one in which

[6] Protecting Kids Online: Testimony from a Facebook Whistleblower Hearing Before the Subcommittee on Consumer Protection, Product Safety, and Data Security of the Committee on Commerce, Science, and Transportation United States Senate, 117th Cong. 1 (2021) (testimony by Frances Haugen).
[7] Protecting Kids Online: Facebook, Instagram, and Mental Health Harm Hearing Before the Subcommittee on Consumer Protection, Product Safety, and Data Security of the committee on Commerce, Science, and Transportation United States Senate; 117th Cong. 1 (2021) (testimony by Antigone Davis).
[8] Rosenberger, Robert. "Multistability and the Agency of Mundane Artifacts: From Speed Bumps to Subway Benches." *Human Studies*. 37, 2014.

biases and perspectives take part. As Melvin Kranzberg famously put it, "Technology is neither good nor bad, nor is it neutral."[9]

A (Very) Brief Explanation of "AI"

"Artificial Intelligence" has hype cycles of waxing and waning interest, and all of these fervent conversations about what "AI" should or will be or do, how it will change the world and make things unrecognizable have happened before, and they very likely will happen again. These cycles are referred to as "springs" and "winters." The first "AI spring" happened in the 1950s and 60s, when the concept of "AI" was brought into public awareness for the first time— the idea of a machine that thinks. In the 1960s, a doctrinal rift formed within "AI" research, with one group interested in furthering the development of computation and cybernetics— feedback loops intended to replicate biological processes— and the other working to replicate human neural networks in electronic, or "artificial," form. Both camps, however, disregarded concerns about the impact specific kinds of bodies, emotions, or socialization might have on cognition; these men firmly believed the key to "AI" was to divorce any messy human social factors from the "purity" of abstracted function, rationality, and intellect.[10]

During this same time, "AI" scientists developed language models (LMs)— a method for determining the probable occurrence of words of particular lengths based on context cues such as their starting letter and the word preceding them. One of the earliest examples of public LMs was ELIZA, a program developed by Joseph Weizenbaum at MIT in 1964 and meant to act as a parody of open-ended psychotherapy. Though the program was only a few hundred lines of code, people told ELIZA their deepest secrets as if she were a human therapist. Next came natural language processing (NLP)— a combination of linguistics, computer science, artificial neural networking, and "AI" research concerned with getting computers to interpret, process, and communicate in "human-like" language. In the 2010s, Global Vectors for Word Representation (GloVe) and Word2Vec were among the most well-known NLP models, working by forming statistical correlations to map the relationships between words, allowing for an easy way to embed layers of associative meaning. In this way, LMs could represent the semantic connections between words like "woman" and "queen" or "drive" and "car." These early models use so-called machine learning ("ML") to encode their varied English language elements as data and then train the system to meet particular predictive targets and reinforce the associations between the embedded datapoints. Those associations are then mapped as vectors— or mathematical representations— of how strongly they're associated.

However, Word2Vec and GloVe had two very large flaws, the first being that their outputs very often contained obviously prejudicial biases. Again, LMs and NLP are, at base, automated word associations in which massive collections of words and phrases known as "language corpora" are turned into mathematical representations known as "tokens." These systems are

[9] Kranzberg, Melvin. "Technology and History: 'Kranzberg's Laws'," *Technology and Culture* 27. 1986.
[10] Wilson, Elizabeth A. *Affect and Artificial Intelligence*. Seattle: University of Washington Press, 2010.

then trained on the tokens to predict the ranges of potential associations between them. When you look at the vector space, the word woman doesn't just appear next to the word king or queen, but vice versa— the word "woman" also appears in the vector space very near to the word "secretary," as well as the words "nurse," and "teacher." And "woman" appears very distant from the words "president," "doctor," "CEO." This happens because the collections of natural language on which they were trained was itself packed full to the brim with error prejudicial bias.[11]

Language corpora have to be built, put together out of existing, human-generated, conversational data because the goal in all of this is the reproduction of "natural" language. But datasets for training modeling systems cost money. To this end, language sets such as the Enron Corpus represent treasure troves of large amounts of free or low-cost public domain natural language interactions; but they are also very often filled with human beings speaking in bigoted, immoral, or even just unconsciously biased. And this is because the Enron corpus is a collection of the more than 600,000 emails generated by 158 employees of the Enron Corporation in the years before the company collapsed, all of which entered into the public domain as the Enron Corporation was on federal trial for public corruption and defrauding their investors. The Enron Corpus contains people trading and rating pictures of women; slurs against people of Muslim background; stereotypical "jokes" about the sexual proclivities of Black and Asian peoples; and more. Due to this, we find that language modeling systems which use this training data replicate, iterate on, and exacerbate the built-in biases and perspectives of the data from which they are trained. And so we find what initially seems like surprisingly gendered bias, racial bias, or ableist bias within the system— and yet, it should not, in fact, be surprising. Because those systems are trained on data— on human interactions— which themselves contain those biases. This also results in outcomes such as automated résumé sorters rejecting the applications of women and certain minorities at higher rates than white men and, more recently, rejecting any résumé that mentions disability in any way.[12]

The second problem with Word2Vec and GloVe was that they couldn't map associations across larger collections of text; in fact, the number of associations they could make actually decreased as the amount of text increased. The solution to this problem wasn't simply adding more and more associations or tokens to parse; this problem space required a different way to parse the tokens themselves. And so we come to the transformer framework.

The "GPT" in ChatGPT stands for "Generative Pretrained Transformer," a phrase describing a system of interoperable algorithms used to weigh, arrange, sort, and return large associations of

[11] Caliskan, Aylin; Bryson, Joanna J.; Narayanan, Arvind. "Semantics Derived Automatically From Language Corpora Contain Human-Like Biases." *Science.* 2017.

[12] Kate Glazko, Yusuf Mohammed, Ben Kosa, Venkatesh Potluri, and Jennifer Mankoff. 2024. "Identifying and Improving Disability Bias in GPT-Based Resume Screening." In Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24). Association for Computing Machinery, New York, NY, USA,

language. The transformer process is the overarching operational algorithmic framework enabling everything from GPT text-based transformers for OpenAI, to Google's Gemini, to Midjourney, to deepfakes, and large language models (LLMs). Their training is about creating a greater fidelity of statistical correlation between text and text, text and image, text and video, or text and audio. In order to do any of this work, you first use the model in an untrained way, and that untrained result is then compared to the desired output. It is then retrained toward the desired output, and then that is put into practice to generate new results. It's a feedback system. Human beings then engage in a process known as reinforcement training. That reinforcement process oftentimes requires a large group of humans to teach these systems whether they've gotten it right or whether they've gotten it wrong and then nudge them closer and closer towards getting it right. That's what reinforcement means in the space, but the language corpora that are used for training still have to come from somewhere. For LLMs, the language corpora used are the internet as a whole.

As the name suggests, LLMs are larger versions of LMs; in fact, the data that LLMs use can be in the billions, or now even trillions, of tokens large. Ideally, these tokens have to be labeled in the "right" way, but what does "right" even mean in this context, and who are the people doing the training which enables all of this? Because this is how biases get embedded within GPT and LLM systems: The data on which they are trained comes from sources with those biases encoded. Over the past several years, this has become a standard understanding throughout LLM systems: biased training data gets biased results. But many people don't know that the biased training data that we're dealing with is, in many cases, decades old, as the old, prejudicially biased corpora get scraped and re-scraped every time the outputs from a system trained on them is scraped. And so these newer LLM and GPT-type "AI" systems retain the same problems of prejudicial bias as we find in GloVe and Word2Vec— but now those problems have iterated faster and become more interwoven with every aspect of our lives.

While the above is not a full history of "AI" or of all of "AI's" "springs" and "winters," it is meant to emphasize that the social values present in each turn have emphasized the automation of speed, efficiency, and the mimicking of human rationality. These values have remained present, even when other, previously disregarded mental and social process elements have been engaged and brought into the "AI" space. It could be argued that we are in the middle of a protracted and segmented "AI" spring, in which both the definition of "AI" and what gets defined ***as*** "AI" change more rapidly and within an ever-shorter timeframe. However, it could be said that the underlying framework characterizing all of these technologies is the automated algorithmic system, a term simultaneously ubiquitous and (often intentionally) opaque. An algorithm is merely a set of instructions— standardized operations within which particular weights and measures can be adjusted in order to change the outcome of a process. But when you alter an algorithm, you must alter every element of the whole— the framing context— to ensure the final product still turns out correctly. And if this is not done intentionally and carefully, then what comes out the other side may be very messy indeed.

Ultimately, the algorithms that enact LLMs and GPT-type systems do not care about giving you answers which correspond to the facts of the world. What these systems do is statistically model the most-likely-to-be-acceptable result. LLM and GPT-type "AI" are, quite simply, trying to tell you a story that will jibe with your preconceptions and biases, based upon the inputs you've given the system; this is not the same as "providing facts" or "telling you the truth." Conversely, when the system generates something that does not correspond with reality— a process anthropomorphically described as "hallucinating"— it isn't "lying" to you or "making a mistake;" in fact it's doing the same thing as it did in providing you an answer that ***did*** corroborate facts about the world. And in epistemology and philosophy of language, we have a very specific word for a speech act that is not at all concerned with falsity or truth.

When we talk about the truth content of speech acts, we are talking about something very specific. If I seek to tell you the truth as it is, I am seeking to tell you what I believe corroborates with the facts of the world as I know (or believe I know) them. If I am mistaken about my beliefs, then I haven't lied to you; I either had an error in my own understanding or in my relation of my understanding to you. And so in order to lie to you, I have to seek to deceive you, and in order to seek to deceive you, I have to tell you something which ***I believe*** runs counter to the facts. To do that, I have to know— or believe I know— what those facts are. But what if I don't care about telling you facts or truth, and I don't care if you're deceived by what I have to say? In that circumstance, what I'm saying isn't the truth, and it isn't a lie— what I'm saying is bullshit.

The Politics Which Frame "AI"

In Harry Frankfurt's 2005 monograph *On Bullshit* he describes the process of "bullshitting" as telling a story and not caring whether it conforms to the facts of the matter (is true), or whether it does not (is false).[13] And for Frankfurt, this is the most potentially harmful kind of speech act because there is no goal for the bullshitter other than telling a good story. And if you're an irascible musician spinning yarns about trains and circus performers, then telling a good story is part of the job. But if you're trying to have a conversation with other people about shared public values, or trying to design a system in which people can make decisions about their daily lives, then a foundation of bullshit is a very dangerous thing indeed. LLM and GPT systems are always making things up, but sometimes what they make up happens to correspond more closely to the facts of the world, and sometimes it doesn't; either way, the "AI" doesn't care. "AI" systems do not care about facts; more to the point, they ***cannot***, so far as they are currently constructed, care about facts. These "AI" systems only care— insofar as they are able "care" about anything— about generating a response that is statistically the most likely to be accepted by the prompter based on the correlational values of their input tokens. This, in turn,

[13] Frankfurt, Harry. *On Bullshit.* New Jersey: Princeton University Press. 2005.

means that there is no way to remove "hallucinations" from these systems— because the "hallucinations" fundamentally are what transformer models does. They are bullshit engines.[14]

And Google's Sundar Pichai, Apple's Tim Cook, and OpenAI's Sam Altman have all admitted this fact; these CEOs have fully admitted that there is no way to eliminate "hallucinations" in practice and that their systems cannot help but bullshit you. You can mitigate the bullshitting, and you can minimize it, but you will never stop an LLM or GPT-type system from, at some point in its operation, fabricating wholecloth something that does not necessarily conform to reality because it figures you might like that answer best. And that "AI" system will present that fabrication to you with the same confidence and certainty as it would any fact that ***does*** conform to reality. This has major implications for the places in which generative "AI" is already interwoven with our lives and understanding this requires first reckoning with the concepts via which we discuss them. As we've discussed, even the term "AI" contains assumptions about the nature of intelligence (or consciousness), about what parts of "human-like" cognitive processes can be replicated in machine processes— but the term also contains and enacts assumptions who or what can have intelligence at all.[15]

In 2022, the Georgetown Center for Privacy and Technology announced they would no longer refer to systems as "artificial intelligence," "AI," or "ML," arguing that these terms obscure and often outright mislead the lay public more than they illuminate. I agree, and also argue that this terminological obfuscation facilitates the hype-laden corporate marketing of those trying to develop and sell "AI" systems and also enables circumstances wherein even supposed critics nonetheless ***uncritically*** reiterate the overblown claims of the technologies' creators. This comingled hype thus allows "AI" proponents to claim that they've built something a) extremely powerful and b) over which they have no responsibility or control. But the basic fact of the matter is that we don't know what "intelligence" is, let alone what we mean by "AI," and so the term has merely become a marketing tool rather than pointing to anything real or concrete. "AI" is a moving cipher seeking a referent.

Simultaneously, this willful occultation of what happens inside these systems further confounds and oppresses the marginalized human beings most often subjected to them. This, too, is why all instances and variants of "AI" are here present in scare quotes or otherwise qualified: To better critically and productively trouble both this occultation and the assumed notions of both "artificiality" and "intelligence." Elsewhere, I also use this to highlight and directly challenge the unwillingness of so many to confront the religious and magical valences of their own thinking about "AI." For instance, when "AI" systems determine what targets to hit in

[14] Williams, Damien P. "On Bullshit Engines: The Socioethical and Epistemic Status of GPTs and other AI." Lecture in the Forum in Ethics, Law, and Society Lecture Series, hosted by the Sonoma State University Philosophy Department's Center for Ethics, Law and Society. 2023

[15] These assumptions have no agreed-upon criteria, and a not-uncontroversial history of application; for more on this see Williams, Damien Patrick. *Belief, Values, Bias, and Agency: Development of and Entanglement with "Artificial Intelligence"*. Diss. Virginia Polytechnic Institute and State University, 2022.

military conflicts, the technology is often given religiously valanced names. In Gaza and Lebanon, the Israeli military uses a system known as "the Gospel."[16] When we use religious language to discuss technology, it taps into a conceptual paradigm that signifies certainty, even for those who might consider themselves atheists. It provides a framework of guarantee and expertise that says, "This is trustworthy, this is correct, this is infallible. It is the gospel."[17]

This happens in much the same way that technologies that rely on code and math garner veneers of "objectivity," which in turn gives people a way to wash their hands of any responsibility for what those very human technologies do: the repeated habit of and inclination toward this process is what Meredith Broussard terms "technochauvinism."[18] This happens not only with GPTs but with "AI," algorithms, and code more broadly as well. "It is math, it is code. It can't have biases in it. Can it?" And yet have long known that it can— that technological systems can have human perspectives embedded in them.[19] And while these systems are increasingly applied in theater of war where a decision is a matter of life and death, as we've discussed above, it can also be about life and death at home.

Neurodivergent individuals, non-native English speakers, women, and racially minoritized individuals who tend to "mask" or seek to "pass," in their communications, be they spoken or written. From the outside, this process often appears like someone over-thinking or being overly formal or precious— even "stilted" or "robotic"— with their word choices. In written form, those phrasing patterns get flagged as "anomalous" by GPT checkers, thus designated as less likely to be "human." This is done not only by automated systems but also by other humans. As we'll discuss below, many people begin to believe that "we can always tell" when utterances are produced by "AI" systems— even when they're not. When we rely on "AI"— when we certify "AI" systems as arbiters of truth, as determiners of fact— what we are doing is giving these systems the power to make judgments for us and safeguard knowledge. In so doing, we are imputing heuristic capabilities to generative "AI" systems that they do not possess. And when we begin as a culture and as a society to rely upon them, we begin to reflexively seek to enact the kinds of judgments that we believe an "infallible" system might make. Thus, we begin to believe that we can see the "markers of AI" where they do not exist.

Telling people who they are rather than heeding their lived experiences is a way of enacting power and control over how people perceive themselves and their world. In this same way, institutions and organizations further concentrate their influence and power via circumventing or

[16] Mariam Barghouti, "The Gospel: Israel Turns to a New AI System in the Gaza War." Al Jazeera. 2023.
[17] For more detailed discussions of the religious values and perspectives embedded in "AI," see Williams, Damien Patrick. "Any sufficiently transparent magic…" *American Religion* 5.1 (2023); and Williams, Damien Patrick. "Religious Belief and Practice in the Age of 'AI,'" hosted by the Emory University Candler School of Theology. 2024.
[18] Meredith Broussard, *More Than a Glitch: Confronting Race, Gender, and Ability Bias in Tech*. Cambridge, MA: MIT Press. 2023.
[19] Porter, Theodore M. *Trust In Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, N.J.: Princeton University Press. 1995.

manufacturing the consent of their users and those made subject to their systems. When "AI" companies suggest the need for nearly fantastical advances in nuclear power or the retasking of existing sources as a way to redirect attention away from their astronomical water use and carbon footprints, it is in an attempt to control the conversation about the human cost of "AI"'s creation, deployment, and operations.[20] When "AI" companies claim to make major advancements in the fields of machine vision or reinforcement learning, but have actually just offloaded that work, some of which has some of the most dangerous effects on mental health of any job on today, to exploited workers in the majority world, those are also attempts by those corporations to control the conversation.[21]

And this is why "AI" CEOs such as Sundar Pichai or Sam Altman, and corporations ranging from Facebook to Zoom, and now even academic publishers like Wiley, Oxford, and Taylor and Francis try to claim the right to scrape anything they want from the internet in general and our creations in particular without our consent and use it for "AI" training data.[22] When these people and groups tell you "we can't do our job if we can't use copyrighted material," they are both lying to you and trying to get you to accept a particular way of thinking and operating within the framework of these technologies. When Meta partners with Ray Ban to make facial recognition-enabled smart glasses that they put on the market and then place that product in every single commercial break of every television or online streaming show you watch, they are trying to lock you into a way of thinking about these technologies. However, these corporations' and organizations' preferred frameworks are, in fact, not necessarily the only options, though they want you to believe that they are. It is not inevitable that these tools, that these systems, are built in the ways that they are built, but believing in their inevitability allows for the people who make the systems to tell us how it then has to be from that point forward, if they're inevitable, then we have to build them. And if "we" build them, then our competitors can't build them— or at least "we" have the most influence over how they're built and what tools and capabilities are available within these systems. "If American companies build AI systems," the argument runs, "then the US controls the market, not China;" or "If we make these chips in Ohio, we don't have to worry about what hidden firmware may be hustled into the system."

All of this is a process of paradigmatic capture. When you capture the paradigm— the way of thinking about, researching, building, deploying, and using something— from top to bottom, you get to then determine how everyone else engages with it from that point forward. However, there are a whole host of people and groups working to regulate and alter how we build "AI" systems,

[20] Verma, Pranshu, and Shelly Tan. "A Bottle of Water Per Email: The Hidden Environmental Costs of Using AI Chatbots." *The Washington Post*. 2024; "Microsoft deal propels Three Mile Island restart, with key permits still needed," Reuters. 2024.
[21] Perrigo, Billy. "Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic." Time.com. 2023; Zeff, Maxwell. "Amazon Ditches 'Just Walk Out' Checkouts at Its Grocery Stores." Gizmodo. 2024; Birhane, Abeba, Sanghyun Han, Vishnu Boddeti, and Sasha Luccioni. "Into the LAION's Den: Investigating Hate in Multimodal Datasets." Advances in Neural Information Processing Systems 36 (2024).
[22] Kwon, Diana. "Publishers are selling papers to train AIs-and making millions of dollars." *Nature* 636.8043 (2024)

including executive orders and laws from the United States and new "AI" acts from places like the EU and China. China, for instance, just decreed that there must be a way to digitally watermark everything that comes out of an "AI" system or that said system cannot be used in China.[23] Digital watermarking and other means of provenancially verifying "AI" outputs was proposed as early as 2022, floated as a means to know what is and is not "AI," and at that time (and even more recently) "AI" companies said it would be completely infeasible.[24] Since then, it has been revealed that not only ***can*** it be done, but since nearly their inception OpenAI has had a system to steganographically embed a marker to identifies text generated by their "AI" with 99.9% accuracy, absent large-scale editing on the system's product by a human.[25] This can be done via simple word correlation choices, which can be weighted within the model, and the fact that they are not widely enacted until companies are forced to do so indicates something about those companies motives.

The Politics Enacted Through "AI"

We are now beginning to see generative "AI" in all the places where algorithmic systems have long been interwoven into our daily lives. We are continually confronted with headlines highlighting some new horrific outcome of "AI," yet people and institutions seem to still be rushing full steam ahead to use it. For instance, during this election cycle, we have seen the use of generative "AI" deployed from and intended to influence political perspectives, from "AI" generated Taylor Swift and her fans being used to claim endorsements for multiple candidates to "AI" generate audio of candidates' voices meant to convince people not to vote. Left unaddressed, the danger that generative "AI" poses to the civic processes is truly massive— but not wholly unprecedented.

In addition to the already-mentioned sexist and ableist prejudicial bias displayed by résumé sorters, researchers have also found that systems entangled with generative "AI" bullshit engines recommend white male job candidates almost 100% of the time.[26] "AI" image generators or "enhancers" consistently create racist and sexist images.[27] Generative "AI" systems which are being used to guide people through getting disability benefits replicate the same kinds of ableist biases that we've seen in previous older versions of the system. Racist and sexist biases show up

[23] Yang, Zeyi. "China's Plan to Make AI Watermarks Happen." 2024. *Wired*; China Law Translate. "Measures for Labeling of AI-Generated Synthetic Content (draft for solicitation of comments)." 2024.
[24] Francis, Matthew R. "The Ethics of Artificial Intelligence-Generated Art" SIAM News. 2022; Hague, Douglas C., et al. Comment #NTIA-2023-0005-1215, National Telecommunications and Information Administration Request for Comment on AI Accountability. 2023. https://www.regulations.gov/comment/NTIA-2023-0005-1215.
[25] Seetharaman, D., and M. Barnum. "There's a Tool to Catch Students Cheating with ChatGPT. OpenAI Hasn't Released It." *The Wall Street Journal* (2024).
[26] Stifler, Lisa. "AI Overwhelmingly Prefers White and Male Job Candidates in New Test of Resume-Screening Bias." GeekWire. 2024.; Wilson, Kyra, and Aylin Caliskan. "Gender, race, and intersectional bias in resume screening via language model retrieval." *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. Vol. 7. 2024.
[27] Ananya. "AI Image Generators Often Give Racist and Sexist Results: Can They Be Fixed?" *Nature*. 2024 Mar;627(8005); Elizabeth Laraki, *Tweet*. http://web.archive.org/web/20240000000000*/https://x.com/elizlaraki/status/1846252781851890026;

in regards "AI" systems' recommendations for who should get which kinds of health care, and even how an "AI" summarizes someone's healthcare interactions.[28] Because facial recognition systems don't see dark skin tones well, and the "AI" systems that are used to sort through surveillance videos are trained on and operating out of repositories of data rife with misidentified Black people, Black people comprise the vast majority of misidentifications, arrests, and erroneous prosecutions by police of people based on "AI"-backed systems of facial recognition.[29]

Combined with generative "AI" policing integrations such as police body cameras with generative text capabilities for writing reports or making predictive frames for resource distribution.[30] The datasets for training such a system will consist of the tagged and parsed collection of often handwritten, post-hoc recollections police reports, reports which are themselves written based on the responding officers' notoriously and prejudicially biased memories.[31] So, while some researchers suggest the use of "AI" tools as a means to review bodycams and check ***against*** officers' memories— possibly even to help determine what patterns and markers might be present in interactions where in which police are more likely to harass people— it is baffling to hear anyone defend any "AI" as "objective" or "impartial."[32] It is especially concerning when that assertion receives no pushback from the interviewing journalist.[33]

Generative "AI" integrations given over 6,000 sample loan applications based on data from the 2022 Home Mortgage Disclosure Act recommended denials for Black applicants more often than statistically identical white counterparts, recommended higher interest rates for Black applicants, and labeled Black and Hispanic borrowers as "riskier," than white counterparts, and results showed that white applicants with low credit were approved 95% of the time, with Black applicants approved less than 80%.[34] These outcomes can be traced to the prior orders of both algorithmic digital and original analog redlining, so the recurrence of it in new technologies should not come as a surprise.

---

[28] Koenecke, Allison, Anna Seo Gyeong Choi, Katelyn X. Mei, Hilke Schellmann, and Mona Sloane. "Careless Whisper: Speech-to-Text Hallucination Harms." In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*. 2024.

[29] Sanford, Alyxaundria. "Artificial Intelligence Is Putting Innocent People at Risk of Being Incarcerated." The Innocence Project. 2024.

[30]Cox, Joseph. "Here Is What Axon's Bodycam Report Writing AI Looks Like." *404 Media*.2024.

[31] McIlwain, Chalton. *Black Software: The Internet and Racial Justice, from the AfroNet to Black Lives Matter. New York: Oxford University Press.* 2019

[32] Kaste, Martin. "Human Reviewers Can't Keep Up with Police Bodycam Videos. AI Now Gets the Job." NPR. 2024.

[33] Ibid.

[34] Gross, Paige. "As AI takes the helm of decision making, signs of perpetuating historic biases emerge." 2024. Utah News Dispatch; Bowen III, Donald E. et al., "Measuring and Mitigating Racial Disparities in Large Language Model Mortgage Underwriting." 2024. Available at SSRN 4812158.

At least equally alarmingly, there's recently been a raft of "AI" generated mushroom foraging guides available for sale online[39] — sometimes wholly generated images based on what the system has been trained a mushroom is statistically supposed to look like, while other guides retrieve existing images and then generate text to describe them by means described above. Shortly after they integrated image search capabilities, I had Google Bard (now Gemini) list out ten wild mushroom species to forage, and recipes to go with them; it correctly identified some, but the ones it misidentified had errors ranging from strange preparation suggestions to outright describing toxic species as edible— complete with tasting notes. Now, when someone asks for a bedtime story about a geese in the woods who had a delicious dinner, a statistically-more-likely-to-be-accepted response is fine; but when a real person is feeding their family based on a bullshit engine's advice, that is a different story. Because a fun thing about mushrooms is that many of them appear very similar, and sometimes what you think is a tasty dinner ingredient is actually a poison that will liquefy your liver; you will die, and it will hurt the entire time that you are dying.[40]

In the realm of academia, we have seen increasing evidence that "AI detectors"— systems that themselves use generative "AI" to purportedly identify text created ***by*** generative "AI"— contain prejudicial bias against both non-native English speakers and neurodivergent students, a thus urging us to further consider differences in how expectations and perceptions of gender and race can affect people's communication.[42] Some people's lived experiences are least likely to be adequately accounted for in either the training data or the weighting architectures of GPTs or other "AI" tools, the implications of which are well exemplified by incidents related by Rua M. Williams: One in which a non-native English-speaking student used ChatGPT to rewrite their work to "sound more white," and another in which Williams was falsely accused of using "AI" for their own writing was. [43] Both of these instances exemplify the normative pressures within both academia and "AI" research.[44]

And yet, even though we recognize all of their varied harms, these systems are still in use, which makes it seem like it is more valuable to have a system that does not work well and which harms marginalized and minoritized individuals and groups than it is to do the work necessary to fix these systems. Again: One can alter how the "AI" generates text or images, thus producing

[39] Trybuch, Elan. "AI generated mushroom books – life and death." New York Mycological Society. 2023.
[40] Erden, A, et al. "Acute Liver Failure Caused by Mushroom Poisoning: A Case Report and Review Of The Literature." Int Med Case Rep J. 2013 Nov 22;6.
[42] Williams, Damien Patrick. "Bias Optimizers." *American Scientist* 111.4 (2023); Liang, Weixin, et al, "GPT Detectors Are Biased Against Non-Native English Writers," *Patterns*, Volume 4, Issue 7, 2023; Davalos, Jackie and Leon Yin. "AI Detectors Falsely Accuse Students of Cheating—With Big Consequences." 2024. *Bloomberg*.
[43] Williams, Damien Patrick. "Disabling 'AI': Biases and Values Embedded in 'Artificial Intelligence.'" *Handbook On the Ethics of Artificial Intelligence*. Edward Elgar Publishing, 2024; Williams, Rua M. Fractal Echo. https://kolektiva.social/@FractalEcho/111949020988947948, https://kolektiva.social/@FractalEcho/110740599615481285.
[44] Maiberg, Emanuel. "Scientific Journals Are Publishing Papers with AI-Generated Text." *404 Media*. 2024; Williams, Damien P. "Scholars are Failing the GPT Review Process." *Historical Studies in the Natural Sciences* 54.5 (2024)

outputs identifiable by reference to the generating system— this is inherent to how LLM and transformer systems function. It may take extra time and require a bit more effort from, thus costing their parent companies more— and these are reasons why these kinds of processes are said to be "infeasible" or "impossible," right up until the moment we learn they have been done. Thus why we have to be on guard for that moment when "AI" companies shift from saying safeguards are "impossible" to complaining about their difficulty and expense. "AI" companies, like most other corporations, absolutely could protect their users' rights, dignity, autonomy, and the value of their creativity and labour— they just don't think the users are worth it.

Clearly, whether we're talking about GPTs and LLMs or just "machine learning" and algorithms, "AI" systems definitely have politics. The problem is not and never has been that what "AI" is and does is "out of step" or "unaligned" with human values, but rather that the values with which "AI" is aligned are a) very, very human, b) woefully under-examined, and often c) terrible. Those values embedded here tend to be racism, sexism, ableism, efficiency, power-over, speed, control, white supremacy, eugenics, hegemony, kyriarchy; what they don't tend be are prosocial consensus reality, intersubjective truth, or facts. Because obfuscation, misinformation, and disinformation are valuative positions— and they are also mechanisms of control.

Epistemic Crisis as Power

When we think about how we construct systems of knowledge, we are often concerned with ideas of objectivity and the determination of fact; but we can also consider intersubjective knowledge and consensus reality—making a shared understanding of our world together.[45] Both Lorraine Daston and Peter Gallison's 2007 book and Jonathan Grunert's 2019 dissertation discuss the history of how we come to think about what is and is not objective truth.[46] When we think about an anatomical drawing or nature drawing of a leaf or a taxidermy representation of an animal, what is the goal of that representation? Is it a perfect re-inscription of one particular animal, one particular plant? Or is it a way of trying to capture something about all forms of that animal, that plant? This is a debate that's been alive within image-making for centuries and must now be extended to what we are seeking to represent when we use generative "AI." Are we seeking to represent the factual truth of the matter, or are we seeking to represent, in a broad way, the generally accepted perceptions of things as they have tended to be recognized by people?

Because what we do every time that we use a generative "AI" system is point to the past: Every single "AI" system that exists uses data ***from*** the past. It is trained on things that have

[45] Williams, Damien Patrick. "Constructing Situated and Social Knowledge: Ethical, Sociological, and Phenomenological Factors in Technological Design," in *Engineering and Philosophy: Reimagining Technology and Social Progress*, edited by Guru Madhavan, Zachary Pirtle, and David Tomblin; Springer, Cham, 2021.
[46] Daston, Lorraine, and Peter Galison. *Objectivity*. New York: Zone Books, 2007; Grunert, Jonathan D. "Strict Fidelity to Nature: Scientific Taxidermy, US Natural History Museums, and Craft Consensus, 1880s to 1930s." (2019).

been, and yet it is used to teach us something or to try to teach us something, about what might be— a task for which they are fundamentally not suited. If you want to look at the way that people thought about race, gender, and sexuality up until even 2021, then "AI" is the system for you. But if you want something that can tell you how things might be tomorrow, or how things ought to be 10 years from now, that is not what these tools are for. Expectation and PR hype shape what the public thinks "AI" tools are, can be, what they should do, and how people use them. Reflexively, people's expectations and use then shape how the researchers, designers, and companies tell the story of how we should understand "AI". Here, we can think about the lessons of the social shaping of technology or social constructivism, which say that though we embed our values in technology, we cannot simply think that reshaping technology will completely reframe our values.

The problems of technochauvinism mean that we are more likely to trust as verifiable that evidence which technology provides. When Google changed the online search process, people began to treat positions in the ranked results as markers of trustworthiness; however, researchers like Safiya Noble and others have shown that trust and veracity had little to do with what we see in our searches.[48] To be clear, the certification of knowledge or the possibility of anything like an "objective truth" is by no means uncontested. Theorists in STS, epistemology, and philosophy of technology have demonstrated that the situatedness of time, place, and culture, from laboratories to peer review, can all have major impacts on what we know and how we know it.[49] But many people still seek certainty and treat technologically-enacted search results as a means to it, and they are thus vulnerable to over-trusting mathematically and technologically enabled tools— integrating generative "AI" into search makes all of this worse. We cannot Technofix our way out of values problems; and when we try, weird stuff happens. When Google started pushing their "AI Overviews" into search, there were immediate problems with the results, ranging from the absurd to the downright dangerous.[50] The pushback online was swift, and yet google changed very little about the process other than the things it wouldn't discuss. At the time of writing, the "AI" overviews are still in place, enacting terrible values through their operation.[51] When Google reframed their Bard LLM system as "Gemini" and started integrating audio, video, and image generation, somebody who was trying to prove a point about "wokeness"[52] prompted the system for an image of "racially diverse Nazis."[53] And so, Google's Gemini system generated a picture of racially diverse Nazis. Because that is what it was asked to

[48] Noble, Safiya U. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press, 2018.
[49] Cf. Madeleine Akrich, Bruno Latour, Steve Woolgar, Karin Knorr Cetina; Williams, D. P. "Constructing Situated and Social Knowledge."
[50] Cole, Samantha. "Google and Amazon AI Say Hitler's Mein Kampf Is 'a True Work of Art'." 404Media. Mar 6, 2025.
[51] Gilbert, David. "Google, Microsoft, and Perplexity Are Promoting Scientific Racism in Search Results." *Wired*. 2024.
[52] A term originating in Black American communities to indicate savviness about structural racism, but which has since been co-opted to indicate a kind of performative and exaggerated social justice mindset.
[53] Gilliard, Chris. "The Deeper Problem With Google's Racially Diverse Nazis." The Atlantic. 2024.

do, and that is all these systems do: Try to provide you for what you asked for. If people place their trust in "AI" generated images as being just the same as facts, we risk shaking or even outright corrupting the project of shared knowledge and consensus reality and corrupting the processes for certifying knowledge. They are not, on their own, a way to create a more equitable or more just society, and believing that they are can result in deeply offensive results that do more harm than good.

If an image is presented with a story as though the image ***is*** that story, exactly as it happened, and that image turns out to have no basis in reality, then we should question the story itself; that is, if the supplemental evidence used to certify a narrative is found to be false, then we should not automatically treat the story as true. Otherwise, the veracity of every conjunction of statement and evidence is assessed only by our feelings about them— by the vibes. And if we ***want*** everything just to be determined by vibes— by confirmation of our preexisting biases within bespoke realities— then that's okay, but we ought not try to call this a process of figuring out "truth" or "facts" through "evidence." We should just admit we like how an image, story, or narrative assertion makes us feel, no matter how spurious. Without at least that level of honesty, all we're doing is bullshitting each other at a speed and scale heretofore unheard of, and frankly, that will just keep degrading our ability to trust and communicate with other people. When generated images get called out as "AI" and members of the public respond with, "ok yes the image is probably A.I. generated ***it doesn't mean the story isn't real***" (emphasis added), then it can credibly be claimed that we are in a moment of cultural epistemic crisis.

Multiple papers in the past two years demonstrate that over-reliance on supposedly objective "AI" tools diminishes critical thinking capacity and prevents students from building the kinds of foundational skills which allow them to learn more complex concepts, adapt to novel situations, and grow into experts.[55] That lack of expertise and capacity has a direct impact on people's ability to discern facts, produce knowledge, and even participate in civic/public life. The diminishment of critical thinking skills makes people more susceptible to the abovementioned dangers of propaganda and other forms of dis- and misinformation. This further erodes trust in shared knowledge, consensus reality, and even expertise, thus exacerbating and worsening the next turn on the cycle when it starts all over again. And all of this creates the very conditions by which authoritarians seek to enrich themselves and cement their control: by obfuscating their intentions, offloading responsibility onto the systems, and undercutting the mechanisms which can empower the populace to understand and oppose dictators, theocrats, fascists, and kleptocrats.

[55] Gonsalves, C. "Generative AI's Impact on Critical Thinking: Revisiting Bloom's Taxonomy." *Journal of Marketing Education*. (2024).; Bastani, Hamsa, et al. "Generative ai can harm learning." *Available at SSRN* 4895486 (2024). Lee, Hao-Ping Hank, et al. "The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers." Microsoft. (2025)

Our reactions to new technologies are often slow, because the ways we think about them are uncertain. They are new, unfamiliar, and so we don't necessarily have a framework in which we can fit them; but many times, this slowness is intentional. How we think about these things are directed by whoever gets to have a conversation about them, and the people who drive these conversations, in turn, get to shape the cultural narrative about the thing. And so, given that we cannot but have values embedded in our technology, we have to take a close look at the values we want embedded in our technology to determine whose values and which values we want embedded in our technology. We will not be able to have an unbiased "AI." We will not be able to have a completely value free, value neutral technological system. It is not possible. We are human beings, and everything we touch, we put our perspectives into. Our artifacts and our systems have politics, and our other values too. So what values do we ***want*** to have in the things that we create?

None of this has to be this way. There are so many points when the prevailing paradigm and general milieu of the "AI" project could have been changed; we didn't have to be here. And we don't have to stay here. However, this will not be solved by using the same frameworks of consent violation, theft, and manipulation in a "public" or "open source model," but rather by reframing to a model that respects the dignity, autonomy, rights, and freely given and meaningfully informed affirmative consent of those whose information gets used as training data, and on whose populations the systems are deployed. The US Federal Communications Commission moved very quickly to make illegal the use of "AI" generated audio in campaign ads. The "Joe Biden" robocall went out January 9, 2024; the FCC ruled in February; and by late September 2024, the person who created the ad was fined $6 million .[56] Taylor Swift specifically noted that part of her reason for publicly endorsing Vice President Kamala Harris was that Swift felt her own autonomy, rights, and consent were violated by "AI" claiming to speak for her.[57] And after public outcry, Zoom had to change their plans, as did Microsoft for their "Co-Pilot" rollout.

Change can be effected when the public is educated and mobilized on these issues at multiple levels. And so we need multiple voices working together to try expand not just the technical education about, but our understandings of the social and human implications of "AI" itself and everything that falls under that very nebulous, empty cipher of a header.

These systems are being widely implemented, and those willing to sow doubt and use confusion to control others will profit more and faster than those who hold to consensus, prosocial norms.[58] In the run-up to university finals season for Spring 2025 OpenAI began

[56] Shepardson, David. "Consultant fined $6 million for using AI to fake Biden's voice in robocalls." Reuters. 2024.
[57] Torek, Belle. "Taylor Swift, Kamala Harris, AI Falsehoods and the Limits of Counterspeech." 2024. Tech Policy Press.
[58] Chesney, Bobby, and Danielle Citron. 2019. "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security." California Law Review 107 (6); Schiff, Kaylyn J, Daniel S Schiff, and Natalia Bueno. 2023. "The Liar's Dividend: Can Politicians Claim Misinformation to Evade Accountability?" OSF.

offering two free months of ChatGPT, marketed specifically to students; Donald Trump signed an executive order mandating the use of “AI” in K-12 education; and that same administration began feeding reams of sensitive personal data into “AI” systems, even as they cut funding in technoscientific research.[59] Norms like diversity, equity, and inclusion, consensus reality, and the processes of shared knowledge-making are under attack, precisely when we could benefit from them. We need to be but are being expressly prevented from considering the situated knowledge and intersectionality of who gets to shape everything from the political sphere to the technical training of the people who build our technologies. Many public policy, research, and grassroots organizing groups are working on ideas for new technologies and new ways to build these technologies, so how do we bridge the different needs and perspectives of the people doing this work?

Suppose we want technological systems that do not depend on predatory, extractive, kyriarchal logics which don't replicate biases against gender, disability, race, or other aspects of identity— which don’t first hustle those harms in under the guise of the next big thing, and then disseminate them to different places. To undo and prevent these harms, we must work very hard to educate, our students, our communities, and ourselves about how these systems work, what they can and cannot do, what impacts they have, and how we might work together to build something more overarchingly just in their place.

[59] Lee, Cydney. “In First Big Marketing Push, OpenAI Makes ChatGPT Free for College Students.” AdWeek. 2025; Trump, Donald J. "Executive Order Advancing Artificial Intelligence Education for American Youth,” April 23, 2025. Larkin, CJ. “100 Days of DOGE: Assessing Its Use of Data and AI to Reshape Government.” TechPolicy.Press. 2025; Garisto, Dan "Exclusive: NSF stops awarding new grants and funding existing ones.” Nature. 2025.