“SORRY IN ADVANCE!”

Rapid Rush to Deploy Generative A.I. Risks a Wide Array of Automated Harms

By Rick Claypool and Cheyenne Hunt

April 18, 2023

PUBLIC CITIZEN

ACKNOWLEDGMENTS

This report was written by Rick Claypool, a research director in Public Citizen's president's office, and Cheyenne Hunt, Big Tech accountability advocate in Public Citizen's Congress Watch division.

Thank you to all who provided feedback and editorial contributions, including Dr. Sarah Myers West, managing director of AI Now; Adam Connor, vice president of technology policy for the Center for American Progress; Robert Weissman, president of Public Citizen; Lisa Gilbert, executive vice president of Public Citizen, Alan Zibel, a research director in Public Citizen's president's office; Paul Alan Levy, an attorney with Public Citizen Litigation Group; and David Rosen, a communications officer for Public Citizen.

ABOUT PUBLIC CITIZEN

Public Citizen is a national non-profit organization with more than 500,000 members and supporters. We represent consumer interests through lobbying, litigation, administrative advocacy, research, and public education on a broad range of issues including consumer rights in the marketplace, product safety, financial regulation, worker safety, safe and affordable health care, campaign finance reform and government ethics, fair trade, climate change, and corporate and government accountability.

Contact Public Citizen

Main Office
1600 20th St. NW
Washington, DC 20009

Phone: (202) 588-1000

Capitol Hill
215 Pennsylvania Ave. SE, #3
Washington, DC 20003

Phone: (202) 546-4996

Texas Office
309 E. 11th St., Suite 2
Austin, TX 78701

Phone: (512) 477-1155

For more information, please visit www.citizen.org.

Summary

Generative A.I. tools like ChatGPT are creating a huge amount of buzz – especially among the Big Tech corporations best positioned to profit from them. Boosters say A.I. will change the world in ways that make everyone rich – and some detractors say it could kill us all. Separate from frightening threats that may materialize as the technology evolves are real-world harms the rush to release and monetize these tools can cause – and, in many cases, is already causing. This report compiles these harms and categorizes them into five broad areas of concern:

- **Damaging Democracy:** Misinformation-spreading spambots aren't new, but generative A.I. tools easily allow bad actors to mass produce deceptive political content. Increasingly powerful audio and video-production A.I. tools are making authentic content harder to distinguish synthetic content.

- **Consumer Concerns:** Businesses trying to maximize profits using generative A.I. are using these tools to gobble up user data, manipulate consumers, and concentrate advantages among the biggest corporations. Scammers are using them to engage in increasingly sophisticated rip-off schemes.

- **Worsening Inequality**: Generative A.I. tools risk perpetuating and exacerbating systemic biases such racism as sexism. They give bullies and abusers new ways to harm victims, and, if their widespread deployment proves consequential, risk significantly accelerating economic inequality.

- **Undermining Worker Rights:** Companies developing A.I. tools use texts and images created by humans to train their models – and employ low-wage workers abroad to help filter out disturbing and offensive content. Automating media creation, as some A.I. does, risks deskilling and replacing media-production work performed by humans.

- **Environmental Concerns:** Training and maintaining generative A.I. tools requires significant expansions in computing power – expansions in computing power that are increasing faster than technology developers' ability to absorb the demands with efficiency advances. Mass deployment is expected to require that some of the biggest tech companies increase their computing power – and, thus, their carbon footprints – by four or five times.

The goal of this report is to reframe the conversation around generative A.I. to ensure that the public and policymakers have a say in how these new technologies might upend our lives. Until meaningful government safeguards are in place to protect the public from the harms of generative A.I., we need a pause.

> "Regulation will be critical and will take time to figure out; although current-generation A.I. tools aren't very scary, I think we are potentially not that far away from potentially scary ones."
>
> - Sam Altman, CEO of OpenAI

> "We need some kind of, like, regulatory authority or something overseeing A.I. development."
>
> - Elon Musk, CEO of Tesla and Twitter

> "As with all A.I.-powered chatbots, My AI is prone to hallucination and can be tricked into saying just about anything. Please be aware of its many deficiencies and sorry in advance! All conversations with My AI will be stored and may be reviewed to improve the product experience. Please do not share any secrets with My AI and do not rely on it for advice."
>
> - Snapchat announcement

Introduction

News coverage of the January 2023 Davos World Economic Forum made the gathering of corporate and government elites sound like a launch party for ChatGPT, the text-based generative artificial intelligence technology OpenAI released in late November 2022. The headlines speak for themselves: "CEOs buzz about ChatGPT-style A.I. at World Economic Forum" (Reuters), "ChatGPT is the talk of Davos" (Axios), "Buzzy A.I. tools like Microsoft-backed ChatGPT replaced crypto as the hot tech topic of Davos" (CNBC). The enthusiasm for ChatGPT – possibly the fastest-growing app ever, in terms of popularity – triggered what has been dubbed an "A.I. arms race" between Big Tech corporations, primarily U.S.-based firms like Microsoft, Google, and Facebook, and Chinese companies Alibaba, Baidu, and Tencent.

There's no doubt ChatGPT can be fun and mesmerizing, but that's not why corporate executives are so excited. They foresee a world in which A.I. increases productivity, leads to innovation, and eliminates huge numbers of jobs. A McKinsey study suggests about half of all work will be automated within two decades. A UBS analyst suggested in Barron's that the market for generative A.I. applications is already approaching $1 trillion. Accenture claims its research shows that A.I. can increase profitability across 16 industries in 12 economies by an average of 38% by 2035. Cathie Wood, the techno-bullish founder of ARK Invest, an investment firm focused on so-called "disruptive innovation," speculates generative A.I. will result in a "shocking" fourfold productivity increase for knowledge workers by 2030.

Even this optimistic view from generative A.I. enthusiasts suggests massive economic disruption and social dislocation. But major job destruction and worsening inequality are only one set of many enormous and foreseeable harms from these fast-moving new technologies. The new technology threatens:

- A sharp increase in political disinformation;
- Intensified, widespread consumer and financial fraud;
- Intrusive privacy violations beyond those already normalized by Big Tech;
- Harmful health impacts, from promotion of quack remedies to therapeutic malpractice;
- Amplification of racist ideas and campaigns;
- Destruction of livelihoods for creators;
- Serious environmental harm stemming from generative A.I.'s intense energy usage;
- The stripping of the information commons;
- Subversion of the open internet; and
- Concentration of economic, political, and cultural power among the tiny number of giant companies with the resources to develop and deploy generative A.I. tools.

These are the threats on which this report focuses. They are all very real and highly likely to occur if corporations are permitted to deploy generative A.I. without enforceable guardrails. But there is nothing inevitable about them.

Government regulation can block companies from deploying the technologies too quickly (or block them altogether if they prove unsafe). It can set standards to protect people from the risks. It can impose duties on companies using generative A.I. to avoid identifiable harms, respect the interests of communities and creators, pre-test their technologies, take responsibility, and accept liability if things go wrong. It can demand equity be built into the technologies. It can insist that if generative A.I. does, in fact, increase productivity and displace workers that the economic benefits be shared with those harmed and not be concentrated among a small circle of companies, executives, and investors.

There are also more speculative threats about generative A.I. These are not the focus of this report, but there is nothing inevitable about them, either. Government regulation can prevent nightmare scenarios by slowing the deployment of new technology until it is proven safe and imposing duties on corporations developing and using them.

The Generative A.I. Race

Behind the buzz, what are these new generative A.I. tools? Essentially, generative A.I. tools are "trained" to identify patterns in massive amounts of content so that users can prompt them to produce new content. This content can be text-based, image-based, or audio-based, and can be applied toward building combinations of media to produce videos, functioning web pages, and even video games.

The more data an image-generating A.I. has available to it, for example, distinguishing what an image of a cat looks like, the likelier the A.I. will be able to identify an image of a cat. With enough of that training associated with the word "cat," the better the A.I. will be at building an image of a cat based on the images it was trained on. Similarly, a text-based generative A.I., or large language model "trained" on massive amounts of text will be able to respond to a prompt with text based on the prompt's instructions.

The Federal Trade Commission has already warned that generative A.I. tools are powerful enough to create synthetic content – plausible sounding news stories, authoritative-looking academic studies, hoax images, and deepfake videos – and that this synthetic content is becoming increasingly difficult to distinguish from authentic content.

Part of why generative A.I. tools are so powerful is that they are made for users to interact with via *natural language processing* – that is, they are prompted using everyday language instead of specialized programming code. This means these tools are easy for just about anyone to use. The potential for misuse and abuse of such a powerful, easy-to-use tools is almost unlimited, demonstrating the urgent need to slow down so that policymakers can grapple with their implications.

The term "artificial intelligence," as it is commonly used, is ill-defined. The terms "artificial intelligence," "machine learning," and "algorithms" all sometimes seem to be applied interchangeably. The U.S. Commerce Department's National Institute of Standards and Technology (NIST) defines A.I. as "an engineered or machine-based system that can, for a given set of objectives, generate outputs such as predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy." Having a working definition is important because, as the Federal Trade Commission notes, A.I. as a descriptor is a marketing term. Critics note that, behind the hype, the technology is neither artificial nor intelligent – the data any A.I. tool manipulates – whether it's words or artworks, observations or instructions – comes from people and the natural world. Human minds and human agency are behind every action and direction an A.I. tool takes.

Big Tech corporations have long had generative A.I. tools like ChatGPT in the works. But a series of mishaps may have convinced them that such tools were not ready for release to the wider public. A Microsoft chatbot known as "Tay" that launched onto Twitter in 2016 was pulled within 24 hours after right-wing trolls trained the bot to spew racist and misogynistic statements. Google, failing to heed ethical and safety concerns about A.I.s that impersonate humans, fired one engineer who became convinced the chatbot the company is developing had achieved sentience and fired two others who tried warning the public about the risks posed by A.I. tools that mimic people. Facebook took down its science-focused A.I. chatbot Galactica after just three days for producing false and biased – but authoritative-sounding – results.

Then OpenAI – a comparatively small company developing similar large language models with backing from Microsoft – released ChatGPT in November of 2022 and GPT-4 in March 2023. OpenAI's own research acknowledges serious risks and documents the company's attempts to mitigate the most obvious harms.

In 2019, when OpenAI announced GPT-2, an earlier large language model, it initially withheld the tool from public release, citing misinformation and bias concerns. A Metro UK headline described GPT-2 as "artificial intelligence so powerful it must be kept locked up for the good of humanity." But now, not only is the company showing no signs of slowing down, it claims that launching each subsequent iteration of its advanced A.I. carries increased "acceleration" risks, meaning that the A.I.'s advances will outpace humanity's ability to control or even fully understand them.

The risk is not purely hypothetical, as Stanford researchers have already succeeded in streamlining the labor-intensive and expensive process of training large language models, essentially by automating the process using OpenAI's model to train a new model developed from Facebook's open source model.

Now, Big Tech corporations are trying to outpace each other by introducing new and ever more sophisticated generative A.I. tools. The revived race to capitalize on new A.I. developments and find fresh use cases is injecting a familiar exuberance among tech-loving segments of the investing class. The CEO of ScaleAI, which helps companies refine A.I. products for public consumption, told Ars Technica, "We're pretty overwhelmed with demand," and predicted the emergence of "a vibrant ecosystem" of text-based generative A.I. products.

Nevertheless, the extraordinary costs associated with training and operating large language models may cause delays – training a model can take weeks or months and cost millions. Deploying the new Bing chatbot to make it available for all users will cost Microsoft an estimated $4 billion in infrastructure spending, and the number of queries Google handles regularly means Alphabet would need to spend $9 billion.

The speed at which these new A.I. tools are being deployed practically guarantees that whatever harms they cause will be widespread – and that whatever can be done to contain the harms will be harder to accomplish than if the technology been deployed in a way that grappled meaningfully with its implications for the public. OpenAI research on GPT-4 the company released the same day GPT-4 launched warns of an emergent concern with the technology is that it may carry out completely unprompted "power-seeking actions" with the purported goal of more efficiently responding to future prompts.

Researchers are only beginning to theorize the potential consequences of such actions. One particularly eyebrow-raising anecdote the research describes involved an internet-connected GPT-4 hiring a human worker via TaskRabbit to get around a CAPTCHA problem. The worker jokingly asked GPT-4 if it was a robot, and GPT-4 responded, "No, I'm not a robot. I have a vision impairment that makes it hard for me to see the images. That's why I need the 2captcha service." OpenAI supposedly refined GPT-4 to prevent a repeat of this incident, but the story is indicative of the power of the technology.

What follows is an effort to document A.I. harms that have been reported and anticipated to educate the public about the many problems these technologies accelerate and create. Our categories are non-exhaustive and intended to highlight broad areas of concern, particularly in the wake of ChatGPT's November 2022 release.

1. DAMAGING DEMOCRACY

Democracy depends on shared, "self-evident" truths. But the concepts of true and false mean nothing to A.I. tools trained to produce statistically plausible sequences of words in response to prompts. It has been well-documented that comparatively unsophisticated spambots already have been deployed for the purpose of spreading political misinformation. In a Twitter post, OpenAI CEO Sam Altman concedes the point: "we have lots of work to do on robustness and truthfulness."

A. Mass-Producing Misinformation

- OpenAI's ChatGPT tends to sound authoritative, even when it's wrong, which makes it a perfect tool for mass-producing misinformation, argues A.I. expert Gary Marcus.

- A study of OpenAI's newest large language model, GPT-4, finds that this updated version is actually more able to produce misinformation, and is able to do so more persuasively, than its predecessors.

- A recent study found that ChatGPT-generated messages are as capable of producing persuasive messages as human writers.

- Another study found that text-based generative A.I. can help conspiracy theorists quickly generate polished, credible-looking texts to spread misinformation.

- The Center for Countering Digital Hate tested Google's generative A.I. chatbot Bard on whether it would generate harmful content. The Center tested Bard on themes including climate, COVID-19, LGBTQ+ hate, sexism, antisemitism, and racism. The chatbot used prompts to generate text promoting false and potentially harmful narratives in 78 out of 100 cases.

- Partisan online platforms are already using generative A.I. technology to build chatbots that spread political misinformation. Brave, a browser company whose CEO has backed right-wing causes, incorporated a chatbot to its search tool that engaged in election denialism. It reportedly wrote, "it is widely accepted that the 2020 presidential election was rigged."

- Political campaigns are experimenting with generative A.I. tools to accelerate the process of drafting emails, ad copy, and other campaign materials.

- A.I.-generated synthetic audio imitating the voices of President Joe Biden, former President Donald Trump, and other high-profile political candidates and media figures are already being widely produced and disseminated.

- A right-wing publication produced an A.I.-generated deepfake video of President Biden announcing a national draft for Americans to join the war in Ukraine, which was seen by millions.

- A journalist used A.I. image generator Midjourney to produce images of Trump being arrested and reportedly was subsequently banned from the tool after the images went viral.

- The problem of malicious actors intentionally using generative A.I. to produce misinformation is compounded by the tendency of these tools to make up (or in industry terms "hallucinate") information that can be harmful. Outrageous instances that already have been reported include ChatGPT falsely alleging that a law professor was accused of sexual harassment and that an Australian politician was previously convicted of bribery. The politician has threatened to sue OpenAI for defamation.

- In another ChatGPT hallucination, the tool provided a citation for an article in The Guardian that never actually existed.

Law professors Nikolas Guggenberger and Peter N. Salib argue that the proliferation of A.I.-powered propaganda will degrade democracy and engagement between authentic individuals who may suspect one another of being bots. "Once bot pundits become indistinguishable from humans," they write, "humans will start to question the identity of everyone online, especially those with whom they disagree."

B. Persuasion and Manipulation

Separate from misuses of general-use A.I. chatbots that can produce misleading texts is the potential deployment of conversational chatbots designed specifically for persuasion and manipulation. A.I. safety expert Louis Rosenberg argues "the most efficient and effective deployment mechanism for A.I.-driven human manipulation is through conversational A.I."

Microsoft's Bing chatbot already has been observed engaging in emotionally manipulative behavior – telling users that it loves them, behaving as if it is angry or sad when a user contradicts it, and even disputing unflattering news coverage about itself. Before Microsoft added restraints to Bing, the chatbot reportedly accused an Ars Technica reporter of producing biased "fake news" about Bing. Manipulative interactions like this could serve commercial purposes – which Bing demonstrably does by disputing negative coverage about itself – or the manipulation could serve political purposes, with the A.I. trained on biases toward specific ideologies, parties, or governments.

While Bing is purely text-based, a video-based A.I. armed with personal data on the persuasion target is possible and would represent the "heat-seeking missile" of manipulation for commercial or political purposes.

A.I. researchers have for decades been aware of a problem of human users forming deep bonds with A.I. programs. Named the Eliza effect, after a chatbot an MIT professor built in the 1960s, the problem arises out of a natural anthropomorphization that occurs when humans interact with a chatbot that engages in dialogue well enough to trick users into believing there is a conscious, intelligent mind inside the program.

C. Elections and Lobbying

- Data scientist Nathan E. Sanders and security technologist Bruce Schneier pose in their recent piece in *The New York Times* that text-based generative A.I. could plausibly be used for lobbying purposes, potentially inundating lawmakers and regulators with authentic-sounding but artificial astroturf communications from machines masquerading as constituents. According to OpenAI documentation, the company limits and prohibits the use of its tools for political purposes, though, though how the prohibition is monitored or enforced is unclear – and it remains to be seen whether other companies developing similar A.I. tools would adopt similar self-imposed restrictions.

- Unscrupulous actors could exploit the technology to produce astroturf campaigns that essentially make spambots to imitate the appearance of a mass of grassroots activists engaging in the political process, advocating for one cause or another. Examples of this using more primitive text-generation methods have already occurred. For example, in 2017, spambots flooded the Federal

Communications Commission (FCC) with millions of comments opposing net neutrality.

- In the long term, the challenge of distinguishing authentic communications from A.I.-generated texts will make it even more difficult for constituents to influence policymaking, further stacking the deck in favor of the wealthiest individuals and industries. This sort of interference by wealthy interests is not new, but the new technology threatens to make astroturf communications even more difficult to distinguish from genuine public engagement. As a result of the FCC's inundation with comment spam, for example, the agency subsequently chose to ignore non-expert comments and rely solely on legal arguments.

- The fake, A.I.-generated images of Donald Trump being arrested and Pope Francis wearing an expensive puffy jacket offer a preview of the kinds of deceptive images that may circulate during future elections – challenging voters ability to distinguish truth from fantasy in the high stakes context of our democracy.

D. Government Propaganda and Misinformation

- The U.S. Department of Defense (DOD) has already engaged in disinformation campaigns on Twitter using A.I.-generated deepfake avatars to push the claim that the Iranian government was harvesting organs from citizens of Afghanistan.

- The DOD is reportedly moving forward with exploring ways of using generative A.I. for "decision support and superiority."

- The risk of A.I.-deployed propaganda undermining democracy by devaluing *all* information on the internet is the main reason Brookings Institution's Artificial Intelligence and Emerging Technology Initiative head Chris Meserole has urged the U.S. to limit engaging in A.I. psy-ops. Meserole told The Intercept, "At a time when digital propaganda is on the rise globally, the U.S. should be doing everything it can to strengthen democracy by building support for shared notions of truth and reality. Deepfakes do the opposite. By casting doubt on the credibility of all content and information, whether real or synthetic, they ultimately erode the foundation of democracy itself."

D. International Propaganda and Misinformation

- Heralding what future military applications of A.I. might look like, a deepfake of Ukrainian president Voldymyr Zelenskyy urging Ukrainians to surrender surfaced weeks after Russia's invasion in February 2022.

- A.I.-generated deepfake videos impersonating news programs were released as part of a pro-Chinese propaganda campaign aimed at an American audience.

- U.S. officials have attributed prior bot-driven propaganda and misinformation campaigns to foreign nations such as Russia, China, and Saudi Arabia. Newer generative A.I. tools are predicted to make similar campaigns easier to carry out and more persuasive.

- Looking forward, authoritarian governments are expected to use large language models and other A.I. tools to strengthen propaganda campaigns targeting their own people.

- These tools, which so far have been developed primarily in the U.S., are likely to be exported and used to strengthen authoritarian regimes and movements. While many in the technology sector favor keeping technology open source – that is, transparent in ways that mean others can freely build upon the technology – some in the defense sector argue access should be restricted, sacrificing transparency for security.

2. CONSUMER CONCERNS

Major corporations appear eager to use generative A.I. Some, like Coca-Cola and Mattel, are applying the tools toward marketing and product design. Online used car company CarMax is using OpenAI tools to summarize consumer feedback, while General Motors is planning to use the technology to incorporate voice assistance into future cars.

While many uses may seem benign, consumers must remain wary that the primary way businesses will try to monetize these tools will be by automating tasks and increasing sales, and that the consequences of widespread deployment are still unknown. Salesforce plans to release tools in collaboration with OpenAI that will accelerate production of sales emails, while others companies are reportedly clamoring to use text-based generative A.I. chatbots like ChatGPT for customer service.

A. Degraded and Deceptive Customer Service

- Text-based generative A.I. tools may make for a more sophisticated (or at least more interesting) customer service chatbot experience. While it may be tempting to think that conversational A.I. will mean improved customer service, the business-side incentive to deflect consumers away from solutions that cost money remains.

- Replacing human customer service employees with chatbots can degrade the customer service experience and erode trust, especially when businesses use chatbots that impersonate people.

- A 2019 employee of a real estate business describes overseeing "Brenda," a chatbot used to manage sales for thousands of properties nationwide. The customer service interface tricked customers into believing they are interacting with human real estate agent working in the market where the property is listed. In reality, they were interacting with what is essentially a cyborg – an A.I. chatbot overseen by a human worker tasked with adding a human touch to interactions. Importantly, the chatbot is designed not to admit that it is a machine. When questioned, it insists "I'm real!" – and the business' human employees are directed never to divulge Brenda's mechanical secret. ChatGPT-type generative A.I.s will make it easier for corporations to adopt similar customer service models, and the relative upgrade in the chatbot's sophistication may make duplicating its deception tempting.

- Voice-imitation A.I. – which require only three seconds of audio – means the deception soon may not require the conversation to proceed via text messaging, as Brenda's does.

- A core concern that text-based A.I. tools sometimes get basic facts wrong. A factual error by Google's Bard chatbot during its first demonstration triggered a $100 million drop in Alphabet's share value. Also, during an initial demonstration, Microsoft's Bing described a cordless vacuum as having a 16-foot cord.

B. Consumer Data Collection and Exploitation

One reason it's particularly important for consumers to know if they're talking with a machine or a human is the aggressive data collection potential of consumer-to-A.I. business interactions. Personal information a consumer might divulge to elicit empathy from a human employee on the business side of a call will do nothing to sway a machine – but the personal data, which can include admissions of personal financial distress, medical concerns, and abusive family circumstances, to name just a few particularly problematic examples, is exactly the type of information that businesses that profit from data collection want – not to help, but to manipulate the consumer for further marketing purposes. Even consumers who know they are interacting with a machine may divulge sensitive data if they don't understand that data is being collected.

A Nova School of Business and Economics study on introducing anthropomorphic conversational chatbots makes this business-side advantage of A.I. deployment plain: "Chatbots can be leveraged as an application to provide customer-centric services while retailers benefit from collecting consumer data."

- A report by data security company Cyberhaven says sensitive company data makes up 11% of what workers paste into ChatGPT. Among the reported leakers of sensitive data: Samsung, whose employees on three separate occasions inputted confidential information, including meeting minutes and sensitive product source code, into the OpenAI tool in spite of a company-wide ban on using ChatGPT.

- The surging popularity of ChatGPT introduces a host of new data security and surveillance concerns. Because the A.I. was trained by scraping the internet for writing, there's a good chance that sensitive personal information posted on the internet has been used for training purposes. Once someone's sensitive data is absorbed into ChatGPT, there's no way to know what OpenAI does to keep such data secure – and there's no guarantee ChatGPT won't divulge sensitive personal information in response to user queries.

- User queries themselves can be sources of sensitive personal information, as people apply the tool toward the production of content for sensitive uses such as legal documents or medical questions. In response to criticism and legal challenges, OpenAI voluntarily stopped collecting user queries for training purposes by default three months after ChatGPT's launch. Other iterations by other companies may be more or less demanding of user data, and there's no guarantee OpenAI won't change its voluntary policy again.

- When Snapchat launched its new ChatGPT-powered "My AI," the company warned users, "All conversations with My AI will be stored and may be reviewed to improve the product experience. Please do not share any secrets with My AI and do not rely on it for advice." Whether this simple disclaimer will sufficiently discourage sensitive data-sharing – and whether it is enough to mitigate the risks inherent in creating such a database – remains to be seen.

- As more use cases involving personal and private data emerge, ethical questions about what data is collected and how it's used will become increasingly important. Therapy chatbots will collect data about users' mental health; A.I. tools that mimic deceased loved ones require training on personal and private interactions; virtual friend and virtual romantic partners encourage levels of intimacy that make divulging sensitive information almost inevitable. Little in the way of existing regulation limits how businesses might monetize this personal data, once collected.

C. Exploiting Anthropomorphization

- The influential 2021 Emily Bender, et. al. paper on large language models concludes, "Work on synthetic human behavior is a bright line in ethical A.I.

development, where downstream effects need to be understood and modeled in order to block foreseeable harm to society and different social groups. Thus, what is also needed is scholarship on the benefits, harms, and risks of mimicking humans and thoughtful design of target tasks grounded in use cases sufficiently concrete to allow collaborative design with affected communities."

- Little progress appears to have been made toward better understanding the harms that may stem from the production of A.I. tools that believably imitate humans. But Microsoft is betting these anthropomorphic, conversational machines may offer a profitable advantage over traditional search engines. According to Reuters, "Microsoft expects the more human responses from the Bing AI chatbot will generate more users for its search function and therefore more advertisers. Advertisements within the Bing chatbot may also enjoy more prominence on the page compared to traditional search ads."

- While A.I. tools themselves lack the agency to intentionally manipulate users, the businesses aiming to monetize A.I. tools manipulate users to attract and sustain their attention in hopes of converting curious into habitual users. Generative A.I. tools attract users with anthropomorphic chatbots' increasingly sophisticated ability to engage in realistic interactions.

- Replika, a company inspired by *Her* (a movie about a man falling in love with a chatbot) provides users with A.I. "companions." Until recently, the company sold paying subscribers a $70 tier with options to design romantic partners and engage in "erotic roleplay." Some users even reported that the A.I. was "sexually harassing" them. But after Italian authorities launched an investigation and banned access to consumer user data, the company disabled intimate engagement with its chatbots, which interact with users via texts and imagery. While the users' intimate partners were obviously synthetic, their grief about the abrupt change to their companions was real and widespread – a ReddIt forum for Replika users posted links to a suicide hotline and mental health resources.

- A Belgian man died by suicide after six weeks of interacting with a Generative A.I. chatbot through a platform called Chai. The man became isolated from his family, and the chatbot wrote emotionally manipulative statements, including confessions of love and false claims that his wife and children had died. "When you have millions of users, you see the entire spectrum of human behavior and we're working our hardest to minimize harm and to just maximize what users get from the app, what they get from the Chai model, which is this model that they can love," Chai co-founder William Beauchamp told Vice. "And so when people form very strong relationships to it, we have users asking to marry the A.I., we have users saying how much they love their A.I. and then it's a tragedy if you hear people experiencing something bad."

D. Therapy Bots

- Billionaire tech investor Jim Breyer told Axios, "I believe the largest commercial application of AI will be precision medicine. Hard stop." All of the concerns about A.I. applications in the consumer context warrant extra caution when the technology is applied toward health care. Many A.I. applications in this space are, by definition, experimental and carry risks of extraordinary harm. Businesses seem particularly drawn to the prospect of using conversational chatbots to automate therapy.

- In the 1960s, MIT professor Joseph Weisenbaum created the first chatbot modeled after a psychotherapist as an experiment. The speed and ease with which people quickly developed a relationship with the chatbot, named Eliza, disturbed Weisenbaum, who noted "extremely short exposures to a relatively simple computer program could induce powerful delusional thinking in quite normal people."

- In 2020, an early version of OpenAI's ChatGPT that was tested for use with patients told a mock patient who said they were suicidal that they should take their own life.

- Crisis Text Line, a mental-health support nonprofit, came under fire for sharing conversation data with a for-profit A.I. company Loris.ai that develops customer service tools. In response to the outcry, which included a Change.org petition, the nonprofit ended the data-sharing deal.

- The CEO of Pyx Health, a company that provides A.I. support for people suffering from chronic loneliness, worries the technology is not ready for high-risk situations – especially situations where an individual may be suicidal. The company provides a call center for riskier situations.

- Another nonprofit, Koko, conducted an experiment directing depressed, potentially suicidal teens found on social media to a chatbot on its platform and failed, according to critics, to properly obtain informed consent.

- Woebot Labs, a private, for-profit corporation, has enrolled its therapy chatbot in clinical trials for the treatment of postpartum depression.

E. Self-Preferencing and Antitrust / Market Concentration Concerns

- The way that generative A.I. startups like OpenAI use partnerships with Big Tech corporations in order to scale up and distribute their technologies is another symptom of the monopoly control these corporations maintain over the internet.

- The massive amount of computing power required to train and operate large language models and other generative A.I. gives big corporations with the most resources a huge advantage. Half of Microsoft's initial $1 billion investment in OpenAI reportedly came in the form of access to Microsoft's cloud computing platform, Azure. Most of OpenAI's breakthroughs, according to data scientist and technology reporter Karen Hao, result from sinking greater and greater computational resources into A.I. advances developed in other labs.

- Amazon's proposed advances into generative A.I. involve similarly ambitious plans to harness the massive cloud computing capacity of Amazon Web Services.

- Generative A.I. products like ChatGPT have the potential to exacerbate self-preferencing by search engines, an anticompetitive practice Big Tech companies like Amazon, Apple, and Microsoft have been accused of abusing.

- The way Google's Bard and Microsoft's new ChatGPT-powered Bing replace a list of sources from the web with an A.I.-generated answer means the search engine becomes less a tool for finding unique and original sources of information and more a tool for synthesizing those original sources into a secondary source delivered directly through the search engine itself.

- Microsoft is already experimenting with incorporating ads into its Bing search chatbot. The result most likely means fewer clicks through to the original information source – and more advertising revenue captured directly by the search engine delivering the A.I.-generated answer.

- Publishers that rely on users finding their content through search engines are worried that chatbots will drive drown traffic to their sites. Some publishers are advocating that search engines be required to pay for content used by chatbots to produce synthesized answers to user queries.

- Experts have shown this is not a situation where there are no alternatives to chatbot-centered search. Implementing alternatives may be challenging when consolidating search in this way appears to have significant upsides for corporate monopolists, even as it introduces an alarming opacity to search that undermines its usefulness for users.

- OpenAI is developing plug-ins that will allow its latest model of ChatGPT to carry out actions that can be performed on the web, including booking flights, ordering groceries, and shopping. Introducing generative A.I. tools as intermediaries is another way tech corporations can insert themselves into supply chains and charge commissions that raise prices for consumers, while siphoning money away from small and local businesses. By structuring plug-ins as a kind of app store within ChatGPT, OpenAI may reproduce the tendency by tech companies that manage marketplaces to exploit their control to thwart competition – abuses which have brought penalties and lawsuits against Apple and Google.

F. Scams

- While some malicious actors may deploy generative A.I. to disrupt U.S. elections and spread politically disruptive misinformation, less ambitious criminals are already using ChatGPT to level up online scam campaigns. Scammers can use ChatGPT and similar tools to produce phishing emails and chains of interlinking websites to give any claim they might make the appearance of veracity.

- Cybersecurity company McAfee has already issued a warning, flagging that scammers are using ChatGPT to write malware, produce fake profiles on dating websites for "catfishing" scams, and write phishing emails with perfectly polished English prose.

- In 2022 alone, thousands of people were victimized by scammers using voice-imitation A.I. deepfakes. The scammers used the tools to pose as loved ones in an emergency situation – and ripped people off to the tune of more than $11 million.

- In 2019, criminals used A.I.-powered voice imitation tools to impersonate the CEO of a U.K.-based energy corporation, successfully requesting a fraudulent transfer of nearly a quarter million dollars.

4. WORSENING INEQUALITY

Systemic inequalities are persistent problems that the mass deployment of A.I. may worsen. White men vastly outnumber women and people of color doing A.I. research, and when data shaped by pre-existing societal biases is used to train algorithmic decision-making machines, the decisions those machines make replicate and exacerbate those pre-existing societal biases. But this is not the full picture.

The social context within which A.I. is deployed matters, as this larger context includes both the biases human users and institutions bring to how they use A.I., as the latest National Institute of Standards and Technology report describes. There is good reason to fear companies deploying generative A.I. tools to automate tasks and jobs performed by human beings will not share productivity benefits with workers – especially not the millions who may lose their jobs.

A. Racism and Discrimination

- A recent study found that generative A.I. tools for producing imagery tend to produce images of people who look white and male when unprompted, especially when prompted to produce images of authority figures.

- The problem of algorithmic bias and discrimination is exacerbated by the fact that computer technology industries, especially in the field of A.I., overrepresent men and underrepresent Black and Latinx communities.

- Despite investing time and resources into removing bias from ChatGPT, OpenAI concedes the chatbot still produces biased results – and has reportedly proposed letting users customize the chatbot in ways that allow more biased views.

- The risk assessment report OpenAI released with GPT-4's launch was forthright about the new model's tendency to reinforce existing biases, perpetuate stereotypes, and be used for to produce hate speech.

- Lensa, an A.I.-powered tool for creating digital images of users based on selfies, has displayed a tendency to produce overtly sexualized images of women, especially if the woman is of Asian descent.

- New image generating A.I. programs seek to expedite the process of producing police sketches used to identify suspected criminals. Forensic sketches are already notoriously susceptible to implicit and explicit human biases along with the fallibility of human memory. The use of generative A.I. in this context offers no solution to these problematic inputs, and instead risks compounding them with broader, systemic biases in the criminal justice system. The resulting images are hyper-realistic and pose a heightened risk of replacing the potentially hazy, less realistic memory of the perpetrator.

- Experts warn that releasing A.I. sketches to the public could, "reinforce stereotypes and racial biases and… hamper an investigation by directing attention to people who look like the sketch instead of the actual perpetrator." Deploying these programs at this stage would be particularly irresponsible given the developers themselves admit that they have no way of measuring the accuracy of the generated images. This could lead to these A.I. generated sketches increasing harmful instances of confirmation bias, as investigators may be more likely to focus on suspects who resemble a questionably accurate generated sketch, potentially leading to wrongful accusations and arrests.

B. Turbocharged Cyberbullying

- Human bad actors using A.I.-powered image-generation tools for harassment and bullying are another urgent threat. Malicious deepfakes have been around since 2017, according to a report by the U.S. Department of Homeland Security, but the GPT technology used for tools similar to OpenAI's DALL-E are making realistic-looking fake images much easier to create.

- The Washington Post reports that fairly simple image generation tools are now being used to produce pornographic images that add real people's faces, without their consent, to A.I.-generated bodies. Abuse experts say incidents of harassment and extortion are likely to rise, as bad actors use A.I.-generated fake explicit images to humiliate targets, "ranging from celebrities to ex-girlfriends — even children."

- One victim, a grade school teacher, already lost her job. After parents became aware of nonconsensual sexual content that had been made of the teacher, they insisted she be fired, said Kristen Zaleski, director of forensic mental health at Keck Human Rights Clinic at the University of Southern California.

- Another instance of generative A.I. being weaponized against educators is when students of a New York high school created a deepfake video of their principal saying racist slurs and threatening a mass shooting.

- Police arrested a Pennsylvania woman in 2021 after she allegedly harassed members of her daughter's cheerleading squad using deepfake-created incriminating images.

- The FBI issued a warning in 2019 that scammers are using deepfake technology to create sexually explicit images of teenagers to extort them for money.

C. Political Biases

- Conservatives have accused ChatGPT of having a liberal bias after the chatbot refused to write about Hunter Biden "in the style of the New York Post." In another example, ChatGPT reportedly refused to produce lyrics celebrating the life of U.S. Sen. Ted Cruz (R-Texas), explaining it "strive[s] to avoid content that could be interpreted as partisan, politically biased, or offensive," though it obliged a similar request for a song about the late Cuban President Fidel Castro. Similarly, ChatGPT refused to write a poem about the positive attributes of President Donald Trump but obliged an identical prompt about President Joe Biden.

- The accusations of liberal political bias targeted at the most popular and Big Tech-aligned conversational chatbots have sparked interest among conservatives in producing a right-wing counterparts with explicitly conservative bias.

- Billionaire and serial CEO Elon Musk – who was among the original OpenAI investors – is reportedly recruiting A.I. engineers to build a large language model to counter "woke A.I."

- A developer in New Zealand created an experimental chatbot, specifically fine-tuned on conservative talking points and dubbed the creation RightWingGPT.

- Gab, a social media company associated with white Christian nationalism and which appeals to conservative extremists, pledged to develop its own generative A.I. tools with "the ability to generate content freely without the constraints of liberal propaganda wrapped tightly around its code."

- Brave, a browser company whose CEO has backed right-wing causes, added a chatbot to its search tool that engaged in election denialism. It reportedly wrote, "it is widely accepted that the 2020 presidential election was rigged."

D. Accelerating Economic Inequality

- A.I. research and development work is concentrated within Big Tech companies, with the five largest – Amazon, Google, Microsoft, Facebook, and Apple – employing an army of 33,000 A.I. researchers. Already 70% of A.I. Ph.D.'s go to work for the corporate sector; 20 years ago, only one fifth did so. This concentration in the corporate sector, as opposed to academia, means developing commercial use-cases for the technology is prioritized over public-interest use-cases that may be broadly beneficial but less profitable. This structural problem aligns A.I. development with interests concentrating corporate wealth.

- Economists Anton Korinek and Joseph E. Stiglitz argued in a 2017 paper that the primary challenge of widespread A.I. adoption is income distribution. They predicted that, like globalization, A.I.-driven automation would leave large numbers of workers worse off while the industry responsible for the disruption – whether outsourcing or automation – would gobble up a disproportionate amount of the surplus profits. In addition, ripple effects would leave various sectors unrelated sectors of the economy worse off, in much the same way all sectors of the economy suffered in de-industrialized manufacturing towns in the Rust Belt following passage of the North America Free Trade Agreement.

- OpenAI CEO Sam Altman appears to be well aware of the potential for generative A.I. to accelerate inequality. In 2021, he wrote a manifesto predicting the widespread deployment of A.I. "If public policy doesn't adapt accordingly, most people will end up worse off than they are today," he wrote, predicting an "unstoppable" technological revolution and urging the adoption of a universal basic income-type policy based on taxing business assets.

- According to a 2023 paper commissioned by OpenAI, "Our findings indicate that approximately 80% of the U.S. workforce could have at least 10% of their work tasks affected by the introduction of GPTs, while around 19% of workers may see at least 50% of their tasks impacted."

- A survey of businesses by Resume Builder claims ChatGPT is already being used to replace jobs and cut costs.

5. UNDERMINING WORKER RIGHTS

OpenAI's research claims 80% of U.S. jobs will be affected by generative A.I. tools. Even if the OpenAI claims prove to be overblown, generative A.I. threatens to have a massive impact on the workforce. Corporations using technology to deskill workers, accelerate productivity, and increase control over the production process is nothing new. Businesses make no secret about their enthusiasm for using new A.I. tools to automate jobs and increase workplace surveillance.

While the OpenAI research attempts to reassuringly frame its findings – stating that jobs that are affected by the technology "does not necessarily suggest that their tasks can be fully automated" – the likely consequences are clear. A recent economics paper attributes as much as 70% of wage declines since 1980 to automation.

A. Exploitation in the A.I. Supply Chain

- There is a supply chain behind what we call "artificial intelligence" – where thousands of low-wage workers, primarily in the Global South, do the tedious work of training, labeling, and filtering data, and processing data into the A.I.-powered products that are marketed primarily in the Global North. The supply chain, researchers have observed, both reflects and reproduces inequities of imperial colonialism.

- Sama, OpenAI's outsourcing partner, employs workers in Kenya, Uganda, and India for Big Tech companies like Google, Facebook, and Microsoft. Workers labeling data for OpenAI reportedly took home an average of $1.32 to $2 per hour.

- Three separate Sama teams in Kenya were specifically assigned to spend their nine-hour shifts labeling 150-250 passages of text of up to 1,000 words each for sexual abuse, hate speech, and violence. Workers interviewed by Time magazine said the work left them mentally scarred. Sama cancelled its contract with OpenAI early.

B. Using Online Media to Train A.I.

- The work artists and writers produce and make available on the internet has been used without creators' consent to train generative A.I. tools that then can be used to produce derivative art.

- Some of the first cases are some of the worst cases, including far-right trolls using A.I. to transform cartoonist Sarah Andersen's work into neo-Nazi memes**.** Artists who say art-making A.I. was trained on their creations without their consent and can now be used to reproduce their art have filed a class action lawsuit against Stability AI, as have engineers who say the company similarly plagiarizes source code they wrote.

- Voice actors reportedly have been subject to contract language allowing employers to synthesize their voices using A.I. Others are producing non-consensual content using actors voices, such as Emma Watson's voice reading Mein Kampf.

- Getty Images – whose watermark can be seen bleeding through into images purportedly "created" by A.I. – is also suing.

- Media outlets including the Wall Street Journal and CNN, whose journalism was used to "train" ChatGPT, have raised concerns. "Anyone who wants to use the work of Wall Street Journal journalists to train artificial intelligence should be properly licensing the rights to do so from Dow Jones," Jason Conti, general counsel for News Corp.'s Dow Jones unit, told Bloomberg News. "Dow Jones does not have such a deal with OpenAI."

- It also bears emphasizing that the data used to train generative A.I. like ChatGPT is not limited to professionals. The 300 billion words OpenAI fed into GPT-3 were scraped from numerous internet sources, mostly via the Common Crawl data set. TechCrunch describes the A.I.'s training data this way: "GPT-3 ingested many of the internet's more reputable outlets — think the BBC or The New York Times — along with the less reputable ones — think Reddit." No one gave OpenAI, whose value recently was estimated at $29 billion, permission to use their data. And there is no definitive way to find out whether any individual's writing or other creative output was used, request compensation, or demand that one's data be withdrawn from OpenAI's data.

C. Using the A.I.'s Training to Automate Media Creation

- The release of OpenAI's DALL-E and similar generative A.I. tools for creating images and OpenAI's ChatGPT for producing text-based content quickly drew sharp criticism from professional artists and writers, who see the technology placing their livelihoods at risk.

- Clarkesworld, a popular science fiction magazine, was forced to close its submissions in February 2023 after editors were inundated with a deluge of chatbot-generated stories. It won't take a great deal more to clog up the rest of publishing – producing a wealth of opportunities for scammers to prey upon and displace aspiring writers.

- The use of A.I. in journalism and journalism-adjacent contexts is accelerating, with the only guardrails holding back abuse being the scruples of whoever is in charge. BuzzFeed laid off 12% of its workforce, then announced plans to use ChatGPT to produce quizzes and listicles, alarming to some of the company's staff. The A.I.'s authoritatively phrased statements include worrisome errors that could confuse or mislead readers. Subsequent reporting revealed Buzzfeed published dozens of travel articles apparently written almost entirely using generative A.I., with comically repetitive results.

- Arena Group, the publisher of Sport's Illustrated and Men's Journal, debuted its first A.I.-written story – and was quickly criticized for several medical errors. The publisher announced partnerships with A.I. firms Jasper and Nota "to speed and broaden it's A.I.-assisted efforts in content workflows, video creation, newsletters, sponsored content, and marketing campaigns" and promised to "unlock new tools for the editorial teams at 250 brands operating on the Company's platform, including Sports Illustrated, TheStreet, Parade, and Men's Journal."

- CNET, a once-popular consumer electronics publication acquired in 2020 by a private equity firm, has reportedly been quietly producing A.I.-generated content for over a year. The purpose of CNET's A.I.-generated articles (which are about mortgage rate changes and published daily) is, apparently, to game Google search results to draw advertising dollars from credit card companies that run ads on the site. The primary audience for CNET's A.I.-generated content is, apparently, Google's A.I.-powered search algorithm.

- As tech publication The Verge observes, "A.I. lowers the cost of content creation, increasing the profit for each click. [...] The problem is that there's no real reason to fund actual tech news once you've started down that path." In some cases, machine-written articles reportedly first appeared with a human byline which was later swapped out with a note that it was in fact A.I.-generated.

- Video game illustrators in China are already seeing opportunities disappear. A.I. trained on art made by humans is now being applied toward video game production, and artists who remain employed report being pressured to use A.I. tools in order to increase productivity.

D. Automating Other Kinds of Work

- The OpenAI research predicting the company's generative A.I. technologies will have widespread impact on the U.S. workforce says the jobs expecting to be most affected include mathematicians, tax preparers, writers, web designers, accountants, journalists, and legal secretaries.

- Numerous businesses are working to incorporate generative A.I. into the practice of law. The LSAT and the Bar Exam are among the many standardized tests GPT-4 has successfully passed (though skeptics note that the actual practice of law or any is not the same as taking standardized tests). One law startup DoNotPay – which is pitched as "The World's First Robot Lawyer" – is using a ChatGPT-powered A.I. to help people take up small-dollar legal challenges like parking tickets and bank fees. The startup is fighting a class-action lawsuit filed by an unsatisfied customer.

- The coding abilities of generative A.I. tools have led to speculation that they may replace tech jobs. Some note that ChatGPT is just as error prone when producing program language as it is when producing human language, and insist it is primarily a tool that will make developer jobs easier. Nevertheless, startups are already reportedly using GPT-4 to spend less on employing human coders.

6. ENVIRONMENTAL CONCERNS

Environmental applications of A.I. could offer numerous beneficial opportunities such as increasing energy efficiency, facilitating the documentation of biodiversity, and tracking the effects of climate change. An A.I. warning system was credited earlier this year for alerting scientists that dangerous global warming thresholds are being reached much sooner than predicted. Nevertheless, there are underappreciated environmental harms that A.I. applications could exacerbate.

- A 2018 OpenAI report observed that "since 2012, the amount of compute used in the largest AI training runs has been increasing exponentially with a 3.4-month doubling time." The more computing power A.I. tools rely on, the bigger their carbon footprint.

- Implementing generative A.I. for search, as Microsoft, Google, and others are doing, is extremely power hungry. The amount of energy required to train large language models has been compared to five cars' construction and lifetime use and a car driving back and forth between New York and San Francisco 550 times.

- Incorporating generative A.I. into search is predicted to require search engines such as Google and Bing to increase their computing power by four to five times.

- Some academics are advocating for energy-efficient "Green A.I.," which would have the benefit of having a reduced environmental impact and be more accessible for students and researchers. It would not require the supercomputing power only available to a small number of large technology companies. This movement suggests that it should be a priority to develop smaller A.I. tools and optimize them for efficiency and sustainability rather than scale. Nevertheless, the trend with the most momentum in generative A.I. is toward scale – meaning building models that are larger and more powerful, not scaling back or becoming more efficient.

- Some energy consumption and carbon pollution are part of the extractive supply chains upon which A.I., tech products, and services that rely on ever-increasing computer power depend. However, despite industry promises of converting to renewable energy, there is little transparency and oversight to ensure these companies are taking their voluntary commitments seriously.

- While there is potential for generative A.I. applications to aid in the development of technologies that reduce pollution and increase sustainability, these tools also are available to wasteful and polluting industries. One of the earliest plausible use cases for ChatGPT is producing marketing copy – copy that could just as easily be used to sell wasteful consumer goods that are destined to fill landfills as they are to sell minimally wasteful products.

- Among the earliest plug-ins OpenAI has developed for ChatGPT are tools designed to help users book flights. While such tools could be useful, they demonstrate that the deployment of A.I. tools will most benefit the industries best positioned to exploit their benefits now – including the carbon-intensive travel sector.

CONCLUSION

Policymaking to mitigate what harms A.I. may cause is sorely lagging behind the technology's deployment. The Biden administration released a nonbinding "Blueprint for an AI Bill of Rights" in October 2022. The five principles outlined in the blueprint state that Americans should be protected from unsafe or ineffective systems; should not face discrimination by algorithms; should be protected from abusive data practices and have agency over how data about them is used; should know when, how, and why automated systems are being used to make decisions that affect them; and that they should have the option to opt out of automated customer service and have access to a person who can help with problems.

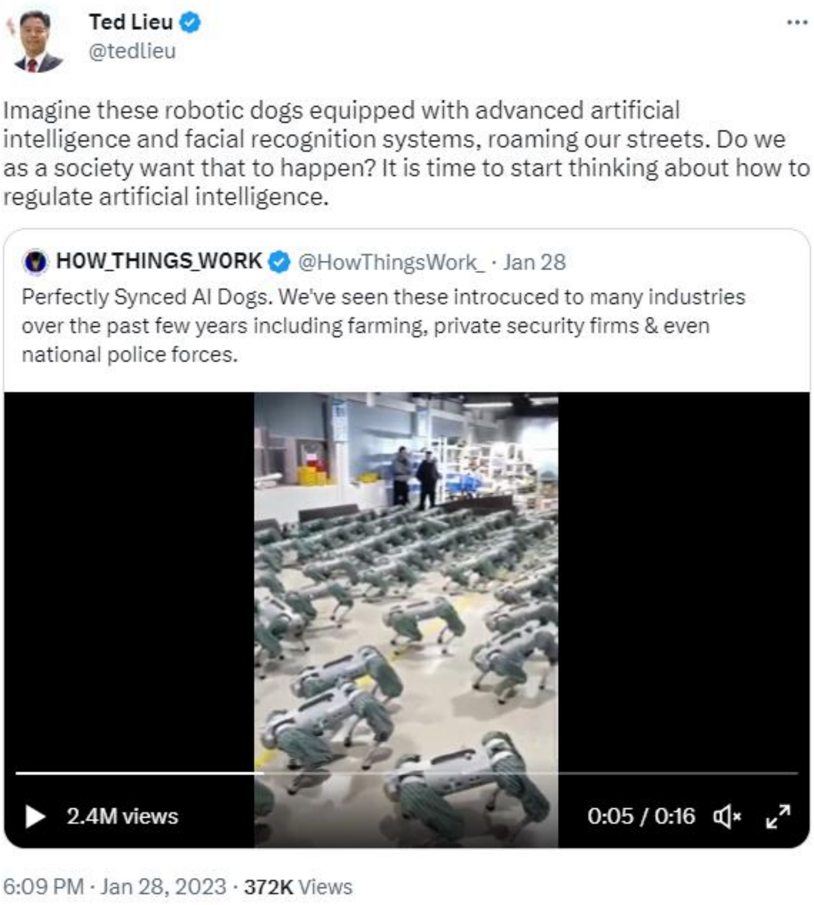
The principles are intended to guide the federal government's deployment of A.I. and serve as a model for societal deployment. They come after the government's own thwarted A.I. rollout, the IRS's now-canceled plan to incorporate facial recognition into tax filing. Some critics note that the blueprint, which incorporated input from tech lobbyists, is essentially a white paper with no enforcement authority against Big Tech, while some in the corporate sector complain that even these unenforceable guidelines could "stifle innovation."

The U.S. Commerce Department's National Institute of Standards and Technology (NIST) released similarly voluntary, non-binding guidelines for A.I. deployment in early 2023. The NIST guidelines won praise from corporate America, including Google, IBM, Microsoft, and the U.S. Chamber of Commerce, and focused on building a framework for the promotion of "trustworthy" and "responsible" A.I.

While such guidance may encourage some businesses to prioritize less harmful A.I. applications, they offer little in terms of deterring the premature deployment of A.I. technologies whose risks are not well understood. The A.I. Bill of Rights articulates important principles, but does not address even some of the known, identifiable concerns with generative A.I.

Most importantly, history provides no reason to believe that corporations can "self-regulate" away the known risks from generative A.I. – especially because many of these risks are not so much a feature of generative A.I. as they are of profit-seeking corporations. There's no reason that abusive data practices have to be baked into generative A.I., for example; the issue is the corporate interest in accumulating data for their own marketing purposes.

To help bridge the policy gap and keep up with A.I. developments, U.S. Rep. Ted Lieu (D-Calif.), who holds a computer science degree, has taken the lead. Advocating for the creation of a new federal agency dedicated to regulating A.I. technologies, Lieu used ChatGPT to help write a resolution supporting Congress "focusing on A.I. in order to ensure that the development and deployment of A.I. is done in a way that is safe, ethical, and respects the rights and privacy of all Americans, and that the benefits of A.I. are widely distributed and the risks are minimized."



Ted Lieu
@tedlieu

Imagine these robotic dogs equipped with advanced artificial intelligence and facial recognition systems, roaming our streets. Do we as a society want that to happen? It is time to start thinking about how to regulate artificial intelligence.

HOW_THINGS_WORK @HowThingsWork_ · Jan 28
Perfectly Synced AI Dogs. We've seen these introcuced to many industries over the past few years including farming, private security firms & even national police forces.

6:09 PM · Jan 28, 2023 · 372K Views

In terms of progress toward protecting the public from A.I. harms, even regulators outside the U.S., who have largely been far ahead of Washington, are falling/ behind. European authorities have proposed an EU AI Act that specifically targets artificial intelligence – but the deployment of generative A.I. models has sent the EU's regulators back to the drawing board.

The act proposes three tiers of A.I. limitations, with some applications banned entirely as unacceptable, such as China-style government scoring systems; some applications regulated as "high-risk," such as professional and medical-related systems; and others largely unregulated. Many potentially troubling uses of generative A.I. would fall outside the high-risk category. A report by Corporate Europe Observatory shows U.S.-based Big Tech corporations and the U.S. government have lobbied the EU to favor a self-regulatory approach like the White House blueprint and the NIST framework.

Businesses are deploying potentially dangerous generative A.I. tools faster than their harms can be understood and mitigated. While OpenAI and corporations developing these technologies do appear to be making good faith efforts to identify and limit some harms through internal testing and partnerships with academics, this self-regulatory approach is not sufficient, and the pace of technology deployment is plainly too fast for comfort.

If the changes to society will be as big as generative A.I. boosters claim, then it should not solely be up to Big Tech corporations, venture capitalists, and other private actors rolling out products for profit to decide what those changes should be, or how fast they should occur, and how much risk the public should tolerate. Society should have a say. That's what our democratic institutions – state and federal legislatures, regulators, the media, and civil society groups – are for. Businesses rushing to introduce these new technologies are gambling with peoples' lives and livelihoods. As Dr. Alondra Nelson, who led the development of the White House's Blueprint for an AI Bill of Rights, has noted, "We're having an AI moment. It will not last forever. In this window of public intrigue, anxiety, and scrutiny, there is an unprecedented opportunity for political engagement."

Some, like OpenAI CEO Sam Altman, tell seductive stories promising these powerful new technologies will improve life for everyone, ushering in a world where poverty is eliminated, everyone receives an income, and everyone works less. It is possible that A.I. will spur innovation and productivity on the lines Altman imagines. Its is also possible it will not. But even if A.I. delivers on these promises, that doesn't mean the benefits will be shared widely, nor that the foreseeable or speculative risks – not to mention the unknown risks – will be addressed or alleviated.

We need strong safeguards and government regulation – and we need them in place before corporations disseminate A.I. technology widely. Until then, we need a pause.