epic.org

ELECTRONIC
PRIVACY
INFORMATION
CENTER

MAY 2024

GENERATING HARMS

Generative AI's New & Continued Impacts

CONTRIBUTIONS BY

Grant Fergusson
Sara Geoghegan
Calli Schroeder
Maria Villegas Bravo

EDITED BY

Chris Frascella, Tom McBrien, Calli Schroeder, Maria Villegas Bravo, Kara Williams, and Enid Zhou

Notes on this Paper:

This paper builds on the structure and issues raised by *Generating Harms: Generative AI's Impact & Paths Forward*, our report released in May 2023.

We have expanded into a sequel report to address additional harm areas that have become clearer in the intervening year. As generative AI continues to develop, we anticipate there may be future editions as well.

While we have largely maintained the typology of harms from the first report (modeled from Danielle Citron and Daniel Solove's Typology of Privacy Harms and Joy Buolamwini's Taxonomy of Algorithmic Harms), there are some issue areas that presented specific harm types outside of those categories. We have identified these where necessary.

Table of Contents

Introduction

In the year since EPIC issued the first Generating Harms Report,[1] generative AI has exhibited near-unfettered growth, expanding across every industry and policy discussion worldwide. Though proponents of generative AI continue to tout these systems as having unlimited and world-altering capabilities, many have challenged this narrative. The public is not only aware of the spread of generative AI: They actively distrust and dislike it,[2] especially because of its impact on privacy.[3] Enforcement bodies have held multiple hearings on generative AI and are rapidly moving forward with targeted regulations. Technologists, civil rights leaders, advocates, and other experts continue to draw attention to existing laws and principles that should curb generative AI. In this report, we intend to add to that effort.

One marked shift that we have seen in the past year lies in the conversation about generative AI harms. Despite industry efforts to pivot harms discussions to hypothetical future existential risks of generative AI, the focus is now squarely on the present: Who are these systems harming, what are the harms, and what can we do about it? The effort to distract us with ideas of a dystopian future has not stopped us from keeping up the pressure to change our dystopian present.

Emphasizing the real-life harms perpetuated by generative AI serves several purposes. It raises basic questions about the fitness and safety of the technology, prompting discussion of why we are permitting unproven technology into sensitive areas of our lives. It forces companies developing and selling generative AI to be accountable for the consequences of their technology's largely unchecked spread. It pushes authorities to look at how generative AI harms their constituents and what they can and should do to stop that harm.

This report expands on our previous work by looking at additional areas of generative AI harms and some of the efforts taken to counter them. The areas of harm that we identified in our initial report remain relevant, with multiple examples of both real-life harms and active pushback against generative AI in those spaces. However, additional harm categories have emerged that merit examination. In addition, a year of expanded generative AI development and discussion has produced several proposed and implemented efforts to counter generative AI harms. This report sheds light on both of these aspects of generative AI.

— Calli Schroeder, EPIC Senior Counsel and Global Privacy Counsel

Endangering Elections

BACKGROUND AND RISKS

Through election-related harms, generative AI can alter the futures of nations. Problems like misinformation, disinformation, foreign influence, and security issues are not new. But generative AI supercharges those problems to a degree that presents a genuine threat to democracy.[4] AI can generate convincing deceptive text, audio, and video content that often takes a great deal of time and technical knowledge to debunk. In addition, the volume of election-related harms enabled by generative AI turns problems that were once minor or manageable into devastating attacks that erode trust in elections. Think of it like a building that previously had the occasional rock thrown at it now being hit by a wrecking ball.

Election-related harms can range from degrading trust in elections to manipulating voter behavior to creating security threats. If elections are successfully altered via AI-generated content, the harms expand to empowering parties who otherwise would not have won and altering policy and political decisions for years to come. In this section, we look at three ways generative AI is damaging elections: mis/disinformation, foreign influence, and security infrastructure. Unless these harms are dealt with swiftly, we will see a massive erosion of trust in the security and accuracy of elections, leading to widespread mistrust and societal unrest with potentially catastrophic consequences.

MISINFORMATION AND DISINFORMATION

Political campaigns are often plagued by rumors, scandals, and falsehoods about candidates, campaigns, and political parties—what we categorize as misinformation and disinformation. Typically, misinformation refers to the spread and creation of unintentionally false information while disinformation is purposeful spread of false information. However, the speed at which AI-generated content spreads blurs this distinction, as people share information without first editing or fact-checking it.[5] Whether false information is created and spread purposefully or unknowingly, the impact is the same.

AI-generated mis- and disinformation can be polished and highly convincing, often mixing accurate and inaccurate information. Even when not deliberate or prompted by a user intending to produce disinformation, generative AI systems may spontaneously generate inaccurate information when they are trained on data that is not reviewed for accuracy prior to its addition to the training dataset or where the algorithm draws the wrong conclusions and connections internally. In some cases, the incorrect information may stem from the phrasing of the end user's prompt and be incorporated into the generated content. Generative AI systems have more than enough content in their training databases to produce wildly inaccurate election information, as there have been years of discussion, conspiracies, and deception related to voting machine security, mail voting, "stolen" election narratives, and biographical information and theories about candidates that these systems can draw from.

AI-generated mis- and disinformation can take many forms, with many repercussions. Audio, video, and images may warp public perception of candidates, their policies, their actions, or broad election practices (such as deepfake images of discarded mail ballots), producing scandals based on entirely false information. The volume of generated content may convince voters that there is broad consensus on a political matter or widespread acceptance of a falsified scandal. AI may be combined with targeted influence campaigns using demographic information held by data brokers, allowing highly tailored and manipulative targeting of individuals or groups with similar characteristics. AI-generated content may flood party representatives with comments from fake “constituents” or put out false poll information to push parties to change their political stances in response to the “will of the people.” Voters could be disenfranchised through AI-generated content, spreading false information about where and how to vote, eligibility requirements, or false risks of voting.

Once created, this AI-generated content will spread like wildfire across social media, in chat groups, and possibly even through news outlets. Reports have demonstrated that mis- and disinformation tend to have *more* engagement than factual information.[6] The risk is even higher in areas without reliable local news, where AI-generated content and websites will take advantage of the gap to seed mis- and disinformation.[7]

The impact of mis- and disinformation is particularly damaging when it is released without adequate time to effectively debunk it. For instance, fake audio recordings of a Slovakian party leader discussing how to rig the election was posted on social media just days before the election—that party then lost the election.[8] The challenge of confirming whether content is real cuts both ways. We have already begun to see politicians claim that actual photos, videos, and audio recordings of them were AI generated.[9] When voters have no clear way to determine when information is true or fake, they are much more vulnerable to manipulation, election interference, and other propaganda that may affect how they vote.

In addition to concerns about accuracy around candidates and political stances, AI-generated mis- and disinformation could be used to directly disenfranchise voters. For example, two days before the primary in New Hampshire, robocalls went out using AI-generated audio of President Biden’s voice to urge Democrat voters to “save their votes” for November and not vote in the primary.[10] Similar ploys could tell voters incorrect poll or voting time information or attempt to intimidate voters by claiming voting would put them at risk. For instance, Jacob Wohl and Jack Burkman made robocalls to Black New Yorkers claiming that their personal information would be sent to law enforcement, debt collectors, and other authorities if they voted by mail.[11] Mimicking the voice or image of a trusted figure through generative AI makes these types of attacks even more potentially damaging.

FOREIGN INFLUENCE

Generative AI's ability to produce content that sounds like it comes from a local speaker of the target language is a boon to foreign actors attempting to influence elections. Textual tells, odd ways of forming sentences, or phrasing have often been giveaways that a piece of content may come from a foreign actor. AI services smooth over these signs, producing content that is indistinguishable from how a local would speak or write it. New York City mayor Eric Adams has already employed this tactic, using AI to generate robocall messages of his voice speaking multiple languages.[12] Further, if a foreign actor wants to make audio or video content to make themselves more convincing, generative AI can create speech in a local accent. Because it is very difficult to confirm where content has been AI generated, much less what human individual prompted the creation, authorities have a very difficult time enforcing against foreign actors who would take advantage of this technology.[13]

SECURITY AND PERSONAL SAFETY

Elections are often a time of dramatically increased communications to voters from candidates, political parties, and interest groups, creating an opportunity for bad actors to solicit personal and sensitive information from individuals under the guise of needing it to process voting information or financial support. For example, someone may thank an individual for their “donation” to a cause. When the individual responds that they made no such donation, that person may then request bank information in order to cancel the transaction. Generative AI systems could be used to generate phishing attacks on individuals, or hackers could embed malware in AI-generated content. While these risks have already been addressed in our initial Generated Harms report under “Turbocharging Information Manipulation,” the unique circumstances of elections may make individuals more likely to believe authority figures are contacting them and that they should respond. In addition, generative AI could be used to directly target election officials for phishing or doxing attacks, endangering election security even further.[14]

HARMS

- **Economic/Economic Loss**: Scams, phishing attacks, and malware may result in direct economic loss for individuals through fraudulent donation funds, gaining access to financial accounts, and other instances of manipulation and deception. Successful attacks in this area may also have long-term impacts on credit.
- **Reputational/Relationship/Social Stigmatization**: Mis- and disinformation about individuals connected to elections (candidates, their families, party leaders, etc.) can permanently impact their public image and interpersonal relationships, potentially on multiple levels—local, national, and international.
- **Psychological**: The individuals implicated in mis- and disinformation may face severe emotional harm, shame, and embarrassment due to the spread of false information, recordings, and images. In addition, individuals tricked by scams or mis- and disinformation campaigns that are later debunked may be embarrassed or feel manipulated and used.

- **Autonomy**: The wide spread of mis- and disinformation and the difficulty of confirming the truth in a timely and effective manner impacts individuals' ability to make a properly informed and non-manipulated decision regarding their vote and other actions.
- **Discrimination**: Many of the harms listed above are specifically targeted at marginalized or vulnerable communities.[15]
- **Societal**: The effects of election-specific generative AI harms erode trust in government and basic voting institutions, threatening civic participation and democracy.

EXAMPLES

- Microsoft has already tracked a marked increase in campaigns from both Russia[16] and China[17] to influence U.S. elections using AI-generated content.
- New Hampshire residents were bombarded three days before the primary with robocalls featuring an AI-generated audio mimicking President Biden's voice, instructing them not to vote in the primary.[18]
- Deepfake videos of U.S. politicians have been widely circulated. This includes a video of "President Biden" reinstating the draft,[19] which was viewed more than 8 million times on Twitter, and a video of "Senator Elizabeth Warren" saying Republicans should be barred from voting.[20]
- Internationally, harmful AI-generated content has thus far included a video of Moldova's president supporting a Russia-friendly political party,[21] audio of Slovakia's liberal party leader discussing vote rigging,[22] and videos of female Bangladeshi opposition politicians wearing a bikini and in a swimming pool (offending many who observe the conservative dressing customs of the Muslim-majority country).[23]
- Politicians are now arguing that authentic text, audio, and video footage of them has been AI-generated, including former President

Trump dismissing attack ads featuring videos of public gaffes,[24] a Taiwanese politician denying the accuracy of video of him entering a hotel with a woman,[25] and an Indian politician denouncing audio of him accusing his party of illegally accumulating $3.6 billion.[26]

- Several[27] analyses[28] have demonstrated how social media algorithms amplify mis- and disinformation over factual information.

INTERVENTIONS

- Enact a law that would make generating deliberate mis- or disinformation regarding elections, candidates, or political parties illegal. Some examples of this approach are the proposed Deceptive Practices and Voter Intimidation Prevention Act and South Carolina's H4660.[29]
- Build on the FCC's decision to outlaw robocalls[30] containing AI-generated voices to combat mis- and disinformation campaigns impersonating politicians. Federal agencies could also mandate that political ads either be barred from using AI-generated content or be forced to make clear disclosures where they are used—including information put out by paid influencers.
- Mandate that generative AI companies adopt baseline precautions to counter AI disruption of elections, such as preventing the technology from generating images or impersonations of known political figures.[31]
- Require AI developers to implement filters for known election falsehoods and continuously update those filters as more falsehoods are developed and spread. This includes both filtering outputs of generative AI systems and constantly checking the training datasets to ensure accuracy.
- Require social media companies to proactively develop frameworks to counter AI-generated election risks. Approaches may include

modifying their algorithms to ensure mis- and disinformation are not promoted or enabling a trust verification on fact-checked information. Some technology companies have already signed on to a voluntary framework as a first step.[32]

Eroding Privacy

BACKGROUND AND RISKS

While several harm categories in our initial report mentioned privacy as a risk factor, none examined how existing generative AI systems and practices damage individual and societal privacy. Generative AI consumes vast amounts of data, from data used to create training datasets, to data created while monitoring user inputs, to the platform's own pattern forming and outputs. Unless carefully and consistently curated, that data will include personal data, sensitive data, and inferred information about individuals. People whose data is processed by generative AI may be unable to get the data removed and, in many cases, may be entirely unaware that their personal data is being used by the system at all. In this section, we look at three key areas of privacy harm: maximalist data use, scraping to train data, and data security issues.

MAXIMALIST DATA USE

Generative AI is built on a data maximalist approach: AI developers are incentivized to collect more and more data to train models. Often, developers will use methods like data scraping to indiscriminately collect information from the internet to feed and train their models with little regard to the quality or accuracy of the data. The National Institute of Standards and Technology ("NIST") has explained:

> The performance of GenAI text-to-image and language models scales with model size and dataset size and quality. For example, scaling laws indicate that training a 500 billion parameter models would require 11 trillion tokens of training data. Thus, it has

> become common for GenAI foundation model developers to scrape data from a wider range of uncurated sources.[33]

Some developers may use filters or parameters to attempt to sanitize or clean the data, but the industry lacks meaningful oversight or requirements to ensure this.

This approach exposes individuals to exorbitant privacy risks and violations. Uncontrolled data scraping will pick up every piece of personal information that can be found online. This includes items that a person may consider public or benign (name, employer, age), more sensitive data that a person may want more tightly protected (address, relationships, location), and highly sensitive data that could expose or endanger an individual (sexuality, health information, religion, political affiliation). It may also include false or illegal personal data, such as revenge porn images or deepfakes of the individual. Most data scraping does not have meaningful checks in place to ensure that the collected data is accurate, high-quality, or even legal (as has been demonstrated by high-profile investigations showing child sexual abuse material ("CSAM") in several datasets).[34] Processing these massive amounts of data also take a significant environmental toll due to the energy and water required for processing data and cooling components.

Not only is indiscriminate scraping potentially harmful on an individual level, the practice also contradicts fundamental data privacy principles like data minimization and purpose limitation. Data minimization requires anyone using personal data to limit its collection, use, disclosure, processing, and retention to only the information that is reasonably necessary to furnish the product or service requested. Purpose limitations, a necessary part of a strong data minimization rule, require that the data only be used for the purpose for which the data was collected. For example, if someone provides their email address to verify their account when making a purchase online, they would likely expect to receive email communications about the status of their order, since that is connected to the service they have requested. However, if they were to post a status and a photo of their children on social media, they would not expect that the photo and text would be scraped by an AI developer and used to train an AI system. Not only is that use wholly unrelated to the purpose of the post, it involves a company they have never willingly interacted with taking and using personal data with no notice to the individual or opportunity for them to refuse. Similarly, a person might interact with a chatbot on a website seeking medical or mental health advice. That user would not expect that their sensitive health information could be used to train the bot and this purpose would contradict assumptions of confidentiality and discretion around this sensitive information.

Generative AI practices contradict several other generally recognized data privacy principles like accountability, accuracy, and transparency. When developers indiscriminately scrape publicly available websites for data to train their models, they may have no actual knowledge of the quality or legality of the data collected. Scraping lacks meaningful oversight to ensure that the data is accurate or without bias. Data collection and training dataset creation practices for generative AI are largely opaque, happening in black boxes without transparency. Without universal requirements to limit this, developers will build, train, and deploy a generative AI system without any independent review, such as third-party audits to test for accuracy and bias or independent privacy impact assessments. California has attempted to correct this with its newest regulations, but the impacts will remain to be seen because the rules have not yet gone into effect.

SCRAPING TO TRAIN DATA

Generative AI systems are often built using training datasets comprising data scraped indiscriminately from the web. That means the data used to feed generative AI systems can be any information that is publicly available online. Most companies developing these models do not release detailed information about the data sets they have used, “but these data sets inevitably include some sensitive personal information, such as addresses, phone numbers, and email addresses.”[35] This may also include data released through breach, illegally shared information like CSAM or revenge porn, or information revealed through doxing. Training datasets have already been shown to include confidential and privileged information that should not have been publicly available, such as private medical record photos.[36]

Often, generative AI systems lack filters to prevent this type of data from being used to build or train models. Foundation models make “heavy use of unsupervised learning” during the pre-training stage when they are created.[37] This means information that was never meant to be public could be used to train a system, which poses serious threats to privacy, security, and accuracy. When a generative AI system inputs this type of information, it may generate a result that includes the very information that fed the system. For example, data from a breach that reveals a consumer’s social security number could be scraped and used in a generative AI system that generates a response that includes the consumer’s social security number. Images that were scraped and fed into the system may generate substantially similar images, revealing intimate details about the person in the scraped image. The lack of oversight and limitations on the collection of publicly available personal information poses a serious privacy threat because sensitive personal information may be revealed in a generative AI system’s output.

The European Union has begun to break ground on regulating the use of web scraping because of its threat to the fundamental rights to privacy and data protection. For example, the Dutch Data Protection Authority (“DPA”) recently published guidelines discussing web scraping.[38] The DPA concluded that web scrapers and web crawlers almost inevitably capture personal data, including special categories of data because of the sheer breadth of data on the internet (including data leaked from data breaches)[39] In most cases, this data scraping violates European law because there is no lawful basis of processing of the data, nor is there notification to the data subjects that their data has been processed.[40] The Artificial Intelligence Act ("AI Act" or "EU AI Act"), set to come into force in 2026, also strictly prohibits algorithms that scrape images off the internet or CCTV footage for the purpose of creating or expanding facial recognition databases.[41]

Finally, the vast amount of data consumed by these systems can lead to extremely revealing inferences. Generative AI systems are built to detect patterns in how information is connected and constructed, from basic sentence structure to much more revealing information. Once those patterns are built into a system, it may solidify inferred information that may be highly sensitive, like analyzing communication patterns to reveal romantic relationships or tracking movement patterns that would reveal an individual's home address.

DATA SECURITY ISSUES

The excessive collection, use, and retention of personal information in generative AI systems makes them ripe for data security issues. As explained previously, generative AI systems rely on massive amounts of data that are retained indefinitely. This includes both the data collected for training datasets and data input into the AI system by users. Security concerns with this second data collection avenue has prompted bans from scores of federal agencies[42] and private businesses, particularly across medicine, finance, journalism, security, and other industries dealing with sensitive confidentiality requirements.[43] The persistent threat of a breach increases when data is excessively collected or retained; data that is deleted after it is no longer needed cannot be subject to breach. These indefinitely retained troves of personal information are vulnerable to attacks by bad actors, either through breach or adversarial machine-learning techniques. One such technique is prompt injection, which can get an AI system to reveal its raw training data, exposing any personal data contained in the dataset. There are two types of prompt injection attacks: direct and indirect. A direct prompt injection happens when an actor inputs text intending to alter the behavior of the Large Language Model (“LLM”).[44] An indirect prompt injection involves an attacker that manipulates the data used in a LLM to remotely inject system prompts without directly interacting with the application.[45] One example of an indirect prompt injection is when an actor changes information on a webpage that the LLM will read when generating a prompt, manipulating the outcome. NIST explained how many parts in the process can be susceptible:

> These security and privacy challenges include the potential for adversarial manipulation of training data, adversarial exploitation of model vulnerabilities to adversely affect the performance of the AI system, and even malicious manipulations, modifications or mere interaction with models to exfiltrate sensitive information about people represented in the data, about the model itself, or proprietary enterprise data.[46]

These attacks may allow either direct access to raw data contained in the datasets, which will almost certainly include personal data, or may basically “short circuit” the generative AI system to provide pieces of personal data in its outputs.[47] Generative AI provides more opportunities for bad actors to manipulate data and prompts to manipulate outputs. Without meaningful limitations, these opportunities threaten the security of the data used to build and train a system.

HARMS

- **Physical**: Generative AI systems may reveal data that could put a person at risk, whether real or perceived. For example, sensitive information, such as a person's email address or home address may be revealed to stalkers, abusers, or other bad actors.
- **Reputational/Relationship/Social Stigmatization**: Generative AI can reveal a person's sensitive information, which may result in damage to reputation or social stigmatization. A person's sexuality may be inferred where sexual images, information related to sexual behaviors, or location information were fed to an LLM.
- **Economic**: Businesses whose trade secrets have been incorporated into training sets or individuals whose economic information has been incorporated face potential economic injuries.
- **Psychological**: Individuals may suffer from anxiety or fear due to the lack of control over removing their personal data from training sets and may fear consequences from the ways in which the data is used. People may also be angry, frustrated, or feel exploited that their information has been used to feed a for-profit LLM, even where the data is anonymized.
- **Autonomy**: Individuals cannot control the collection and use of their personal information, including whether it is used to train datasets.
- **Discrimination**: Biased data can be scraped and fed into an LLM, which will likely then produce biased results built from historic discrimination.

EXAMPLES

- ChatGPT falsely accused a law professor of sexual harassment by including his name on a generated list of professors that had sexually harassed someone, citing a non-existent news article.[48]
- Researchers at Google forced ChatGPT to reveal some of its training data, revealing an email address and phone number. The researchers said that when instructed to repeat a single word forever, the bot often released personal information and raw training data.[49]
- A recent report[50] found that "an organization can expect around 660 daily prompts to ChatGPT for every 10,000 users, with source code being the most frequently exposed type of sensitive data, posted by 22 out of 10,000 enterprise users and generating, on average, 158 incidents monthly. This is ahead of regulated data (on average, 18 incidents), intellectual property (on average, four incidents), and posts containing passwords and keys (on average, four incidents) every month."[51]
- A ChatGPT vulnerability may have revealed users' payment-related information and titles from users' chat histories.[52]
- Microsoft, a strong proponent of generative AI and the developer of one of the most popular generative AI systems, accidentally incorporated two employees' back up computers—including passwords, encryption keys, and Teams threads—into its training data.[53]

INTERVENTIONS

- Enact a general, comprehensive federal privacy law that limits the collection, use, and retention of personal information to that which is reasonably necessary to fulfill the service requested. This will include purpose limitations to ensure that personal information is not used in an out-of-context way to train generative AI systems.

- NIST has suggested several mitigation techniques to minimize harms from prompt injections, while explaining that these measures do not provide full immunity to all attacker techniques: training for alignment, prompt instruction and formatting techniques, detection techniques, reinforcement learning from human feedback, filtering retried inputs, an LLM moderator, interpretability-based solutions.[54]
- Enforce laws that prohibit unfair and deceptive trade practices, mandate consent requirements for child users, and require justification for data processing.
- Build support tools by only using a limited and disclosed set of data.
- Adopt a strict data minimization standard by developers to help mitigate the privacy harms of creating, tweaking, and updating models to train AI. Data minimization is a standard that, depending on the precise definition, should only allow collection of personal data to the extent that it is necessary to carry out the service requested by the user. The tenets of data minimization are fundamentally at odds with the large-scale creation of generative AI datasets from public info without disclosure or consent.

Data Degradation

BACKGROUND AND RISKS

Generative AI companies have flooded the digital public square with content of varying quality that has demonstrably worsened the experience of daily internet use. The influx of synthetic content has fundamentally stripped the utility from the internet as a service, leaving end users locked in to an ever-worsening system. This phenomenon mirrors what Cory Doctorow describes as "enshittification," which we refer to as "data degradation" in this context.[55] The data degradation process is gradual with consistent, identifiable steps. First, generative AI companies hook businesses, governments, and users on the idea that they will make life better and easier by performing some simple tasks (like text generation) well and claiming that this success can extend to unlimited other use cases. Second, the companies abuse the primary creators and audiences of synthetic content to provide value to business clients—in this case, by flooding search engine results to brute force search engine optimization and game social media algorithms. Third, and finally, the AI company will claw back more value for themselves by cutting costs, downgrading the quality of their offerings and exploiting the business clients. Software that was once open source will be taken off the market and deemed proprietary;[56] previously curated datasets will be taken over by datasets built by web scrapers. This process ends in a flood of inaccurate, non-sensical, and low value synthetic content that takes over the digital ecosystem so completely that it is no longer possible to meaningfully sort the good from the bad.

This data degradation problem is two-pronged, partly because the quality of AI-generated content varies wildly. On one hand, much of the content being generated is low quality—incoherent sentence structures, inaccurate data, discriminatory outputs, and a penchant for hallucinations.[57] Generating low quality content creates a feedback loop where the system that scrapes the internet and its own outputs for new training data ingests the low quality content, leading to eventual model decay and continually worsening output quality.[58] On the other hand, synthetic content that is indistinguishable from authentic content is distorting reality and overwhelming human-generated content. Convincing and higher-quality outputs make humans unable to distinguish between authentic content and generated content, which not only erodes trust but also means people will unknowingly amplify AI generated content to the detriment of human created content. This lack of trust and ability to differentiate content carries over into settings that require objective, authentic evidence, like trials, mediations, election campaigns, and more. Every corner of the internet, and beyond,[59] has been infiltrated by generative AI to the detriment of everyday people.

A FULL-SCALE INVASION

The actual scale of the generative AI problem is impossible to quantify, but more and more of the internet is being taken over by synthetic content. Generative AI developers and deployers are seeing unprecedented growth. In February 2023, ChatGPT boasted 100 million monthly users after only two months in business—a feat that took Facebook almost four and a half years after launch.[60] Nine months later, ChatGPT reported over 100 million weekly users after only a year in business.[61] Over two million developers have taken advantage of OpenAI's model with ChatGPT and Whisper's (a speech recognition tool) application programming interfaces ("API"), which allow entities to use and iterate on software for the entity's own use.[62] In total, the top 50 generative AI models received over 24 *billion* site views in the course of 12 months.[63] Retail platforms like Amazon[64] and Etsy[65] are being flooded by AI generated listings, and generative AI is creating full websites in seconds.[66]

Generative AI companies are not the only culprits. Platforms like social media websites and search engines are boosting synthetic content across the internet. For example, AI generated news articles are topping search engine results, often beating out authentic news outlets, including those whose content was scraped to create the synthetic article.[67] In a preprint study by the Stanford Internet Observatory, researchers found that Facebook's recommendation algorithm pushes AI generated content because AI-generated content appears to generate more engagement.[68] In fact, one of the most viewed pieces of content on Facebook in Q3 of 2023 (boasting 40 million views and almost 2 million interactions) was an AI generated image.[69]

One reason that the actual scale of the synthetic content issue is hard to calculate is that there are currently no consistent requirements on generative AI companies or platforms to label or otherwise distinguish synthetic content from authentic content. In the United States, many of the proposals around labelling AI-generated content are tied to election-related content.[70] In addition, NIST[71] and several state legislatures have proposed more general watermarking requirements, but none have been enacted as of the release of this publication.[72] The European Union's AI Act requires any fully synthetic or partially manipulated content to be labelled and obvious to the user as AI-generated before or during interaction with the content.[73] However, the AI Act will not come into force until 2026. The United Kingdom Information Commissioner's Office is also exploring possible watermarked or other labeling requirements for generative AI content.[74]

Watermarking may be a helpful first step, but cannot be a full solution to the problem of identifying AI content since it does not address audio or text based synthetic content and can likely be copied, manipulated, removed, or otherwise defeated.[75] Researchers at the University of Maryland have already found ways to break watermarking methods and insert false watermarks onto images.[76] Few other options for identifying generated content have been put forth, so watermarking remains the most commonly proposed labelling method. Other methods of checking for synthetic material (such as counting fingers or checking image consistency) are already losing utility.[77] Some companies advertise services where AI algorithms detect synthetic content,[78] but this technology is still in its infancy and is easy to fool. Even OpenAI threw in the towel and decommissioned its own synthetic content detection software.[79] Without methods to reliably identify synthetic content, the full scale of the issue remains unknown and the true extent of the harms stemming from generative AI remains unaddressed.

AN OUROBOROS OF DATA DEGRADATION

LLMs and other generative AI algorithms continuously evolve by incorporating new data into the model's training set and "learning." To satisfy the enormous quantity of data required to build and maintain training datasets, AI companies use web crawlers to scrape images, text, and all other types of data from the internet.[80] This data is incorporated into the training dataset and the algorithm "learns" common patterns, data structure, type, classification, and other categories from it. The patterns and information "learned" from the constantly-updated training data in turn show up in the generated outputs. While the data collection method itself comes with several issues, this section of the report will focus on model collapse.

The increasing volume of AI generated content means more and more of the data collected by web scrapers will be from AI feeding AI. Humans may be able to distinguish where content is clearly AI generated, obviously incorrect, or incomprehensible. However, web scrapers have no such discernment. Accuracy and quality issues in training datasets could be addressed by using smaller amounts of carefully curated and fact-checked data or employing human review before data is added to a training dataset; however, these responsible practices take resources that generative AI developers thus far seem unwilling to expend.

Where scraped data is automatically put into training datasets with no quality checks or human oversight, the models deteriorate rapidly until all new outputs mirror the incomprehensible, inaccurate, and hallucinatory inputs.[81] This deterioration creates a never-ending spiral of ever worsening inputs and outputs until the AI model is entirely useless, and the ouroboros of data degradation is complete.

REALITY DISTORTION

The more common AI generated content becomes, the harder it is to discern what is authentic and what is synthetic content. With images, synthetic content is often identified once viewers count fingers, look at ears and other complex body parts, and use the relative scale of objects to discern authenticity. Models like DALL-E initially struggled to generate these intricate, complex features, giving synthetic images a number of "tells." However, these synthetic content giveaways are quickly disappearing. Generative AI systems are consuming exponential amounts of content, including more and more synthetic content as the volume generated grows. This is leading to a situation where generative AI systems either quickly spiral into model collapse or begin to produce much higher quality outputs that are increasingly difficult to distinguish from real images.

Furthermore, the increase in volume of generated content has not prompted a similar increase in the public's ability or inclination to verify every piece of content they see. The lack of information verification means false information spreads quickly and easily, while true information can be dismissed as AI generated.[82] This lack of certainty creates a warped reality where no one is fully confident in fundamental facts and news, experts, and even our own eyes and ears can no longer be trusted.

The low quality of synthetic content ruins institutions that rely on truthful, authentic information like legal investigations, academic research, and journalism. If a bad actor sends the Federal Bureau of Investigation a deepfake of a person saying they are going to bomb a bank, the FBI will waste valuable resources evaluating the veracity of the bomb threat and could begin to erroneously investigate a person who was impersonated by the bad actor. This ability to impersonate other people and recreate evidence in painstaking detail creates reasonable doubt in a jury's mind, whether the evidence itself is authentic or synthetic. Humans also over rely on visual evidence, leaving juries more likely to believe visual synthetic content such as deepfakes. Even for non-legal contexts, like journalism, synthetic content impedes the spread of truth and creates doubt and confusion for the investigative process. Finally, many types of academic research rely heavily on data pulled from the internet for sources, analysis, and more. If there are hallucinations and an increased volume of inaccurate information, the data found by researchers is rendered useless because it introduces noise and skews results.

HARMS

- **Economic/Loss of Opportunity**: Journalists, musicians, actors, artists, and other humans engaged in intellectual property creation are losing employment opportunities to generative AI. The ever-lowering quality of outputs also causes economic harms to the companies using these systems as they must correct the unusable outputs.
- **Psychological**: The volume of AI generated content and its widely varied quality makes individuals less able to distinguish between authentic content and GAI content. This causes frustration and

helplessness when trying to ascertain the veracity of online content such as videos of political candidates or divisive news stories.

- **Reputation**: The continued degradation of both training datasets and generative AI outputs leads to more and more inaccuracy and hallucinations that incorporate real individuals' data. This wrongly links people to damaging acts and behaviors in AI generated content and adequately countering this false information is nearly impossible.
- **Environmental**: The volume of electricity and water for cooling required to run generative AI systems will only increase as they are made to process more and more data, in training and in outputs.
- **Autonomy/Loss of Consumer Choice**: Because of the network effects of the internet and the sheer scale of generative AI garbage created, individuals cannot reasonably avoid generative AI content nor systems. End users will eventually be left with virtually no alternatives untouched by generative AI if this continues.
- **Autonomy/Loss of Service Quality**: The degrading quality of generative AI content means that the overall quality of online platforms and services is falling as well.
- **Autonomy/Behavior Manipulation**: Algorithmic feeds, buzzword-laden headlines generated by AI, and other generative AI integrations actively effect what information individuals can find, allowing the companies employing generative AI to manipulate human behavior[83] by distorting the lens through which they take in information.
- **Statutory Harms/Constitutional**: The inability to discern between authentic content and synthetic content leads to the degradation of the criminal justice system by calling into question the veracity of every piece of evidence. Evidence presented at trial or used as "probable cause" for investigations will be continuously called into

question and authenticating that evidence will take significant time and resources.

- **Societal**: The inability to differentiate real from generated information leads to expanding mistrust of any evidence, calling into question historic events, current events, and our own perception.

EXAMPLES

- Google search results have become so clogged with AI spam that they have had to modify the ranking systems specifically to downrank generative AI content.[84]
- AI generated “obituary spam” has become a chronic problem, flooding the internet with strange obituaries filled with search keywords and often entirely incorrect. In some cases, the obituaries were generated for still-living people, causing panic with misinformation.[85]
- Studies on model collapse caused by AI generated content in training datasets reveal “irreversible defects in the resulting models” across all forms of generative models.[86]
- Low-quality scam books generated by AI and capitalizing on newly released human-penned books are flooding Amazon and confusing consumers into incorrect purchases.[87]
- AI-generated news and information sites are flooding the internet with inaccurate information with no human oversight.[88]
- Many newsrooms have shifted to using AI-generated content that is often full of errors, partly or wholly plagiarized, and badly written.[89]
- In several instances, generative AI has hallucinated entire court cases that people then cited in court briefs without reviewing whether those cases existed.[90]

INTERVENTIONS

- Enforce existing consumer protection and product safety laws on generative AI systems and their outputs. This includes, but is not limited to, product liability standards, defamation laws, intellectual property rights, and criminal enforcement for non-consensual deepfakes.
- Institute synthetic content identification requirements mandating that generative AI models label content as fully synthetic or otherwise manipulated to end users. Several regulations at the state, federal, and international level have included watermarking requirements to address the identification problem. However, watermarking is an incomplete and flawed solution that is not viable for audio and text-based content. Furthermore, researchers have already found ways to remove, copy, or otherwise defeat watermarking software like Google's SynthID. Regulators must engage in a multifactor approach to identify synthetic content.[91]
- Require generative AI companies to regularly review and curate the data being added to their training datasets. One way to enforce this would be to remind generative AI companies that they have strict liability for any illegal content that could be scraped and added to datasets, like CSAM. The prospect of prison time for irresponsible data scraping practices may incentivize better practices.

AI Content Licensing

BACKGROUND AND RISKS

Since our first generative AI report, *Generating Harms*, the debate over AI copyright and the online intellectual property landscape has reached a fever pitch. While some lawsuits had already emerged in 2022—mainly focused on breaches of contract[92]—2023 and 2024 saw an explosion of generative AI lawsuits filed by artists and content creators alleging copyright and privacy violations.[93] Most targeted the largest AI developer in the market: OpenAI.

While the core practice of training generative AI models like GPT-4, Midjourney, and Sora on unlicensed, copyrighted works remains unchanged, the prominence of AI-generated content skyrocketed. Plans to displace actors, writers, and other entertainment workers with AI sparked major labor strikes from the Writers Guild of America ("WGA") and the Screen Actors Guild-American Federation of Television and Radio Artists ("SAG-AFTRA").[94] Even when AI-powered chatbots spout misinformation, companies and the public sector have widely adopted these chatbots.[95] And across the web, AI-generated content is rapidly becoming the most common form of content we see.[96]

Amidst this changing AI landscape, we are returning to a topic we recommended in last year's report: content licensing. As we described in *Generating Harms*, AI developers rarely, if ever, have permission to use copyrighted content to train their AI models. Today, unlicensed web scraping remains common.[97] Given these trends, one would think formal content licensing agreements would be strong contractual protections for content creators. However, recent examples of AI content licensing agreements between AI developers and companies like Shutterstock,[98] the Associated Press,[99] and Reddit[100] suggest that AI content licensing may raise new concerns that undermine the benefits they promise.

There are three main issues emerging from today's AI content licensing practices: (1) restricting competition, (2) avoiding judicial review, and (3) exploiting the division between content creation and content ownership.

RESTRICTING COMPETITION

Most major AI developers have already scraped the web for publicly accessible content, but the value of web scraping has gone down tremendously since generative AI tools were first released due to a curious feature of generative AI development: generative AI models appear to collapse when trained on AI-generated content, generating less and less accurate content (as discussed in our "data degradation" section above).[101] To increase AI model accuracy today, AI developers are competing for the last vestiges of purely human-developed content online.

While licensing agreements have been proposed as a way to make sure content creators have adequate control over this new AI training paradigm, *exclusive* content licensing arrangements can be used as a tool to further entrench major AI developers. For example, if a major content provider signed an exclusive licensing deal with an AI developer like OpenAI, that provider's content—news articles, images, videos, etc.—would be unavailable to other competitors unless they risk unlicensed web scraping themselves. Even if competitors did try to scrape or otherwise access the licensed content, major AI developers with exclusive licenses would be incentivized to enforce their exclusive licenses against competitors. The result: smaller AI developers will not be able to compete against larger AI developers, creating an anticompetitive environment that may increase consumer costs and decrease the quality of generative AI products and services.

AVOIDING JUDICIAL REVIEW

Thus far, major AI developers have raked in billions of investment dollars by training their AI models on content scraped indiscriminately from the public internet.[102] Until recently, the practice of training AI systems via web scraping has remained a legal gray area. It is still being actively debated whether AI developers are protected by fair use, an exception to copyright protections, or liable for infringing copyright law.

As of this writing, there are over a dozen copyright lawsuits against AI developers like OpenAI, each arguing, among other things, that the core business model behind generative AI development is illegal under copyright law.[103] A single court ruling that AI developers violated the law could undermine core assumptions about the future of generative AI; without easy access to large amounts of data, current generative AI training practices are unsustainable.

Given the current litigation landscape around generative AI, content licensing agreements implicate the same risk as overreliance on judicial settlements: AI developers may pursue licensing as a way to foreclose active or potential litigation before a court ruling comes down. While content licensing agreements are a net positive outcome for the involved copyright owners, avoiding court rulings means that web scraping and other unlicensed uses of copyrighted material by AI systems will remain broadly unsettled law.

EXPLOITING THE DIVISION OF CONTENT CREATION AND OWNERSHIP EXPLOITATION

Lastly, the recent Reddit-Google AI licensing agreement raises the specter of one final risk: those profiting from AI content licensing may not be the content creators themselves.[104] Across the internet, millions of people post text, images, audio, and video on online platforms like Reddit, Instagram, TikTok, and the platform formerly known as Twitter every day.[105] But most platforms include broad licensing arrangements in their Terms of Service where any user content posted on a platform is available for licensing without the content creator's express consent.[106] Because of these broad licensing arrangements, Reddit and other online platforms have full rights to license what users post to AI developers without users having a say. Even worse, artists may have their work posted on sites like Reddit by *other* users. A content creator who actively avoids online platforms or rejects the licensing arrangements in these platforms' Terms of Service may still be impacted.

HARMS

- **Economic Loss/Competition**: Exclusive and otherwise anticompetitive AI content licensing arrangements will lead to anticompetitive effects in the AI market, reducing the quality of products and services while increasing costs.
- **Economic Loss**: Artists who have provided their content to an aggregator like Shutterstock—or had their content posted on an online platform like Reddit—without any conditions on that content being licensed to AI developers, will not receive any money from lucrative AI content licensing deals the platforms are engaging in.

- **Economic Loss/Demand**: Artists who manage to avoid their content being licensed by others may see decreased demand for their work, as AI models continue to reproduce art in their style at low cost.
- **Reputational:** As generative AI models become more sophisticated, consumers will find it more difficult to differentiate between an artist's real work and AI-generated content designed to look similar to their distinctive styles. On the flip side, legitimate artists may be wrongfully accused of using generative AI models, such that demand for their work decreases.
- **Psychological:** Many artists have expressed pain, sadness, and anger about their creative works being used to train AI models without their consent or knowledge.
- **Autonomy/Lack of Control**: Under many online platforms' Terms of Service, an artists' work may still be licensed to AI developers when posted by users other than the artist, even if the artist otherwise avoids AI content licensing.
- **Behavior**: Facing sophisticated generative AI tools and few avenues for lucrative content licensing arrangements, fewer artists and other content creators will continue to invest the time and resources into creating and publicly sharing non-AI content.[107] To a lesser extent, the threat of AI being trained on social media posts may chill users' behavior online as well.
- **Autonomy**: For artists and other content creators incorporated—consensually or not—into AI content licensing agreements, the terms of such agreements will restrict what legal recourse artists can pursue against AI developers themselves and what options content creators have to protect their own work from new AI developers and others interested in mimicking their work.

EXAMPLES

- Artists have found their works or artistic style explicitly recreated in major generative AI models.[108]
- Reddit and Google have signed a $60 million licensing deal for Reddit user content. Because Reddit's terms of service require users to give Reddit broad licensing rights, users will have no control or compensation from the deal.[109]
- OpenAI has been negotiating content licensing agreements with news organizations even as it faces a series of copyright lawsuits from the industry—sometimes even with the very plaintiffs suing them.[110]

INTERVENTIONS

- Prohibit AI developers from entering into exclusive AI content licensing agreements, such that competing developers would be unable to compete.
- Prohibit online content platforms like Reddit, Instagram, and X/Twitter from licensing user-generated content without explicit, affirmative user consent and/or sharing profits with users.[111]
- Require AI developers to acquire the explicit, affirmative consent of artists before permitting their AI models to mimic and/or wholly recreate an artist's style.
- Content creators can use technological tools like Glaze[112] and Nightshade[113] to disrupt AI developers' ability to train models on the content—tools that may be removed when a content creator explicitly agrees to license their content as AI training data.
- Corporate owners of content licensing rights can implement an opt-in process—or less optimally, an opt-out process—to give creators

control over whether their content is included in AI content licensing agreements.

- While an imperfect solution, require AI developers to add labels or watermarks to AI-generated content to reduce the likelihood that laypeople misconstrue AI-generated content as real or vice versa.[114]

Remedies and Enforcement

Since the *Generating Harms* report, we have seen countless harms come to fruition, including scams involving fraudulent firms sending fake Digital Millennium Copyright Act ("DMCA") notices,[115] the viral spread of AI-generated war propaganda,[116] embedded racial bias in AI resume scanners,[117] AI-generated child pornography,[118] using AI to identify military targets,[119] and more.[120] In the wake of those harms and increasing calls to counter them before generative AI becomes so ingrained in society that it cannot be slowed, a slew of proposals for how to combat generative AI harms have emerged. We want to draw attention to these efforts and what more can be done. The proposals and actions we have seen thus far typically fall into one of three categories: enforcement, regulations and policies, and technical and private-sector remedies.

ENFORCEMENT

Generative AI is a multi-faceted technology with several use and risk areas, making it challenging to regulate absent a high level, AI specific regulation. However, the novelty of the technology does not mean it is exempt from existing consumer protection, civil rights, and product safety laws. Many enforcement bodies have been using their authority to address generative AI harms that fall within their jurisdiction. In the absence of a federal regulation directly addressing generative AI harms, we anticipate continued efforts to enforce existing regulations applicable to these harms.

The Federal Trade Commission ("FTC") has used its authority to launch an inquiry into five companies providing generative AI services, looking to ensure there are not competition and antitrust violations taking place.[121] The FTC has also finalized a rule barring AI impersonations of governments or businesses,[122] proposed expanding the impersonation fraud rule to address individuals and tech companies engaged in AI deepfakes and voice cloning,[123] and launched an inquiry into Reddit's deals licensing data to AI companies for model training.[124]

The Federal Communications Commission ("FCC") issued a Declaratory Ruling confirming that calls made with AI-generated deepfake voices are illegal under the Telephone Consumer Protection Act.[125] The U.S. Copyright Office repeatedly denied registration of AI-generated artwork, explaining in two[126] decisions[127] and in a subsequent statement of policy[128] that only content created by "creative contribution from a human actor" could be granted copyright registration.

Finally, some private companies and individuals have taken enforcement into their own hands. Several media outlets, including The New York Times,[129] The Intercept, and more, have sued generative AI companies for copyright violations.[130] Getty Images,[131] Universal Music Group,[132] various authors,[133] and several artists[134] are doing the same. These cases are not yet decided, but the snowball effect of lawsuits indicates both immense dissatisfaction with the technology and the lack of clarity around how existing laws apply.

REGULATION AND POLICIES

FEDERAL

While several federal AI laws have been proposed (20 in the most recent legislative session alone),[135] the U.S. does not currently have a federal law on AI. Because the U.S. also lacks a comprehensive federal privacy law, even the most basic protections regarding AI system use of personal data are absent. Though the FTC,[136] FCC,[137] Consumer Financial Protection

Bureau,[138] and other agencies have made attempts to fill gaps by applying various consumer protection rules to AI, these efforts are no substitute for a law designed specifically to address AI's many harms.

It is always a challenge to regulate technology where the full scope of use and harms is yet unclear. However, building a framework of consistent and comprehensive protections and rights would not only address the issues we see in AI, but put in place guardrails for future technologies. For example, the EU was able to pass its AI Act fairly quickly by building on established rights, principles, and frameworks within its General Data Protection Regulation.[139] The U.S. has actively proposed many privacy bills—55 in the 118th Congress alone—including broad privacy bills and bills specific to health, financial, children's, biometric, and other privacy matters.[140] However, until privacy or AI laws are actually passed, the protections in this area remain a patchwork.

While not a regulation, the Executive Order "Safe, Secure, and Trustworthy Development and use of Artificial Intelligence" contains a number of measures to regulate federal use of AI technology, directs the Attorney General to mitigate algorithmic discrimination and other civil rights violations tied to AI, provide guidance on AI use, require AI developers to share information on training and safety tests with the government, and promote privacy rights.[141] The Executive Order also includes measures on hiring AI experts and training existing employees on AI issues, developing guidance on technical AI labeling and content authentication, setting testing and transparency standards, and more.

The Executive Order establishes many practical and responsible requirements around AI that should serve as a baseline for the AI industry and any users of AI systems: use policies, regular testing and audits, transparency around process and training, and ensuring compliance with existing regulations on consumer protection, federal tool standards, and civil rights. Though it is not as stable as a regulation—after all, Executive Orders may be overturned by later administrations. Hopefully it lays a foundation of knowledge and practice that can be built upon in later policy.

STATE

Several states have put forth regulations that would either directly address or tangentially extend to AI systems (generally through comprehensive consumer privacy laws). Of the six AI-related laws that went into effect in 2023, five were comprehensive privacy laws and one was an AI-specific regulation addressing use of AI in hiring.[142] Comprehensive privacy laws often impact how AI systems function, affecting how AI systems collect, share, and use personal data, providing individual rights, like opting out of automated processing decisions or profiling, or mandating regular audits and assessments. AI companies often behave as if the technology is somehow beyond or exempt from existing consumer protection and other laws. This attitude has permeated AI company actions to the extent that Attorneys General have had to explicitly state that AI systems must comply with existing laws.[143]

States have already enacted several laws that directly address AI harms. While many of the broad state privacy laws also affect AI systems, some states have pushed for much more AI-specific regulation. Utah's Artificial Intelligence Policy Act mandates clear and conspicuous disclosures where a person is interacting with an AI system rather than a human and creates the Office of Artificial Intelligence Policy to implement further rules on AI.[144] New York, Maryland, and Illinois have laws regarding use of AI in employment interviews.[145] California has a similar law mandating disclosure when a person is interacting with a bot to incentivize a sale or transaction or influence a vote.[146] Other enacted state laws address use of algorithms and predictive models for insurance,[147] the use of deepfakes in elections,[148] and use of assessment mechanisms related to prescriptions for vision correction.[149]

The many proposed state laws (54 that are currently active as of this drafting) address AI risks including disclosures on AI training datasets,[150] AI in elections,[151] an AI Bill of Rights,[152] AI use in publishing,[153] and many, many more.[154] The state law process tends to be more agile than federal regulation-making and so may be able to address specific AI harms more quickly. However, we do not yet know how many of the proposed state laws will pass and, even where they are passed, they will only protect residents of the applicable state, not individuals across America.

INTERNATIONAL

Similar to what we have seen in the U.S., the international approach to AI is varied. Some countries have doubled down on existing regulations that touch on data use and consumer safety, some have adopted a "wait and see" approach, some have issued vague policy statements, and some have moved forward with regulation.[155] Brazil, Canada, and China all have draft laws in ready-to-pass state, building on years of work and existing privacy and data regulations within the countries (noting that China also has already-enacted regulations on recommendation algorithms,[156] deepfakes,[157] and generative AI in particular[158]).[159] The EU's Artificial Intelligence Act, passed in March 2024, looks at AI systems through a product safety lens, categorizing systems into prohibited, high-risk, and low or no-risk systems with varying requirements according to the risk level.[160] The Act is broad, covering all forms of AI and adding in some measures related to generative AI late in the debate process.[161] It remains to be seen whether others will adopt a risk-based approach in proposed regulations or attempt different structures.

TECHNICAL AND PRIVATE-SECTOR REMEDIES

Some proposals to address generative AI harms try to approach the problem at the source with technical remedies linked directly to the AI-generated content or the systems. For example, several regulations[162] and discussions have addressed watermarking and labeling images created by AI, including pledges by companies to implement these measures.[163] This technology is currently an imperfect fix—not only is it largely inapplicable to text and audio creations,[164] but the marks may be spoofed or removed by bad actors once the public is aware of them.[165]

Another step to ensure basic product safety and fitness, along with establishing more trust and transparency in generative AI systems, is implementing regular, independent audits and assessments of those systems.[166] This would require clarity and transparency from the companies and independent auditors and assessors with the technical expertise to analyze multiple areas of the systems (the algorithms, the training data, the outputs, layers of decision criteria, and legal and regulatory exposures).[167]

Finally, the generative AI industry could set its own industry standards for quality and consumer protection until specific regulations are in place. Some generative AI companies have already signed on to voluntary commitments on managing identified risks.[168] Similarly, individual companies have implemented policies and restrictions on their technology to head off some common problems (for example, Google paused its image Gemini AI system's ability to generate images of humans after reports of issues).[169] However, both voluntary commitments and individual company practices and policies can be easily changed,[170] so this does not adequately address generative AI harms long-term.

CONCLUSION

The continuous, expanding damages of generative AI are still being explored and discovered, but are in no way slowing down. While industry is still focused on the technology's theoretical potential, people are facing real-world harms to their privacy, civil rights, livelihood, and sanity. It is never too late to take action against harmful technology, but it becomes harder the longer we wait and the more that technology imbeds itself throughout society. By highlighting the impacts and risks of the technology, we hope to shift the conversation from AI's theoretical potential to its actual harms and how we can protect individuals.

References

[1] EPIC, Generating Harms: Generative AI's Impact & Paths Forward (2023), https://epic.org/gai [hereinafter "EPIC Generative AI Report"].

[2] *Poll Shows Overwhelming Concern About Risks From AI as new Institute Launches to Understand Public Opinion and Advocate for Responsible AI Policies,* A.I. Pol'y. Inst. (Aug. 2023), https://theaipi.org/poll-shows-overwhelming-concern-about-risks-from-ai-as-new-institute-launches-to-understand-public-opinion-and-advocate-for-responsible-ai-policies/ (Revealing that 72% of voters want to slow down AI development, 62% are primarily concerned about AI, and 82% don't trust tech executives to regulate AI); Alec Tyson & Emma Kikuchi, *Growing public concern about the role of artificial intelligence in daily life,* Pew Rsch. Ctr. (Aug. 28, 2023), https://www.pewresearch.org/short-reads/2023/08/28/growing-public-concern-about-the-role-of-artificial-intelligence-in-daily-life/ (52% of respondents are more concerned than excited by AI and only 10% are more excited than concerned).

[3] Steven Overly, *What really worries people about AI,* Politico (Feb. 29, 2024), https://www.politico.com/newsletters/digital-future-daily/2024/02/29/what-really-worries-people-about-ai-00144224 (79% of experts and 66% of the public worry that AI will have a negative impact on privacy).

[4] Mekela Panditharatne & Noah Giansiracusa, *How AI Puts Elections at Risk—And the Needed Safeguards,* The Brennan Ctr. for Just. (July 21, 2023), https://www.brennancenter.org/our-work/analysis-opinion/how-ai-puts-elections-risk-and-needed-safeguards.

[5] The World Economic Forum notes misinformation and disinformation as the most severe global risk in the next two years, "Global Risks Report 2024: Insight Report," World Econ. at F. 8 (Jan. 2024), https://www.weforum.org/publications/global-risks-report-2024/.

[6] Jeff Allen, *Misinformation Amplification Analysis and Tracking Dashboard,* Integrity Inst. (Oct. 13, 2022), https://integrityinstitute.org/blog/misinformation-amplification-tracking-dashboard.

[7] Mekela Panditharatne, *'News Deserts' Could Impact Midterm Elections,* Brennan Ctr. For Just. (Oct. 31, 2022), https://www.brennancenter.org/our-work/analysis-opinion/news-deserts-could-impact-midterm-elections.

[8] Morgan Meaker, *Slovakia's Election Deepfakes Show AI Is a Danger to Democracy,* Wired (Oct. 3, 2023), https://www.wired.com/story/slovakias-election-deepfakes-show-ai-is-a-danger-to-democracy/ (note that Slovakia has a legal bar on media or politicians discussing politics for 48 hours prior to polls opening, making addressing the audio even more difficult).

[9] Pranshu Verma & Gerrit De Vynck, *AI is destabilizing 'the concept of truth itself' in 2024 election,* Wash. Post (Jan. 22, 2024), https://www.washingtonpost.com/technology/2024/01/22/ai-deepfake-elections-politicians/; Nilesh Christopher, *An Indian politician says scandalous audio clips are AI deepfakes. We had them tested,* Rest of World (Jul. 5, 2023), https://restofworld.org/2023/indian-politician-leaked-audio-ai-deepfake/.

[10] John Hendel, *AI-generated Biden robocall linked to Texas companies, officials say,* Politico (Feb. 6, 2024), https://www.politico.com/news/2024/02/06/robocalls-fcc-new-hampshire-texas-00139864#:~:text=The%20calls%20included%20an%20artificial,to%20participate%20in%20their%20primary.

[11] Lola Fadulu, *2 Men Fined $1.25 Million for Robocall Scheme to Suppress Black Vote,* N.Y. Times (Apr. 9, 2024), https://www.nytimes.com/2024/04/09/nyregion/robocalls-black-voters-wohl-burkman.html.

[12] Emma Fitzsimmons & Jeffery Mays, *Since When Does Eric Adams Speak Spanish, Yiddish and Mandarin?,* N.Y. Times (Oct. 20, 2023), https://www.nytimes.com/2023/10/20/nyregion/ai-robocalls-eric-adams.html.

[13] *See* EPIC et al., Comments on FTC Rule on Impersonation of Government, Businesses, and Individuals (SNPRM) (Apr. 30, 2024), https://epic.org/documents/epic-and-partner-organizations-comments-on-ftc-rule-on-impersonation-of-government-businesses-and-individuals-snprm/.

[14] "Risk in Focus: Generative A.I. and the 2024 Election Cycle," Cybersecurity and Infrastructure Sec. Agency (Jan. 18, 2024), https://www.cisa.gov/sites/default/files/2024-01/Consolidated_Risk_in_Focus_Gen_AI_ElectionsV2_508c.pdf.

[15] Spencer Overton, *Overcoming Racial Harms to Democracy from Artificial Intelligence*, Iowa L. Rev. (Forthcoming) (Mar. 14, 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4754903.

[16] *Microsoft finds Russian influence operations targeting U.S. election have begun,* Reuters (April 17, 2024), https://www.reuters.com/world/us/microsoft-finds-russian-influence-operations-targeting-us-election-have-slowly-2024-04-17.

[17] Dustin Volz, *China is Targeting U.S. Voters and Taiwan With AI-Powered Disinformation,* Wall St. J. (April 5, 2024), https://www.wsj.com/politics/national-security/china-is-targeting-u-s-voters-and-taiwan-with-ai-powered-disinformation-34f59e21.

[18] Maggie Astor, *Behind the A.I. Robocall That Impersonated Biden: A Democratic Consultant and a Magician,* N.Y. Times (Feb. 27, 2024), https://www.nytimes.com/2024/02/27/us/politics/ai-robocall-biden-new-hampshire.html.

[19] Stuart A. Thompson, *Making Deepfakes Gets Cheaper and Easier Thanks to A.I.,* N.Y. Times (Mar. 12, 2023), https://www.nytimes.com/2023/03/12/technology/deepfakes-cheapfakes-videos-ai.html.

[20] Aleks Phillips, *Deepfake Video Shows Elizabeth Warren Saying Republicans Shouldn't Vote,* Newsweek (Feb. 27, 2023), https://www.newsweek.com/elizabeth-warren-msnbc-republicans-vote-deep-fake-video-1784117.

[21] Madalin Necsutu, *Moldova Dismisses Deepfake Video Targeting President Sandu,* Balkan Insight (Dec. 29, 2023), https://balkaninsight.com/2023/12/29/moldova-dismisses-deepfake-video-targeting-president-sandu/.

[22] Morgan Meaker, *Slovakia's Election Deepfakes Show AI Is a Danger to Democracy,* Wired (Oct. 3, 2023), https://www.wired.com/story/slovakias-election-deepfakes-show-ai-is-a-danger-to-democracy/.

[23] *Pakistanis, Bangladeshi politicians are new targets of deepfake, 90 per cent of videos online are pornographic,* Trib. India (Dec. 14, 2023), https://www.tribuneindia.com/news/trending/from-rashmika-mandanna-to-bangladeshi-politician-filmed-in-a-bikini-90-per-cent-of-deepfake-videos-online-are-pornographic-571782.

[24] Matt Novak, *Donald Trump Falsely Claims Attack Ad Used AI to Make Him Look Bad,* Forbes (Dec. 4, 2023), https://www.forbes.com/sites/mattnovak/2023/12/04/donald-trump-falsely-claims-attack-ad-used-ai-to-make-him-look-bad/.

[25] Weber Lai, *Deepfakes pose risk for the election,* Taipei Times (Dec. 11, 2023), https://www.taipeitimes.com/News/editorials/archives/2023/12/11/2003810443.

[26] Nilesh Christopher, *An Indian politician says scandalous audio clips are AI deepfakes. We had them tested,* Rest of World (Jul. 5, 2023), https://restofworld.org/2023/indian-politician-leaked-audio-ai-deepfake/.

[27] Allen, *supra* note 6.

[28] Jim Fournier, *How algorithms are amplifying misinformation and driving a wedge between people,* The Hill (Nov. 10, 2021), https://thehill.com/changing-america/opinion/581002-how-algorithms-are-amplifying-misinformation-and-driving-a-wedge/.

[29] Deceptive Practices and Voter Intimidation Prevention Act of 2021, S.1840, 117th Cong. (2021), https://www.congress.gov/bill/117th-congress/senate-bill/1840; H.4660, 125th Sess. (S.C. 2024), https://www.scstatehouse.gov/sess125_2023-2024/bills/4660.htm.

[30] Ali Swenson, *AI-generated voices in robocalls can deceive voters. The FCC just made them illegal,* Assoc. Press (Feb. 8, 2024), https://apnews.com/article/fcc-elections-artificial-intelligence-robocalls-regulations-a8292b1371b3764916461f60660b93e6; *FCC makes AI-Generated Voices in Robocalls Illegal,* Federal Communications Commission (Feb. 8, 2024), https://www.fcc.gov/document/fcc-makes-ai-generated-voices-robocalls-illegal; Fed. Trade Comm'n, Trade Rule on Impersonation of Government and Businesses, Supplemental Notice of Proposed Rulemaking; Request for Public Comment, 89 Fed. Reg. 15,072 (Mar. 1, 2024), https://www.federalregister.gov/documents/2024/03/01/2024-03793/trade-regulation-rule-on-impersonation-of-government-and-businesses.

[31] Matt O'Brien, *AI image-generator Midjourney blocks images of Biden and Trump as election looms,* Assoc. Press (Mar. 13, 2024), https://apnews.com/article/midjourney-ai-imagegenerator-biden-trump-deepfakes-bc6c254ddb20e36c5e750b4570889ce1.

[32] Matt O'Brien and Ali Swenson, *Tech companies sign accord to combat AI-generated election trickery,* Assoc. Press (Feb. 16, 2024), https://apnews.com/article/ai-generated-election-deepfakes-munich-accord-meta-google-microsoft-tiktok-x-c40924ffc68c94fac74fa994c520fc06.

[33] Apostol Vassilev, et al., *Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations*, NIST AI 100-2e2023 (Jan. 2014), https://doi.org/10.6028/NIST.AI.100-2e2023 at 40.

[34] Chad de Guzman and Will Henshall, *As Tech CEOs Are Grilled Over Child Safety Online, AI Is Complicating the Issue,* TIME (Feb. 2, 2024), https://time.com/6590470/csam-ai-tech-ceos/; David Thiel, *Investigation Finds AI Image Generation Models Trained on Child Abuse,* Stan. Univ. (Dec. 20, 2023), https://cyber.fsi.stanford.edu/news/investigation-finds-ai-image-generation-models-trained-child-abuse.

[35] Vassilev, et al., *supra* note 33 at 3.

[36] Benj Edwards, *Artist finds private medical record photos in popular AI training data set,* ArsTechnica (Sept. 21, 2022), https://arstechnica.com/information-

technology/2022/09/artist-finds-private-medical-record-photos-in-popular-ai-training-data-set/.

[37] Vassilev, et al., *supra* note 33 at 36.

[38] Autoriteit Persoonsgegevens, Richtlignen scraping door private organisaties en particulieren (May 1, 2024), https://www.autoriteitpersoonsgegevens.nl/uploads/2024-05/Handreiking%20scraping%20door%20particulieren%20en%20private%20organisaties.pdf.

[39] *Id.*

[40] *Id.*

[41] Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, European Commission, COM(2021) 206 final, 2021/0106 at § 5e, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206 [Hereinafter "EU AI Act"].

[42] Jason Koebler, *National Archives Bans Employee Use of ChatGPT,* 404 Media (May 1, 2024), https://www.404media.co/national-archives-bans-employee-use-of-chatgpt/; Rebecca Heilweil, *More federal agencies join in temporarily blocking or banning ChatGPT,* Fedscoop (Jan. 9, 2024), https://fedscoop.com/more-federal-agencies-join-in-temporarily-blocking-or-banning-chatgpt/.

[43] Jonathan Gillham, *Company AI Policy Examples and Template—Who Has Banned ChatGPT,* Originality.ai (Apr. 18, 2024), https://originality.ai/blog/ai-policy; Andrea Park, *Two-thirds of top 20 pharmas have banned ChatGPT-and many in life sci call AI 'overrated,' survey finds,* Fierce Pharma (Apr. 19, 2024), https://www.fiercepharma.com/marketing/two-thirds-top-20-pharmas-have-banned-chatgpt-and-many-life-sci-call-ai-overrated-survey.

[44] *Id.* at 40.

[45] *Id.* at 44–45.

[46] Vassilev, et al., *supra* note 33 at 1.

[47] Tiernan Ray, *ChatGPT can leak training data, violate privacy, says Google's DeepMind,* ZDNet (Dec. 4, 2023), https://www.zdnet.com/article/chatgpt-can-leak-source-data-violate-privacy-says-googles-deepmind/.

[48] Vishwam Sankaran, *ChatGPT Cooks Up Fake Sexual Harassment Scandal And Names Real Law Professor As Accused*, Indep. (Apr. 6, 2023), https://www.independent.co.uk/tech/chatgpt-sexual-harassment-law-professor-

b2315160.html; Pranshu Verma & Will Oremus, *ChatGPT invented a sexual harassment scandal and named a real law prof as the accused*, Wash. Post (Apr. 5, 2023), https://www.washingtonpost.com/technology/2023/04/05/chatgpt-lies/.

[49] Milad Nasr et al., "Extracting Training Data from ChatGPT," arXiv (Nov. 28, 2023), https://arxiv.org/abs/2311.17035; Milad Nasr et al., Extracting Training Data from ChatGPT, Not Just Memorization https://not-just-memorization.github.io/extracting-training-data-from-chatgpt.html?ref=404media.co) (summary of the academic research paper); Beatrice Nolan, *Google Researchers Say They Got OpenAI's ChatGPT To Reveal Some Of Its Training Data With Just One Word*, Bus. Insider (Dec. 4, 2023), https://www.businessinsider.com/google-researchers-openai-chatgpt-to-reveal-its-training-data-study-2023-12.

[50] Cloud and Threat Report 2024, Netskope (Aug. 2023), https://www.netskope.com/netskope-threat-labs/cloud-threat-report/cloud-and-threat-report-2024.

[51] Paolo Passeri, *The Risk of Accidental Data Exposure by Generative AI is Growing,* Infosecurity Mag. (Aug. 16, 2023), https://www.infosecurity-magazine.com/blogs/accidental-data-exposure-gen-ai/.

[52] James Coker, *ChatGPT Vulnerability May Have Exposed Users' Payment Information*, InfoSecurity Mag. (Mar. 29, 2023), https://www.infosecurity-magazine.com/news/chatgpt-vulnerability-payment/.

[53] David Barry, *Microsoft's AI Data Leak Isn't the Last One We'll See*, Reworked (Sept. 29, 2023), https://www.reworked.co/information-management/microsofts-ai-data-leak-isnt-the-last-one-well-see/.

[54] *See* Vassilev, et al., *supra* note 33 at 44, 48-49.

[55] Cory Doctorow, *The 'Enshittification' of TikTok,* Wired (Jan. 23, 2023), https://www.wired.com/story/tiktok-platforms-cory-doctorow/.

[56] Steven J. Vaughan-Nichols, *ChatGPT, how did you get here? It was a long journey through open source AI*, The Reg. (Mar. 24, 2023), https://www.theregister.com/2023/03/24/column/.

[57] Generative AI hallucinations occur when the system fully invents information to fit a prompt, such as in the multiple cases where lawyers have submitted documents to the court that contain made-up cases that fit the fact pattern they were searching for. *See, e.g.,* Dan Mangan, *Judge sanctions lawyers for brief written by A.I with fake citations,* CNBC (Jun. 22, 2023), https://www.cnbc.com/2023/06/22/judge-sanctions-lawyers-whose-ai-written-filing-contained-fake-citations.html; Leyland Cecco, *Canada lawyer*

under fire for submitting fake cases created by AI chatbot, The Guardian (Feb. 29, 2024), https://www.theguardian.com/world/2024/feb/29/canada-lawyer-chatgpt-fake-cases-ai.

[58] Ina Fried and Scott Rosenberg, *AI could choke on its own exhaust as it fills the web,* Axios (Aug. 28, 2023), https://www.axios.com/2023/08/28/ai-content-flood-model-collapse.

[59] While this report focuses mainly on the internet, the harms from generative AI span the entire digital ecosystem. Robocalls, advertising seen in real life such as billboards, TV shows and movies, mobile apps, and several other venues are equally bombarded with synthetic content.

[60] Krystal Hu, *ChatGPT sets record for fastest-growing user base-analysis*, Reuters (Feb. 2, 2023), https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/; Jon Porter, *ChatGPT continues to be one of the fastest-growing services ever,* The Verge (Nov. 6, 2023), https://www.theverge.com/2023/11/6/23948386/chatgpt-active-user-count-openai-developer-conference.

[61] Porter, *supra* note 60.

[62] *Id.*

[63] Sujan Sarkar, *AI Industry Analysis: 50 Most Visited AI Tools and Their 24B+ Traffic Behavior,* Writerbuddy (Sept. 2023), https://writerbuddy.ai/blog/ai-industry-analysis.

[64] John Herrman, *The Junkification of Amazon*, N.Y. Mag. (Jan. 30, 2023), https://nymag.com/intelligencer/2023/01/why-does-it-feel-like-amazon-is-making-itself-worse.html.

[65] Kaitlyn Tiffany, *AI-Generated Junk Is Flooding Etsy*, The Atlantic (Jun. 15, 2023), https://www.theatlantic.com/technology/archive/2023/06/ai-chatgpt-side-hustle/674415/.

[66] James Vincent, *AI is being used to generate whole spam sites*, The Verge (May 2, 2023), https://www.theverge.com/2023/5/2/23707788/ai-spam-content-farm-misinformation-reports-newsguard.

[67] Joseph Cox, *Google news Is Boosting Garbage AI-Generated Articles,* 404 Media (Jan. 18, 2024), https://www.404media.co/google-news-is-boosting-garbage-ai-generated-articles/; Jason Koebler, et al., *We Need Your Email Address,* 404 Media (Jan. 26, 2024) https://www.404media.co/why-404-media-needs-your-email-address/.

[68] Renee DiResta & Josh A. Goldstein, “How Spammers and Scammers Leverage AI-Generated Images on Facebook for Audience Growth,” arXiv (Mar. 19, 2024), https://arxiv.org/abs/2403.12838.

[69] *Id.* at 7.

[70] Public Citizen, *Tracker: State Legislation on Deepfakes in Elections* (Nov. 20, 2023), https://www.citizen.org/article/tracker-legislation-on-deepfakes-in-elections/; *see e.g.* N.Y. Elec. § 14-106(5)(b) (2024) (requiring disclosure of the use of generative AI in relation to elections. For visual synthetic content, the following disclosure needs to be clearly and easily legible: "This (image, video, or audio) has been manipulated." For synthetic audio such as a phone call or radio, the disclosure must be made before playing the audio.)

[71] *Synthetic Content*, NIST (Apr. 29, 2024), https://www.nist.gov/artificial-intelligence/executive-order-safe-secure-and-trustworthy-artificial-intelligence-2.

[72] Public Citizen, *supra* note 70.

[73] EU AI Act at Art. 50; Titus Wu, *California Lawmakers Push for Watermarks on AI-Made Photo, Video,* Bloomberg Law (Jan. 26, 2024), https://news.bloomberglaw.com/artificial-intelligence/california-lawmakers-push-for-watermarks-on-ai-made-photo-video; Susan Haigh, *Connecticut Senate passes wide-ranging bill to regulate AI. But its fate remains uncertain,* Assoc. Press (Apr. 24, 2024), https://apnews.com/article/artificial-intelligence-ai-connecticut-regulation-b004b4477ac20cc365317edff9f7351b.

[74] Information Commissioner's Office, *Generative AI third call for evidence: accuracy of training data and model outputs* (Apr. 12, 2024), https://ico.org.uk/about-the-ico/what-we-do/our-work-on-artificial-intelligence/generative-ai-third-call-for-evidence/.

[75] EPIC, Comments on the NTIA Request for Comment: Dual Use Foundation Artificial Intelligence Models with Widely Available Model Weights (Mar. 27, 2024), https://epic.org/wp-content/uploads/2024/03/EPIC_Comment_NTIA_Dual_Use_Foundation_Models_with_Appendix.pdf.

[76] Mehrdad Saberi et al., *Robustness of AI-Image Detectors: Fundamental Limits and Practical Attacks*, arXiv (Sept. 29, 2023) (preprint), https://arxiv.org/pdf/2310.00076.pdf.

[77] Gerrit De Vynck, *The AI deepfake apocalypse is here. These are the ideas for fighting it.,* Wash. Post (Apr. 5, 2024), https://www.washingtonpost.com/technology/2024/04/05/ai-deepfakes-detection/.

[78] Jonathan Gillham, *ContentAtScale AI Content Detection Review*, Originality.ai (Dec. 22, 2023), https://originality.ai/blog/contentatscale-ai-content-detection-review.

[79] Francisco Pires, *OpenAI Sunsets Generative AI Text Detection Tool*, tom'sHARDWARE (Jul. 26, 2023), https://www.tomshardware.com/news/openai-sunsets-generative-ai-text-detection-tool.

[80] Gary Drenik, *Data Privacy and ownership To Remain Key Concerns In Web Scraping Industry Next Year,* Forbes (Dec. 18, 2023), https://www.forbes.com/sites/garydrenik/2023/12/18/data-privacy-and-ownership-to-remain-key-concerns-in-web-scraping-industry-next-year/.

[81] *See* EPIC, Comments to NIST on Information Related to NIST's Assignments under Sections 4.1, 4.5, and 11 of the Executive Order Concerning Artificial Intelligence (Feb. 2, 2024), https://epic.org/wp-content/uploads/2024/02/EPIC-Comment-on-NIST-AI-Executive-Order-Mandates-RFI-02.02.24.pdf (describing value and limitations of watermarking); Makena Kell, *Watermarks Aren't the Silver Bullet for AI Misinformation*, Verge (Oct. 31, 2023), https://www.theverge.com/2023/10/31/23940626/artificial-intelligence-ai-digital-watermarks-biden-executiveorder; Mehrdad Saberi et al., *Robustness of AI-Image Detectors: Fundamental Limits and Practical Attacks*, arXiv (Sept. 29, 2023) (preprint), https://arxiv.org/pdf/2310.00076.pdf; David Pierce, *Google Made a Watermark for AI Images That You Can't Edit Out*, Verge (Aug. 29, 2023), https://www.theverge.com/2023/8/29/23849107/synthid-google-deepmind-ai-image-detector; Ilia Shumailov et al., *The Curse of Recursion: Training on Generated Data Makes Models Forget*, arXiv (Cambridge Univ. Working Paper, 2023), https://www.cl.cam.ac.uk/~is410/Papers/dementia_arxiv.pdf.

[82] Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy Democracy, and National Security*, 107 Cal. L. Rev. 1753 (2019), https://www.californialawreview.org/print/deep-fakes-a-looming-challenge-for-privacy-democracy-and-national-security.

[83] Neil Richards, *Why Privacy Matters* 35–37 (2022) ("But the important lesson of [Target using data analytics to infer pregnancy status to tailor ads to individuals] is not actually about the power of human information analytics to find surprising correlations like the one between lotion and pregnancy. Instead, the real lesson is about the power those insights confer to control human behavior. The reason Target wants to know about pregnancy is because Target wants consumers to buy as much as possible of everything they sell at their big box stores—not just diapers and baby clothes, but lawn furniture and underwear, wine and electronics.")

[84] David Pierce, *Google is starting to squash more spam and AI in search results,* The Verge (Mar. 5, 2024), https://www.theverge.com/2024/3/5/24091099/google-search-high-quality-results-spam-ai-content.

[85] Mia Sato, *The unsettling scourge of obituary spam,* The Verge (Feb. 12, 2024), https://www.theverge.com/24065145/ai-obituary-spam-generative-clickbait.

[86] Shumailov et al., *supra* note 81.

[87] Andrew Limbong, *Authors push back on the growing number of AI 'scam' books on Amazon,* NPR (Mar. 13, 2024), https://www.npr.org/2024/03/13/1237888126/growing-number-ai-scam-books-amazon; Kate Knibbs, *Scammy AI-Generated Book Rewrites Are Flooding Amazon,* Wired (Jan. 10, 2024), https://www.wired.com/story/scammy-ai-generated-books-flooding-amazon/.

[88] McKenzie Sadeghi et al., *Tracking AI-enabled Misinformation: 811 'Unreliable AI-Generated News' Websites (and Counting), Plus the Top False Narratives Generated by Artificial Intelligence Tools,* NewsGuard (Apr. 29, 2024), https://www.newsguardtech.com/special-reports/ai-tracking-center/; *Proliferating 'news' sites spew AI-generated stories,* France24 (Mar. 11, 2024), https://www.france24.com/en/live-news/20240311-proliferating-news-sites-spew-ai-generated-fake-stories; Stuart Thompson, *A.I.-Generated Content Discovered on News Sites, Content Farms and Product Reviews,* N.Y. Times (May 19, 2023), https://www.nytimes.com/2023/05/19/technology/ai-generated-content-discovered-on-news-sites-content-farms-and-product-reviews.html; Matthew Cantor, *Nearly 50 news websites are 'AI-generated', a study says. Would I be able to tell?,* The Guardian (May 8, 2023), https://www.theguardian.com/technology/2023/may/08/ai-generated-news-websites-study.

[89] Victor Tangermann, *Gizmodo and Kotaku Staff Furious After Owner Announces Move to AI Content,* Futurism (June 30, 2023), https://futurism.com/gizmodo-kotaku-staff-furious-ai-content; Jade Drummond, *Newsrooms around the world are using AI to optimize work, despite concerns about bias and accuracy,* The Verge (Sep. 28, 2023), https://www.theverge.com/2023/9/28/23894651/ai-newsroom-journalism-study-automation-bias.

[90] Benjamin Weiser & Jonah Bromwich, *Michael Cohen Used Artificial Intelligence in Feeding Lawyer Bogus Cases,* N.Y. Times (Dec. 29, 2023), https://www.nytimes.com/2023/12/29/nyregion/michael-cohen-ai-fake-cases.html; Dan Mangan, *Judge sanctions lawyers for brief written by A.I with fake citations,* CNBC (Jun. 22, 2023), https://www.cnbc.com/2023/06/22/judge-sanctions-lawyers-whose-ai-written-filing-contained-fake-citations.html; Leyland Cecco, *Canada lawyer under fire for submitting fake cases created by AI chatbot,* The Guardian (Feb. 29, 2024), https://www.theguardian.com/world/2024/feb/29/canada-lawyer-chatgpt-fake-cases-ai.

[91] *Supra* note 81.

[92] *See, e.g.*, *Doe 1 et al. v. GitHub, Inc. et al.*, 672 F. Supp. 3d 837 (N.D. Cal. 2023); *Doe 3 et al. v. GitHub, Inc. et al.*, No. 4:22-CV-07074, (N.D. Cal. Nov. 10, 2022).

[93] *See* Jonathan Gillham, *OpenAI and ChatGPT Lawsuit List*, Originality.ai (May 1, 2024), https://originality.ai/blog/openai-chatgpt-lawsuit-list;

[94] *See* Matt Scherer, *The SAG-AFTRA Strike is Over, But the AI Fight in Hollywood is Just Beginning*, CDT (Jan. 4, 2024), https://cdt.org/insights/the-sag-aftra-strike-is-over-but-the-ai-fight-in-hollywood-is-just-beginning/.

[95] *See, e.g.*, Maria Yagoda, *Airline Held Liable for its Chatbot Giving Passenger Bad Advice—What This Means for Travelers*, BBC (Feb. 23, 2024), https://www.bbc.com/travel/article/20240222-air-canada-chatbot-misinformation-what-travellers-should-know; *Introducing New AI Experiences Across Our Family of Apps and Devices*, Meta Blog (Sept. 27, 2023), https://about.fb.com/news/2023/09/introducing-ai-powered-assistants-characters-and-creative-tools/; Colin Wood, *After giving wrong answers, NYC chatbot to stay online for testing,* State Scoop (Apr. 3, 2024), https://statescoop.com/nyc-mayor-eric-adams-chatbot-wrong-answers/; Jessica Nix, *AI-Powered World Health Chatbot Is Flubbing Some Answers,* Bloomberg (Apr. 18, 2024), https://www.bloomberg.com/news/articles/2024-04-18/who-s-new-ai-health-chatbot-sarah-gets-many-medical-questions-wrong.

[96] *See* Jules Roscoe, *A 'Shocking' Amount of the Web is Already AI-Translated Trash, Scientists Determine*, Vice (Jan. 17, 2024), https://www.vice.com/en/article/y3w4gw/a-shocking-amount-of-the-web-is-already-ai-translated-trash-scientists-determine; Fried & Rosenberg, *supra* note 58.

[97] *See, e.g.*, Bryson Masse, *OpenAI Launches Web Crawling GPTBot, Sparking Blocking Effort by Website Owners and Creators*, VentureBeat (Aug. 8, 2023), https://venturebeat.com/ai/openai-launches-web-crawling-gptbot-sparking-blocking-effort-by-website-owners-and-creators/.

[98] Emma Roth, *OpenAI's DALL-E Will Train on Shutterstock's Library for Six More Years*, The Verge (July 11, 2023), https://www.theverge.com/2023/7/11/23791528/openai-shutterstock-images-partnership.

[99] Matt O'Brien, *ChatGPT-Maker OpenAI Signs Deal with AP to License News Stories*, Assoc. Press (July 13, 2023), https://apnews.com/article/openai-chatgpt-associated-press-ap-f86f84c5bcc2f3b98074b38521f5f75a.

[100] Anna Tong et al., *Exclusive: Reddit in AI Content Licensing Deal with Google*, Reuters (Feb. 21, 2024), https://www.reuters.com/technology/reddit-ai-content-licensing-deal-with-google-sources-say-2024-02-22/.

[101] *See* Carl Franzen, *The AI Feedback Loop: Researchers Warn of 'Model Collapse' as AI trains on AI-Generated Content*, VentureBeat (June 12, 2023),

https://venturebeat.com/ai/the-ai-feedback-loop-researchers-warn-ofmodel-collapse-as-ai-trains-on-ai-generated-content/; Shumailov et al., *supra* note 81.

[102] *See* Hayden Field & Kif Leswing, *Generative AI 'FOMO' is Driving Tech Heavyweights to Invest Billions of Dollars in Startups*, CNBC (Mar. 30, 2024), https://www.cnbc.com/2024/03/30/fomo-drives-tech-heavyweights-to-invest-billions-in-generative-ai-.html.

[103] *See* Gillham, *supra* note 93.

[104] *See* Annelise Gilbert, *Google-Reddit AI Deal Heralds New Era in Social Media Licensing*, BL (Mar. 7, 2024), https://news.bloomberglaw.com/ip-law/google-reddit-ai-deal-just-the-start-for-social-media-licensing.

[105] *See, e.g.*, Bell Wong, *Top Social Media Statistics and Trends of 2024*, Forbes Advisor (May 18, 2023), https://www.forbes.com/advisor/business/social-media-statistics/.

[106] *See, e.g.*, *Reddit User Agreement*, Reddit (Sept. 25, 2023), https://www.redditinc.com/policies/user-agreement-september-25-2023.

[107] *See, e.g.*, Kaitlyn Nguyen, *AI is Causing Student Artists to Rethink Their Creative Career Plans*, KQED (Apr. 26, 2023), https://www.kqed.org/arts/13928253/ai-art-artificial-intelligence-student-artists-midjourney.

[108] *See* Will Knight, *Algorithms Can Now Mimic Any Artist. Some Artists Hate It*, Wired (Aug. 19, 2022), https://www.wired.com/story/artists-rage-against-machines-that-mimic-their-work/; Sarah Andersen, *The Alt-Right Manipulated My Comic. Then A.I. Claimed It.,* N.Y. Times (Dec. 31, 2022), https://www.nytimes.com/2022/12/31/opinion/sarah-andersen-how-algorithim-took-my-work.html; Nick Cave, *Issue #218,* The Red Hand Files (Jan. 2023), https://www.theredhandfiles.com/chat-gpt-what-do-you-think/; Beatrice Nolan, *Artists say AI image generators are copying their style to make thousands of new images – and it's completely out of their control,* Bus. Insider (Oct. 17, 2022), https://www.businessinsider.com/ai-image-generators-artists-copying-style-thousands-images-2022-10.

[109] Gilbert, *supra* note 104.

[110] *See* Benjamin Mullin, *Inside the News Industry's Uneasy Negotiations with OpenAI*, N.Y. Times (Dec. 29, 2023), https://www.nytimes.com/2023/12/29/business/media/media-openai-chatgpt.html.

[111] Shutterstock has already adopted a content contributor fund approach, where they will directly compensate content creators if their content was used to develop generative AI models. *See AI-Generated Content on Shutterstock: Contributor FAQ*, Shutterstock Contributor Support (Apr. 18, 2024),

https://support.submit.shutterstock.com/s/article/Shutterstock-ai-and-Computer-Vision-Contributor-FAQ?language=en_US.

[112] Shawn Shan et al., *About the Glaze Project*, Sand Lab at U. Chi., https://glaze.cs.uchicago.edu/aboutus.html (last visited May 1, 2024).

[113] Shawn Shan et al., *What is Nightshade?*, Sand Lab at U. Chi., https://nightshade.cs.uchicago.edu/whatis.html (last visited May 1, 2024).

[114] *Supra* note 81.

[115] Kevin Purdy, *Fake AI law firms are sending fake DMCA threats to generate fake SEO gains,* Ars Technica (April 4, 2024), https://arstechnica.com/gadgets/2024/04/fake-ai-law-firms-are-sending-fake-dmca-threats-to-generate-fake-seo-gains/.

[116] *See, e.g.,* Miles Klee & Nikki McCann Ramirez, *AI Has Made the Israel-Hamas misinformation Epidemic Much, Much Worse,* Rolling Stone (Oct. 27, 2023), https://www.rollingstone.com/politics/politics-features/israel-hamas-misinformation-fueled-ai-images-1234863586/; Isabelle Frances-Wright & Moustafa Ayad, *Misleading and manipulated content goes viral on X in Middle East conflict,* Inst. for Strategic Dialogue (April 14, 2024), https://www.isdglobal.org/digital_dispatches/misleading-and-manipulated-content-goes-viral-on-x-twitter-in-middle-east-conflict-iran-israel-strikes; Amanda Hoover, *Hulu Shows Jarring Anti-Hamas Ad Likely Generated With AI,* Wired (Jan. 30, 2024), https://www.wired.com/story/hulu-anti-hamas-ad-generative-ai/.

[117] Leon Yin, Davey Alba, and Leonardo Nicoletti, *OpenAI's GPT is a Recruiter's Dream Tool. Tests Show There's Racial Bias,* Bloomberg (Mar. 7, 2024), https://www.bloomberg.com/graphics/2024-openai-gpt-hiring-racial-discrimination.

[118] Clayton Vickers, *Law enforcement struggling to prosecute AI-generated child pornography, asks Congress to act,* The Hill (Mar. 13, 2024), https://thehill.com/homenews/house/4530044-law-enforcement-struggling-prosecute-ai-generated-child-porn-asks-congress-act/.

[119] Bethan McKernan and Harry Davies, *'The machine did it coldly': Israel used AI to identify 37,000 Hamas targets,* The Guardian (Apr. 3, 2024), https://www.theguardian.com/world/2024/apr/03/israel-gaza-ai-database-hamas-airstrikes.

[120] Will Oremus, *Hackers competed to find AI harms. Here's what they found.,* Wash. Post (Apr. 4, 2024), https://www.washingtonpost.com/politics/2024/04/04/hackers-competed-find-ai-harms-heres-what-they-found/.

[121] *FTC Launches Inquiry into Generative AI Investments and Partnerships,* Fed. Trade Comm'n (Jan. 25, 2024), https://www.ftc.gov/news-events/news/press-releases/2024/01/ftc-launches-inquiry-generative-ai-investments-partnerships.

[122] Rule on Impersonation of Government and Businesses, 16 C.F.R. § 461 (2024).

[123] K.C. Halm et al., *Who's Liable for Deepfakes? FTC Proposes to Target Developers of Generative AI Tools in Addition to Fraudsters,* Davis Wright Tremaine LLP (Feb. 22, 2024), https://www.dwt.com/blogs/artificial-intelligence-law-advisor/2024/02/ftc-targets-tech-companies-for-generative-ai-fraud.

[124] Alex Weprin, *Reddit Says FTC Inquiring About Deals to License Data and Train AI Models,* The Hollywood Rep. (Mar. 15, 2024), https://www.hollywoodreporter.com/business/digital/reddit-ftc-investigation-license-data-train-ai-models-1235853771/.

[125] *FCC Makes AI-Generated Voices in Robocalls Illegal,* Fed. Commc'n. Comm'n. (Feb. 8, 2024), https://www.fcc.gov/document/fcc-makes-ai-generated-voices-robocalls-illegal.

[126] U.S. Copyright Office Review Board, *Decision Affirming Refusal of Registration of a Recent Entrance to Paradise* (Feb. 14, 2022), https://www.copyright.gov/rulings-filings/review-board/docs/a-recent-entrance-to-paradise.pdf.

[127] U.S. Copyright Office, *Cancellation Decision re: Zarya of the Dawn (VAu001480196)* (Feb. 21, 2023), https://www.copyright.gov/docs/zarya-of-the-dawn.pdf.

[128] U.S. Copyright Office, *Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence,* 88 FR 16190 (Mar. 16, 2023), https://www.federalregister.gov/documents/2023/03/16/2023-05321/copyright-registration-guidance-works-containing-material-generated-by-artificial-intelligence.

[129] Michael M. Grynbaum & Ryan Mac, *The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work,* N.Y. Times (Dec. 27, 2023), https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html.

[130] Emilia David, *The Intercept, Raw Story, and AlterNet sue OpenAI and Microsoft,* The Verge (Feb. 28, 2024), https://www.theverge.com/2024/2/28/24085973/intercept-raw-story-alternet-openai-lawsuit-copyright.

[131] James Vincent, *Getty Images sues AI art generator Stable Diffusion in the US for copyright infringement,* The Verge (Feb. 6, 2023), https://www.theverge.com/2023/2/6/23587393/ai-art-copyright-lawsuit-getty-images-stable-diffusion.

[132] Emilia David, *Universal Music sues AI company Anthropic for distributing song lyrics,* The Verge (Oct. 19, 2023), https://www.theverge.com/2023/10/19/23924100/universal-music-sue-anthropic-lyrics-copyright-katy-perry.

[133] Cat Zakrzewski et al., *OpenAI prepares to fight for its life as legal troubles mount,* Wash. Post (Apr. 9, 2024), https://www.washingtonpost.com/technology/2024/04/09/openai-lawsuit-regulation-lawyers/.

[134] Shanti Escalante-De Mattei, *Artists Are Suing Artificial Intelligence Companies and the Lawsuit Could Upend Legal Precedents Around Art,* Art in America (May 5, 2023), https://www.artnews.com/art-in-america/features/midjourney-ai-art-image-generators-lawsuit-1234665579/.

[135] Artificial Intelligence Legislation Tracker, Brennan Ctr. For Just. (last updated April 1, 2024), https://www.brennancenter.org/our-work/research-reports/artificial-intelligence-legislation-tracker.

[136] Press Release, *FTC Launches Inquiry into Generative AI Investments and Partnerships,* Fed. Trade Comm'n. (Jan. 25, 2024), https://www.ftc.gov/news-events/news/press-releases/2024/01/ftc-launches-inquiry-generative-ai-investments-partnerships; Press Release, *FTC Proposes New Protections to Combat AI Impersonation of Individuals,* Fed. Trade Comm'n. (Feb. 15, 2024), https://www.ftc.gov/news-events/news/press-releases/2024/02/ftc-proposes-new-protections-combat-ai-impersonation-individuals.

[137] Declaratory Ruling, "Implications of Artificial Intelligence Technologies on Protecting Consumers from Unwanted Robocalls and Robotexts," Fed. Commc'n Comm'n. FCC-24-17, Doc. No. 23-362 (Feb. 8. 2024).

[138] Press Release, *CFPB Issues Guidance on Credit Denials by Lenders Using Artificial Intelligence,* Consumer Fin. Prot. Bureau (Sept. 19, 2023), https://www.consumerfinance.gov/about-us/newsroom/cfpb-issues-guidance-on-credit-denials-by-lenders-using-artificial-intelligence/.

[139] Reid Blackman & Ingrid Vasiliu-Feltes, *The EU's AI Act and How Companies Can Achieve Compliance,* Harv. Bus. Rev. (Feb. 22, 2024), https://hbr.org/2024/02/the-eus-ai-act-and-how-companies-can-achieve-compliance.

[140] US Federal Privacy Legislation Tracker, Int'l Assoc. of Priv. Pro. (last updated Mar. 2024), https://iapp.org/resources/article/us-federal-privacy-legislation-tracker/.

[141] *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence,* Biden White House (Oct. 30, 2023),

https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/; Kara Williams, *Summary: What Does Biden's Executive Order on Artificial Intelligence Actually Say?,* EPIC (Nov. 7, 2023), https://epic.org/summary-what-does-bidens-executive-order-on-artificial-intelligence-actually-say/.

[142] Katrina Zhu, *The State of State AI Laws: 2023,* EPIC (Aug. 3, 2023), https://epic.org/the-state-of-state-ai-laws-2023/.

[143] Press Release, "AG Campbell issues Advisory Providing Guidance On How State Consumer Protection And Other Laws Apply To Artificial Intelligence," Off. of the Mass. Att'y Gen. (April 16, 2024), https://www.mass.gov/news/ag-campbell-issues-advisory-providing-guidance-on-how-state-consumer-protection-and-other-laws-apply-to-artificial-intelligence.

[144] SB 149, Artificial Intelligence Amendments, Utah, 2024 General Session, https://le.utah.gov/~2024/bills/sbillint/SB0149.pdf.

[145] Local Law 2021/144, Automated Employment Decision Tools, New York, Ch. 5, Title 20, Admin Code of City of New York, https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4344524&GUID=B051915D-A9AC-451E-81F8-6596032FA3F9; Artificial Intelligence Video Interview Act 820 Ill. Comp. Stat. 42/1 (2020), https://www.ilga.gov/legislation/ilcs/ilcs3.asp?ActID=4015&ChapterID=68; Use of Facial Recognition Services Prohibited -- Consent by Applicant , Md. Lab. & Empl. § 3-717 (2023).

[146] Bolstering Online Transparency Act (BOT), Cal. Bus. & Prof. Code § 17940 *et seq.* (2023)

[147] Restrict Insurers' Use of External Consumer Data, Col. Rev. Stat. 10-3-1104.9 (2021).

[148] Public Citizen, *Tracker: State Legislation on Deepfakes in Elections* (Nov. 20, 2023), https://www.citizen.org/article/tracker-legislation-on-deepfakes-in-elections/; *see e.g.* N.Y. Elec. § 14-106(5)(b) (2024) (requiring disclosure of the use of generative AI in relation to elections. For visual synthetic content, the following disclosure needs to be clearly and easily legible: "This (image, video, or audio) has been manipulated." For synthetic audio such as a phone call or radio, the disclosure must be made before playing the audio.)

[148] *Synthetic Content*, NIST (Apr. 29, 2024), https://www.nist.gov/artificial-intelligence/executive-order-safe-secure-and-trustworthy-artificial-intelligence-2.

[149] Ga. Code Ann. § 31-12-12 (2023).

[150] California AB-2013, https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240AB2013.

[151] Florida SB 850, https://m.flsenate.gov/Bill/850/; New York Assembly Bill A7904A, https://www.nysenate.gov/legislation/bills/2023/A7904/amendment/A.

[152] New York Assembly Bill A8126, https://www.nysenate.gov/legislation/bills/2023/A8129.

[153] New York Assembly Bill A8098, https://www.nysenate.gov/legislation/bills/2023/A8098/amendment/original; New York Assembly Bill A8158, https://www.nysenate.gov/legislation/bills/2023/A8158.

[154] US State-By-State AI Legislation Snapshot, Bryan Cave Leighton Paisner (last updated Jan. 26, 2024), https://www.bclplaw.com/en-US/events-insights-news/2023-state-by-state-artificial-intelligence-legislation-snapshot.html.

[155] Mikhail Klimentov, *From China to Brazil, here's how AI is regulated around the world,* Wash. Post (Sept. 3, 2023), https://www.washingtonpost.com/world/2023/09/03/ai-regulation-law-china-israel-eu/.

[156] China "Provisions on the Management of Algorithmic Recommendations in Internet Information Services," translation available at https://www.chinalawtranslate.com/en/algorithms/.

[157] China "Provisions on the Administration of Deep Synthesis Internet Information Services," translation available at https://www.chinalawtranslate.com/en/deep-synthesis/.

[158] China "Interim Measures for the Management of Generative Artificial Intelligence Services," translation available at https://www.chinalawtranslate.com/en/generative-ai-interim/.

[159] Brazil Bill 2338/2023, covered by Rob Rodrigues, et al., *Brazilian Lawmaker Introduces Bill to Allow AI as Inventor,* Lexology (Feb. 29, 2024), https://www.lexology.com/library/detail.aspx?g=6f1183c4-7670-4208-bf64-5a5d1221ff47; Canada Bill C-27, First Session, Forty-fourth Parliament, https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading; China "Measures for the Management of Generative Artificial Intelligence Services (Draft for Comment)," translation available at https://digichina.stanford.edu/work/translation-measures-for-the-management-of-generative-artificial-intelligence-services-draft-for-comment-april-2023/.

[160] EU AI Act, *supra* note 41.

[161] Maria Villegas Bravo, *What U.S. Regulators can Learn from the EU AI Act,* EPIC (Mar. 22, 2024), https://epic.org/what-u-s-regulators-can-learn-from-the-eu-ai-act/.

[162] EU AI Act Art. 50(2) ("Providers of AI systems, including GPAI systems, generating synthetic audio, image, video or text content, shall ensure the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated."); Xuezi Dan & Yan Luo, *Labeling of AI Generated Content: New Guidelines Released in China,* Lexology (Aug. 25, 2023), https://www.lexology.com/library/detail.aspx?g=656b7b0d-9b82-4fa9-9b50-34d6ceeb4ba9.

[163] Nihal Krishan, *AI watermarking could be exploited by bad actors to spread misinformation. But experts say the tech still must be adopted quickly,* FedScoop (Jan. 3, 2024), https://fedscoop.com/ai-watermarking-misinformation-election-bad-actors-congress/; Kris Holt, *Meta plans to ramp up labeling of AI-generated images across its platforms,* Engadget (Feb. 6, 2024), https://www.engadget.com/meta-plans-to-ramp-up-labeling-of-ai-generated-images-across-its-platforms-160234038.html.

[164] Bob Gleichauf & Dan Geer, *Digital Watermarks Are Not Ready for Large Language Models,* LawFare (Feb. 29, 2024), https://www.lawfaremedia.org/article/digital-watermarks-are-not-ready-for-large-language-models.

[165] Nihal Krishan, *AI watermarking could be exploited by bad actors to spread misinformation. But experts say the tech still must be adopted quickly,* FedScoop (Jan. 3, 2024), https://fedscoop.com/ai-watermarking-misinformation-election-bad-actors-congress/; Kat Tenbarge & Kevin Collier, *Big Tech says AI watermarks could curb misinformation, but they're easy to sidestep,* NBC News (Mar. 19, 2024), https://www.nbcnews.com/tech/tech-news/watermark-deepfake-solution-ai-misinformation-cant-stop-de-rcna137370.

[166] Press Release, *Hickenlooper Proposes AI Auditing Standards, Calls for Protecting Consumer Data, Increasing Transparency,* U.S. Sen. Hickenlooper for Colo. (Feb. 5, 2024), https://www.hickenlooper.senate.gov/press_releases/hickenlooper-proposes-ai-auditing-standards-calls-for-protecting-consumer-data-increasing-transparency/.

[167] Mark Dangelo, *Auditing AI: The emerging battlefield of transparency and assessment,* Thomson Reuters (Oct. 25, 2023), https://www.thomsonreuters.com/en-us/posts/technology/auditing-ai-transparency/.

[168] *Fact Sheet: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI,* Biden White House (Jul. 21, 2023), https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/.

[169] Sabrina Ortiz, *Why Google just banned Gemini from generating images of people,* ZDNet (Feb. 22, 2024), https://www.zdnet.com/article/why-google-just-banned-gemini-from-generating-images-of-people/; *see also* Gerrit De Vynck, *AI companies agree to limit election 'deepfakes' but fall short of ban,* Wash. Post (Feb. 13, 2024), https://www.washingtonpost.com/technology/2024/02/13/google-ai-elections-deepfakes-open/; Matt O'Brien, *AI image-generator Midjourney blocks images of Biden and Trump as election looms,* Assoc. Press (Mar. 13, 2024), https://apnews.com/article/midjourney-ai-imagegenerator-biden-trump-deepfakes-bc6c254ddb20e36c5e750b4570889ce1.

[170] Sam Biddle, *OpenAI quietly deletes ban on using ChatGPT for "military and warfare,"* The Intercept (Jan. 12, 2024), https://theintercept.com/2024/01/12/open-ai-military-ban-chatgpt/.