

# FTC Report Signals Caution On AI Tools For Online Content

By **Anthony DiResta, Kwamina Thomas Williford and Wanqian Zhang** (July 25, 2022)

An interesting report by the Federal Trade Commission, released on June 16, revealed that the agency is critical of the use of artificial intelligence to combat online harms. In fact, the agency found that the use of AI has not significantly curtailed online harms overall, and may even be creating biased and discriminatory practices.



Anthony DiResta

Online harms that are of particular concern include online fraud, impersonation scams, fake reviews and accounts, bots, media manipulation, illegal drug sales and other illegal activities, sexual exploitation, hate crimes, online harassment and cyberstalking, and misinformation campaigns aimed at influencing elections.[1]

While the FTC's report recognizes AI's use in combating harmful content and other positive outcomes, it also cautions against overreliance on the technology.



Kwamina Thomas Williford

While some believe that the report has shortcomings, it is nevertheless important for companies to familiarize themselves with the FTC's concerns about the use of AI, and heed its guidance.

For example, data minimization is critical. Companies should collect only the information necessary to provide the service or product to the consumer.

In addition, companies should be transparent, and provide all material information upfront to the consumer that is relevant to the nature of the transaction and the purchasing decision.



Wanqian Zhang

Finally, human oversight and monitoring should be enhanced. Robust complaint management and awareness of regulatory compliance developments are critical.

## **AI's Shortcomings, and Other Problems**

The FTC report finds AI to be effective in combating harms for which detection requires no context — including illegal items sold online and child pornography — and recognizes effective AI systems in preventing the inadvertent release of harmful information.

AI can be used for intervention or friction purposes before the release of harmful content, including labeling, adding interstitials and sending warnings. But the FTC does not believe these strategies prevent maliciously spread information.[2]

Platforms can also use AI tools to address online harms by finding the networks and actors behind them. AI tools can facilitate cross-platform mapping of certain communities spreading harmful contents. However, these strategies can also inadvertently ensnare marginalized communities using protected methods to communicate about authoritarian regimes.

Notwithstanding the inevitability of AI use, the FTC is concerned with using AI to combat online harms, and cautions against overreliance on it for several reasons.

First, AI tools have built-in imprecision. The FTC report cautions that data sets used to train AI systems are often not sufficiently large, accurate and representative, and the classifications can be problematic.

The report explains that AI tools are generally deficient at detecting and including new phenomena, and the operation of AI tools is subject to platform moderation policies that may be substantially flawed. The report also suggests that AI tools are often unreliable at understanding context, and therefore typically cannot effectively detect frauds, fake reviews and other implicitly harmful contents.

The report further suggests that use of AI cannot solve — and instead, can exacerbate — bias and discrimination. It explains that inappropriate data sets and a lack of diverse perspectives among AI designers can exacerbate discrimination against marginalized groups.

The report cautions that big technology companies can influence institutions and researchers, and set the agenda for which AI research the government funds.

Additionally, the report warns that AI tools used to uncover networks and actors behind harmful contents may inadvertently stifle minority groups. The FTC's research indicates that AI development can incentivize invasive consumer surveillance — because improving AI systems requires amassing large amount of accurate, representative training data.

Finally, the report notes that bad actors can easily escape AI detection by hacking, using their own developing AI technology, or simply using typos and euphemisms. It also warns that the massive amount of ordinary and pervasive posts that express discriminatory sentiments cannot be detected effectively by AI, even under human oversight.

## **Proposed Recommendations**

The report identifies the need to increase the transparency and accountability of those deploying AI as a top priority. It stresses the importance of increasing data and AI designer/moderator diversity to combat bias and discrimination. The report also finds that human oversight is a necessity.

### ***Transparency***

The FTC report stressed that to increase transparency, platforms and other entities should do the following:

- Make sufficient disclosure to consumers about their basic civil rights and how their rights are influenced by AI. The report points out that consumers have the right to be free from being subjected to inaccurate and biased AI, the right to be free from pervasive or discriminatory surveillance and monitoring, and the right to meaningful recourse if the use of an algorithm harms them.
- Give researchers access to sufficient, useful, intelligible data and algorithms for them to properly analyze the utility of AI, and the spread and impact of misinformation.

- Keep auditing and assessment independent, while protecting auditors and whistleblowers who report illegal AI use.

### ***Accountability***

The FTC report stressed that to increase accountability, platforms and other entities should conduct regular audits and impact assessments, should be held accountable for the outcome and impact of their AI systems, and provide appropriate redress for erroneous or unfair algorithmic decisions.

### ***Assessing Through a Diverse Lens***

The FTC report recommends increasing diversity in data sets, AI designers and moderators. Firms need to retain people with diverse perspectives, and should strive to create and maintain diverse, equitable and inclusive cultures.

AI developers should be aware of the context where the data is being used, and the potential discriminatory harm it could cause, and mitigate any such harm in advance.

### ***Human Oversight***

The FTC stresses the importance of proper training and workplace protection of AI moderators and auditors. The training should correct human moderators' implicit biases and moderators' tendency to be overly deferential to AI decisions.

The FTC encourages platforms and other internet entities to use algorithmic impact assessments, or AIAs, and audits, as well as document the assessment results in a standardized way. AIAs allow for the evaluation of an AI system's impact before, during or after its use.

Companies can mitigate bad outcomes in time with AIAs, and the FTC and other regulators can obtain information from AIAs for investigations into deceptive and unfair business practices. An audit focuses on evaluation of an AI model's output.

### **Criticism of the Report**

FTC Commissioner Noah Phillips issued a dissenting statement, and Commissioner Christine Wilson also listed several disagreements that she had with the report in a concurring statement. The two commissioners based their criticisms on three grounds.

First, the agency did not solicit sufficient input from stakeholders. The dissenting commissioners perceive the FTC report as a literature review of academic articles and news reports on AI.

They note that the report's authors did not consult any internet platforms about how they view AI efficacy, and they find that the report frequently cites to the work and opinions of current FTC employees, holding that the quantity of self-reference calls the objectivity of the report into question.

Second, they believe that the report's recommendation might produce the countereffect of subjecting compliant entities to FTC enforcement actions.[3]

Third, they conclude that the report's negative assessment of AI use in combating online harms lacks foundation. They find that conclusions of AI inefficiency are sometimes based on the fact that harmful contents are not completely eliminated by AI tools.

The dissenting commissioners say that the report lacks a cost-benefit analysis of whether the time and money saved by using AI tools to combat harmful contents outweigh the costs of the AI tools missing some percentage of these contents.

### **Key Takeaways**

AI has tremendous benefits that companies leverage every day. But when doing so, it is prudent to be mindful of the FTC's cautions, and take steps to fortify practices related to AI.

### ***Data Minimization***

The FTC is not against implementing innovative AI tools to prevent frauds or fake reviews. However, the agency encourages data minimization.

Companies should collect only the information necessary to provide the service or product. And companies should tailor data collection to their need to render services or products.

### ***Transparency***

The FTC may require social media platforms and other internet entities to disclose sufficient information to allow consumers to make intelligible decisions about whether to and how to use certain platforms.

The FTC may also require entities to grant researchers access to information and algorithms, to a certain extent.

### ***Accountability***

The FTC may hold platforms and other internet entities responsible for impact of their AI tools, especially if the AI harms the rights of marginalized groups — even if the tools are intended for combating harmful contents.

### ***Human Oversight***

Companies should enhance human oversight. The FTC may encourage standardization of appropriate training of AI moderators/auditors and enhancement of their workplace protection.

### ***Consumer Privacy***

Companies should take care to refrain from invasive consumer surveillance. Consumer privacy interests outweigh accuracy and utility of AI tools.

### ***Free Speech Concerns***

Companies should be cautious about potential free speech disputes when prebunking alleged misinformation.

## **Further Guidance**

The FTC may conduct more research on using AI to combat online harms. Its guidance may be subject to significant change, based on the sources it decides to consult.

---

*Anthony E. DiResta and Kwamina Thomas Williford are partners, and Wanqian Zhang is a summer associate, at Holland & Knight LLP.*

*The opinions expressed are those of the author(s) and do not necessarily reflect the views of the firm, its clients or Portfolio Media Inc., or any of its or their respective affiliates. This article is for general information purposes and is not intended to be and should not be taken as legal advice.*

[1] In legislation enacted in 2021, Congress directed the FTC to examine ways that AI "may be used to identify, remove, or take any other appropriate action necessary to address online harms." See Statement of Commissioner Alvaro M. Bedoya Regarding Report to Congress on Combatting Online Harms Through Innovation, FTC (June 16, 2022) (acknowledging that in the 2021 Appropriations Act, Congress asked the commission to report on the use of AI to detect or address harmful online content including fake reviews, opioid sales, hate crimes and election-related disinformation).

[2] Commissioners Christine Wilson and Noah Phillips are concerned about "prebunking misinformation" recognized as effective in the report. Both point out in their statements that prebunking information that is not verifiably false, but may be false, might create free speech issues. See Dissenting Statement of Commissioner Noah Joshua Phillips Regarding the Combatting Online Harms Through Innovation Report to Congress and Concurring Statement of Commissioner Christine S. Wilson Report to Congress on Combatting Online Harms Through Innovation, FTC Public Statements (June 16, 2022).

[3] In 2021, the FTC brought a case against an ad exchange company for violations of the Children's Online Privacy Protection Act and Section 5 of the FTC Act. The company claimed to take a unique human and technological approach to traffic quality, and employed human review to assure compliance with its policies and to classify websites. The company's human review failed. But it was only the human review that provided the "actual knowledge" needed for the commission to obtain civil penalties under COPPA. If the company had relied entirely on automated systems, it might have avoided monetary liability. *U.S. v. OpenX Technologies Inc.*, Civil Action No. 2:21-cv-09693 (C.D. Cal. 2021).