

# FTC Report Warns Against Overconfidence in AI Tools to Combat Online Harm

**Minding Your Business Blog** on July 13, 2022

Last month, the FTC issued a [report](#) to Congress advising governments and companies to exercise “great caution” in using artificial intelligence (“AI”) to combat harmful online content. The report responds to Congress’s request to look into whether and how AI may be used to identify, remove, or otherwise address a wide variety of specified “online harms.” Among the “harms” covered by Congress’s request were impersonation scams, fake reviews and accounts, deepfakes, illegal drug sales, revenge pornography, hate crimes, online harassment and cyberstalking, and misinformation campaigns aimed at influencing elections.

The FTC concluded that the adoption of AI to address these harms could itself introduce additional harms because the tools can be inaccurate, biased, and discriminatory by design. They are also vulnerable to hacking and manipulation, and may rely on increasingly invasive forms of commercial surveillance. These findings add another dimension to the FTC’s recent warnings against using biased algorithms that produce discriminatory outcomes and fuel misinformation campaigns.

For instance, the report focused on automated deletion tools, which may cause unfair outcomes due to unrepresentative datasets, faulty classifications, failure to identify new phenomena, missing context, and flawed design. On the one hand, automated deletion tools can result in false positives or overblocking, reducing freedom of expression and access to valued content, while on the other hand, underblocking can do the same by allowing online harassment to silence its targets.

To avoid these outcomes, companies may consider the following suggestions offered in the FTC report when deploying AI tools to combat online harms:

(1) **Human oversight** to monitor use and decisions of AI tools. Moderators should have adequate training, time, and agency to make decisions, be diverse, and have an understanding of many different cultures and languages.

(2) **Transparency and accountability.** Provide consumers with “meaningful information” about AI tools such that they are “explainable and contestable.” Implement consumer feedback mechanisms as well as independent audits and algorithmic impact assessments.

(3) **Data scientists** and their employers are responsible for inputs and outputs and therefore should hire and retain diverse teams to avoid unintentionally biased or unfair results and avoid using training data and classifications that “reflect existing societal and historical inequities.”

(4) Use machine learning tools to address harmful content **beyond just detection and removal**, such as to mitigate the impact of certain harmful content and to uncover the networks and actors behind the harm.

(5) **Provide individuals** with AI tools so they can limit their personal exposure to certain harmful or otherwise unwanted content.

(6) Provide **greater access to AI tools** intended to combat online harm. This is important for smaller platforms and organizations that may not have the resources to create them. However, the report cautions that this must be balanced against the risks in sharing algorithms, one of them being that the more widely a detection tool is shared the easier it is for bad actors to exploit it.

(7) Supplement detection and removal tools with **authentication tools** that can identify the source of particular content and whether it has been altered. This is important given the challenges of keeping detection technology at a level equal to that of deepfake technology.

(8) Develop **legal frameworks** to help ensure the use of AI tools does not itself cause harm, including laws relating to algorithmic transparency and accountability, as opposed to creating legislation that requires companies to deploy AI tools to detect harmful content.

While the report and its suggestions are somewhat limited in scope, the report serves the FTC’s broader goal to expand the public’s understanding of the consumer protection and competition risks associated with AI and algorithms and demonstrates the FTC views AI as an incomplete solution.

[View Original](#)

**Related Professionals**

---

- **Amy B. Gordon**  
Associate