FUTURE OF
PRIVACY
FORUM

1350 Eye Street NW, Suite 350, Washington, DC 20005 | 202-768-8950 | fpf.org

Filed online at regulations.gov

May 29, 2024

Dr. Laurie E. Locascio
Director and Under Secretary of Commerce for Standards and Technology
National Institute of Standards and Technology
100 Bureau Drive
Gaithersburg, MD 20899

Re: NIST AI 100-4, Reducing Risks Posed by Synthetic Content: An Overview of Technical Approaches to Digital Content Transparency

The Future of Privacy Forum (FPF) welcomes the opportunity to submit comments in response to the National Institute of Standards and Technology (NIST)'s draft for public comment on NIST AI 100-4, "Reducing Risks Posed by Synthetic Content: An Overview of Technical Approaches to Digital Content Transparency."[1] NIST AI 100-4 was drafted in response to the White House Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, directing NIST to develop a report regarding standards, tools, methods, and practices for authenticating, labeling, or detecting synthetic content.[2] FPF is a global non-profit organization dedicated to advancing privacy leadership, scholarship, and principled data practices in support of emerging technologies.

The growth of widely-available generative artificial intelligence (GenAI) tools has brought both exciting benefits as well as heightened risks, including harassment, mis/disinformation, malicious impersonation,[3] and the increased production of child sexual abuse material (CSAM) and non-consensual intimate images (NCII).[4] These risks are likely to disproportionately impact those from marginalized communities, who may face even greater harm in the event that they are targeted. In finalizing NIST AI 100-4, and in further recommending a strategy for addressing harms associated with synthetic content, FPF recommends that NIST should:

[1] National Institute for Standards and Technology, *Reducing Risks Posed by Synthetic Content: An Overview of Technical Approaches to Digital Content Transparency* (Apr. 2024), https://airc.nist.gov/docs/NIST.AI.100-4.SyntheticContent.ipd.pdf.

[2] The White House, *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* (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/.

[3] Electronic Privacy Information Center (EPIC), *Generating Harms: Generative AI's Impact & Paths Forward* (May 2023), https://epic.org/wp-content/uploads/2023/05/EPIC-Generative-AI-White-Paper-May2023.pdf. *See also* EPIC, *Generating Harms II: Generative AI's New & Continued Impacts* (May 2024), https://epic.org/wp-content/uploads/2024/05/EPIC-Generative-AI-II-Report-May2024-1.pdf.

[4] *Supra* 1.

I. Collaborate and align with other federal agencies addressing issues related to AI-driven synthetic content, impersonation, and fraud.

II. Ensure any recommended technical approaches to digital content transparency appropriately account for privacy and security implications, and retain the flexibility to evolve with technological developments.

NIST should collaborate and align with other federal agencies addressing issues related to AI-driven synthetic content, impersonation, and fraud

To ensure NIST's recommended approach to addressing harmful synthetic content is compatible with other similar government efforts, NIST should coordinate with other federal agencies that are already working on issues like AI-driven impersonation and fraud. Notably, the Federal Trade Commission (FTC) has recently proposed a trade regulation rule generally prohibiting the impersonation of individuals.[5] As FPF mentioned in its comment to the FTC, the advent of GenAI tools has made it easier for malicious actors to engage in impersonation using synthetic content, necessitating technical and regulatory safeguards to prevent and mitigate harms associated with AI-driven fraud.[6] Additionally, in response to the White House's AI Executive Order, the National Science Foundation and Department of Energy established a research coordination network to advance research on privacy-enhancing technologies (PETs),[7] and the Department of the Treasury produced a report on managing fraud and cybersecurity harms stemming from AI in the financial services sector.[8] Each of these efforts is likely to address the creation and/or use of synthetic content in some way, and it is imperative they deal with this topic consistently to avoid unnecessary confusion in a new and rapidly evolving issue area.

By coordinating with other government agencies addressing similar issues, NIST can ensure that any final documents and recommendations it develops are consistent, cohesive, and interoperable. The strategies that NIST and other agencies develop are the results of in-depth investigations, requiring different equities like privacy, safety, security, and technological development to be weighed. In harmonizing their strategies, agencies can avoid providing conflicting recommendations, presenting a unified approach to AI governance that is clear to both industry and the public.

[5] Federal Trade Commission, *Supplemental notice of proposed rulemaking; request for public comment: Trade Regulation Rule on Impersonation of Government and Businesses* (Mar. 1, 2022), https://www.federalregister.gov/documents/2024/03/01/2024-03793/trade-regulation-rule-on-impersonation-of-government-and-businesses.

[6] Future of Privacy Forum, *FTC SNPRM Impersonation Comment* (Apr. 29, 2024), https://www.regulations.gov/comment/FTC-2023-0030-0057.

[7] U.S. National Science Foundation, *NSF and DOE establish a Research Coordination Network dedicated to enhancing privacy research* (Feb. 26, 2024), https://new.nsf.gov/news/nsf-doe-establish-research-coordination-network.

[8] U.S. Department of the Treasury, *Managing Artificial Intelligence-Specific Cybersecurity Risks in the Financial Services Sector* (Mar. 2026), https://home.treasury.gov/system/files/136/Managing-Artificial-Intelligence-Specific-Cybersecurity-Risks-In-The-Financial-Services-Sector.pdf.

NIST should ensure any technical approaches to digital content transparency that it recommends appropriately account for privacy and security implications, and retain the flexibility to evolve with technological developments

Technical approaches to data management necessarily involve tradeoffs. For example, a technique like provenance data tracking, while useful for combating certain harms associated with synthetic content by tracking and verifying the origin of the data, may also reveal personal data if deployed without appropriate safeguards. Because provenance reveals the origin and history of digital content, it can make personal information available to actors beyond whom an individual intended to share the data. In other cases, provenance data tracking can in fact promote privacy, by facilitating accountability and compliance with privacy preferences and policies, strengthening access and usage controls, and protecting against data leakage.[9] As NIST notes in the draft report, data does not need to be personal in nature to reveal sensitive information. In some cases metadata can provide information about content's "properties, structure, origin, purpose, time and date of creation, author, location, standards, file size, quality, versions, editing history, and other details."[10] Similarly, digital watermarking may also reveal personal information, if such data is embedded within the watermark.[11]

While NIST's draft report correctly recognizes the need to balance the utility of provenance techniques with privacy and data protection, more research is needed to develop successful methods for doing so. For instance, some studies have examined the possibility of creating secure, privacy-preserving data provenance techniques, but significantly more examination is needed to bring these concepts to maturity.[12] Additionally, it is currently unclear how to balance certain content transparency mechanisms—such as collecting and sharing metadata—with privacy and data protection mechanisms such as an individual's right to control or delete data about them. NIST, in collaboration with other agencies and organizations, should explore the potential ways PETs could contribute to privacy-preserving provenance techniques, in line with recommendations in the White House's Executive Order on AI.[13]

[9] Elisa Bertino et. al., *A roadmap for privacy-enhanced secure data provenance*, Journal of Intelligent Information Systems, Vol. 13 (May 31, 2014), https://profsandhu.com/journals/misc/jiis_provenance_2014.pdf.

[10] *Supra* 1.

[11] Center for Democracy & Technology, *Privacy Principles for Digital Watermarking* (May 2008), https://cdt.org/wp-content/uploads/copyright/20080529watermarking.pdf.

[12] Bofeng Pan, Natalia Stakhanova, and Suprio Ray, *Data provenance in security and privacy*, ACM Computing Surveys, Vol. 55, Iss. 14s (Jul. 17, 2023), https://cyberlab.usask.ca/papers/ACMSurvey23.pdf.

[13] The White House, *FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence* (Oct. 30, 2023), https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/.

In that respect, it is also important that NIST's recommendations account not only for changing approaches but also for changing technologies, in order to remain relevant in a rapidly-evolving technological landscape. For example, immersive technologies that combine elements of both the physical and digital world may be more prone to impersonation and manipulation than other online spaces due to their more realistic, embodied nature.[14] Studies have shown that people may have difficulty distinguishing human faces from computer-generated faces, and even find computer-generated faces *more* trustworthy than real ones.[15] Experiences in immersive environments can be so visceral that people forget the content they're interacting with is virtual, not physical,[16] and the large amount of personal data available on the Internet makes it easier to target attacks to specific individuals. As a result, AI-driven impersonations and disinformation campaigns may be more effective in immersive environments, creating both individual and societal-level risks. It's important that any technical, organizational, and regulatory approaches to the harms of synthetic content acknowledge the potentially different ways these issues might manifest in immersive technologies, and how the technology itself should evolve based on new standards needed to mitigate harms in new technological contexts.

FPF appreciates the opportunity to comment on these issues, and NIST's ongoing efforts to address the harmful impacts of AI-driven synthetic content. NIST's technical recommendations should form part of a comprehensive, holistic strategy for tackling synthetic content harms. We welcome any further opportunity to provide resources or information to assist in this vital effort. If you have any questions regarding these comments and recommendations, please contact Jameson Spivack at jspivack@fpf.org (cc: info@fpf.org).

Sincerely,

Jameson Spivack, Senior Policy Analyst, Immersive Technologies

The Future of Privacy Forum
https://fpf.org/

[14] "Immersive technologies" refer generally to a collection of hardware and software products that substitute, enhance, or alter users' individual, physical-world experiences. This includes extended reality (XR), virtual world and social gaming applications, neurotechnologies, and other related technologies. *Supra* 6.

[15] Louis Rosenberg, *Evil twins and digital elves: How the metaverse will create new forms of fraud and deception*, Big Think (Apr. 25, 2022), https://bigthink.com/the-future/metaverse-fraud-digital-twins/.

[16] Jameson Spivack and Daniel Berrick, *Immersive Tech Obscures Reality. AI Will Threaten It*, WIRED (Sep. 27, 2023), https://www.wired.com/story/immersive-technology-artificial-intelligence-disinformation/.