CENTER FOR DEMOCRACY & TECHNOLOGY

October 17, 2023

Re: Consumer Privacy Should Be Featured at the Hearing on *Safeguarding Data and Innovation: Building the Foundation for the Use of Artificial Intelligence*

Representatives McMorris Rodgers, Pallone, Bilirakis, and Schakowsky,

We commend the Subcommittee on Innovation, Data, & Commerce of the House Energy & Commerce Committee for hosting a hearing focused on Artificial Intelligence (AI) and privacy. This hearing reinforces this Committee's commitment to privacy and civil rights, and comes at a time when Congress has been leading an important and growing effort on developing AI policy.

Comprehensive federal privacy legislation is a foundational pillar of AI governance and is essential for responsible, rights-respecting AI innovation. Last year, this Committee passed the American Data Privacy and Protection Act (ADPPA) with strong bipartisan support. We urge the Committee again to consider and pass that legislation, which would provide individuals with critical privacy protections and provide a key underpinning for Congress's further work on AI.

AI presents or exacerbates a number of privacy issues. Among those issues is the vast datasets containing personal identifiable information on which AI models, such as Generative AI models, are often trained.[1] That data may come from multiple sources: for example, it may be publicly available, acquired from companies that specialize in developing AI training sets or from data brokers, or first party data collected directly by the company engaged in training an AI model.[2] That data is often collected, processed, and transferred without the knowledge or permission of individuals whose data is included or any other transparency. Clearview AI, for example, took it upon itself to collect billions of images online of people's faces and built an AI system that can identify essentially anyone, creating the "Google of facial recognition."[3] Large collections of training data may also be a "honey pot" that attracts hackers and other malicious actors. Data minimization requirements such as those included in

[1] Joe McKendrick, *The Data Paradox: Artificial Intelligence Needs Data; Data Needs AI*, Forbes (June 27, 2021), https://www.forbes.com/sites/joemckendrick/2021/06/27/the-data-paradox-artificial-intelligence-needs-data-data-needs-ai.

[2] *See, e.g.*, AI & ML Training Data, Datarade, https://datarade.ai/data-categories/ai-ml-training-data (last visited October 17, 2023).

[3] Nilay Patel, *Clearview AI and the End of Privacy, with Author Kashmir Hill*, Verge (Oct. 17, 2023), https://www.theverge.com/23919134/kashmir-hill-your-face-belongs-to-us-clearview-ai-facial-recognition-privacy-decoder. A private version of this database exists as well, called PimEyes. Bobby Allyn, *'Too Dangerous': Why Even Google Was Afraid to Release This Technology*, NPR (Oct. 11, 2023), https://www.npr.org/2023/10/11/1204822946/facial-recognition-search-engine-ai-pim-eyes-google.

ADPPA can help address some of the privacy harms that can arise from the indiscriminate collection and use of training data.

AI can exacerbate the harms that result from how a person's data is used. Many companies now use AI-driven systems to make decisions about who is hired, who receives a loan, or who is approved for housing.[4] These AI systems make decisions about candidates with little, if any, transparency and no clear standards for appropriate or fair design. Time and again, we have seen such systems discriminate against older people, women, people of color, and other under-represented groups based on inferences made from their data. For instance, Xerox once famously analyzed its employees' likelihood of retention and found that workers who lived far from the office were more likely to quit. Realizing that workers with a longer commute time were those from lower-income neighborhoods, the company had to adopt a conscious policy *not* to screen job candidates based on commuting time because it would have systematically discriminated against them.[5] This example shows how people's private data (in this case, their addresses) can give rise to unfair treatment by automated systems. AI systems must be transparent so people know what factors the system is considering and can challenge the fairness and appropriateness of those factors, as well as robustly tested for bias and other potential harms.[6]

With the power of AI, the harms related to targeting of content also could significantly increase. Consumers are already targeted based on their online and offline behavior for ads and other content, and the ease, speed, and scale with which AI functions will make such personalized content more frequent, intrusive, and harmful. For instance, an AI system may flag a consumer researching weight loss, and then may target that person with any number of personalized predatory ads ranging from harmful drugs to extreme diets—not based on the most effective medical intervention, but based on which company is willing to pay the most money for that ad impression, as well as a complicated array of AI-powered predictions about that person.[7] Generative AI could also enable easy and cheap creation

---

[4] *See, e.g.*, Emmanuel Martinez & Lauren Kirchner, *The Secret Bias Hidden in Mortgage-Approval Algorithms,* Markup (Aug. 25, 2021), https://themarkup.org/denied/2021/08/25/the-secret-bias-hidden-in-mortgage-approval-algorithms; Pranshu Verma, *AI Is Starting to Pick Who Gets Laid Off,* Wash. Post (Feb. 20, 2023), https://www.washingtonpost.com/technology/2023/02/20/layoff-algorithms.

[5] Lauren Weber, *Growing Use of Tests Sparks Scrutiny Amid Questions of Effectiveness and Workplace Discrimination*, ProgramBusiness (Sept. 30, 2014), https://programbusiness.com/news/Growing-Use-of-Tests-Sparks-Scrutiny-Amid-Questions-of-Effectiveness-and-Workplace-Discrimination.

[6] *See* Testimony of CDT CEO Alexandra Givens, Before the U.S. House of Representatives Energy & Commerce Committee, Mar. 1, 2023, https://cdt.org/wp-content/uploads/2023/02/HHRG-118-IF17-Wstate-GivensA-20230301-final.pdf, at 13.

[7] *See generally* Liza Gak, Seyi Olojo, & Niloufar Salehi, *The Distressing Ads That Persist: Uncovering the Harms of Targeted Weight-Loss Ads Among Users with Histories of Disordered Eating* at 11 (2022), https://arxiv.org/abs/2204.03200 ("[O]nce users indicate interest in topics 'relevant' to dieting and weight loss, and are part of 'relevant' demographics, they are inundated with weight-loss ads.").

of personalized "phishing emails" seeking to entice consumers into giving away sensitive information such as passwords or account numbers.[8]

The need for copious amounts of data to train AI systems also presents national security risks in the absence of comprehensive privacy protection. Today, adversarial or competing foreign nations can easily purchase detailed information about Americans from data brokers and use that data to train their AI models, as well as to target Americans with personalized, AI-generated content.

The first step in protecting against these and other AI-related harms is to pass comprehensive privacy legislation such as ADPPA. ADPPA was negotiated extensively among many stakeholders and is a strong, bipartisan privacy bill that can bring about real protections and change for consumers. These protections extend to AI, including ADPPA's requirements related to data minimization, protecting civil rights, and algorithmic impact assessments.

Thank you for the effort and time you have invested on privacy and AI. We look forward to engaging with the Committee and Subcommittee on these important issues.

Respectfully submitted,

Eric Null
Co-Director, Privacy & Data Project
Center for Democracy & Technology

Samir Jain
Vice President of Policy
Center for Democracy & Technology

[8] Jessica Lyons Hardcastle, *AI-Generated Phishing Emails Just Got Much More Convincing,* Register (Jan. 11, 2023), https://www.theregister.com/2023/01/11/gpt3_phishing_emails.