CENTER FOR DEMOCRACY & TECHNOLOGY

Statement of Samir Jain
Vice President of Policy, Center for Democracy & Technology

U.S. Senate AI Insight Forum: Privacy & Liability
November 8, 2023

Senators, thank you for inviting me to this AI Insight Forum. I am the Vice President of Policy at the Center for Democracy & Technology (CDT), a 28-year old nonprofit, nonpartisan organization that works to protect users' civil rights, civil liberties and democratic values in the digital age. One of our key areas of focus is protecting user privacy, in contexts from commercial data practices to government surveillance.

My key point is straightforward: **the most important step Congress can and should take to protect people's privacy in the context of AI is to pass legislation—ideally in the form of comprehensive privacy legislation but otherwise as part of AI legislation—that shifts the burden of protecting privacy away from individuals and to the companies that collect and profit from individuals' data**. Such legislation has long been overdue, but is even more urgent today. AI presents a number of privacy-related harms, and thus basic privacy protections for all are essential for responsible, rights-respecting AI innovation. A privacy law will not curb AI innovation. Rather, it will help ensure that the technology is developed in accordance with American values, and engenders the trust necessary to promote the adoption and use of AI.

AI-Related Privacy Harms

Data collection has run rampant in the digital age in large measure because companies have economic incentives to amass large pools of data, such as using data to target advertising based on people's behavior. The need for large datasets to train AI systems provides yet another reason for companies to collect or repurpose extensive data about everyone online. That data may come from multiple sources: for example, it may be publicly available, be acquired from companies that specialize in developing AI training sets or from data brokers, or be first party data collected directly by the company engaged in training an AI model. That data is often collected, processed, and transferred solely at the direction of the company, without any attention paid to potential privacy harms. 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."[1] Large collections of training data may also be a "honey pot" that attracts hackers, foreign adversaries, and other malicious actors.

[1] Nilay Patel, *Clearview AI and the End of Privacy, with Author Kashmir Hill*, Verge (Oct. 17, 2023). 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).

The power of AI also threatens to exacerbate the harms related to targeting of advertising and other content. 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, all without the company or the individual knowing how it happened.[2] Generative AI can enable easy and cheap creation of seemingly authentic personalized "phishing emails" seeking to entice consumers into giving away sensitive information such as passwords or account numbers.[3]

AI also can compound the harms that result from how a person's data is used. Many companies now use AI-driven systems that leverage immense amounts of data to make decisions about who is hired for a job, who receives a loan, or who is approved for housing. 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.[4] As this example illustrates, people's private data (in this case, their addresses) can give rise to unfair treatment by automated systems.

Generative AI and applications such as chatbots can also raise privacy concerns. They can, for example, effectively "memorize" personal information contained in the training data. If left unchecked, this can result in such personal information being included in the output of generative AI systems and revealing such information to unauthorized third parties.[5] Moreover, while search queries can be highly revealing, the prompts and other information that users enter when interacting with generative AI applications can reveal even more intimate personal details,

---

[2] *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).
[3] Jessica Lyons Hardcastle, *AI-Generated Phishing Emails Just Got Much More Convincing,* Register (Jan. 11, 2023).
[4] Lauren Weber, *Growing Use of Tests Sparks Scrutiny Amid Questions of Effectiveness and Workplace Discrimination*, ProgramBusiness (Sept. 30, 2014).
[5] Earlier this year Samsung banned the use of ChatGPT after its employees entered confidential source code as part of a query and that code was then transferred to OpenAI's servers and incorporated into its training data. Siladitya Ray, *Samsung Bans ChatGPT Among Employees After Sensitive Source Code Leak*, Forbes (May 2, 2023). OpenAI subsequently rolled out ChatGPT Enterprise for companies that have at least 150 users, which does not use conversations for training by default. Cloey Callahan, *OpenAI hopes ChatGPT Enterprise will answer employers' data privacy concerns*, Digiday (Sept. 15, 2023). Consumer queries, however, are still incorporated into training data unless they activate a setting to turn off chat history. OpenAI, *New Ways to Manage Your Data in ChatGPT* (Apr. 25, 2023).

such as health symptoms and financial details. As a result, the use or disclosure of those interactions can cause significant privacy harms.

Finally, at a time when we are concerned that adversarial or competing nations may outrace us in developing sophisticated AI systems, we are in fact helping them due to the absence of comprehensive privacy legislation. Those 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.

Steps Congress Can Take

1. Privacy Protections. Congress should help alleviate these and other AI-related privacy harms by passing comprehensive privacy legislation. Such legislation must move from a notice–and-consent regime—which relies on the fiction that users read lengthy privacy policies and make informed choices about their data—to one that places responsibility on the companies that collect, use, and profit from individuals' data.

In the absence of comprehensive privacy legislation, **any AI legislation should at minimum encompass protections that would address the privacy harms created and exacerbated by use of AI**. Key privacy provisions that should be incorporated in any AI bill include:

- Data minimization—the principle that companies should collect and process only the data needed to provide the service or product they are offering. A robust data minimization provision would result in less personal data sloshing through the ecosystem that can then end up in training datasets or otherwise be used in ways that harm individuals.
- Heightened protections for the collection, use, and transfer of sensitive information such as biometric, genetic, and location data.
- Appropriate limitations on targeted advertising, such as a ban on such advertising to individuals known to be minors and a right to opt out of such ads.
- An obligation not to collect, use, or transfer data in ways that discriminate on the basis of protected classes.
- A national registry of data brokers and a centralized mechanism for individuals to request deletion of their data and opt out of collection of further information.
- As part of any impact assessment or auditing framework for AI systems, an assessment of how the company minimized the data it used in its training datasets and an assessment of how the company ensured that the model is not biased against protected classes.

2. Protections for Training Data. AI models need to be trained on data to be effective; and that training data needs to be appropriately representative. At the same time, users need protections

with respect to personally identifiable information and sensitive information that may be included in training datasets. For this, Congress must establish baseline protections.

Protections for training data cannot rely on the outdated model of notice and consent. A training data set based only on data from individuals who have opted in or not opted out may be biased or skewed (e.g., if individuals of a certain demographic were less likely to opt in or more likely to opt out). Moreover, relying on notice and consent as the primary method to protect privacy in connection with creating training datasets risks replicating the very weaknesses of today's privacy regime. Instead, **model developers themselves should have the obligation to take steps to reduce privacy risks posed by the use of large-scale training data for AI systems**. For example, organizations can be far more thoughtful about the data used to train AI systems in the first place. Developers should exclude sources of data known to have a significant amount of identifiable information and exclude particularly sensitive forms of data such as biometric data.

Automated tools can be developed to scrub instances of personally identifiable information from training datasets, such as full names, email addresses, Social Security numbers, and credit card information.[6] Machine learning engineers are also exploring ways to incorporate privacy enhancing technologies into training AI systems, such as using differential privacy (adding noise to the model training process to make sure private information is not inadvertently memorized and reflected in the output) or using encrypted data. Synthetic data is increasingly utilized as an alternative source to real-world data. Synthetic data, although artificial, statistically mimics the patterns and characteristics of real-world data and thus allows AI systems to learn key relationships, while not "memorizing" real personally identifiable information.

Developers also should conduct red team testing to determine if models they built memorized or are at risk of revealing private information. For example, consider an AI model designed to recommend HIV treatments. Under what is termed a "membership inference" attack, an adversarial attacker could determine that a particular person was (or at least likely was) part of the training dataset for the model, thereby revealing their positive HIV status. Testing can reveal whether a model is at risk of such an attack and, if so, developers should then take steps to mitigate that risk.

Congress should ensure that federal AI research and government-supported standards processes advance these and other types of methods for protecting privacy and testing for privacy risks. It should also consider whether laws, particularly in sectors such as health and finance that inherently involve sensitive information, should **require that model developers**

[6] While useful, such tools remain imperfect: they will tend to reflect similar gaps as AI models themselves (e.g., performing better on common data formats and in majority languages), and so must be combined with additional efforts to reduce the surface area of privacy risks.

take these types of reasonable steps based on current, state-of-the-art techniques in order to protect individuals' privacy.

*3. **Accountability and Liability Framework.*** Given the range of privacy and other harms that AI can cause, accountability is a critical piece of AI governance. One key way of ensuring accountability is the promulgation of laws and regulations that set standards for AI systems and impose potential liability for violations. Such liability provides for redress for harms suffered by individuals, creates incentives for AI system developers and deployers to take reasonable steps to minimize the risk of those harms from occurring in the first place, and helps to engender trust in the use and deployment of AI.

Congress need not create a new liability regime out of whole cloth. Existing laws, from civil rights laws barring discrimination in multiple sectors to the prohibition on unfair and deceptive trade practices enforced by the Federal Trade Commission to torts such as fraud and defamation, continue to apply even if AI is involved. In some respects, however, these existing laws may not be fit for purpose. For example, in some cases, the entities developing and deploying AI are not always readily recognized as entities that traditionally have been covered under existing civil rights and consumer protection laws. This ambiguity helps entities responsible for AI harms claim that existing laws do not apply to them. Furthermore, due to the lack of transparency around the use of AI, a person who is harmed may not have the information needed to even know if they have a claim. They may not know whether or how an AI system was used, let alone have information about training data, how a system works, or what role it plays.

Both as to existing laws and new laws, a further complication is apportioning liability across the AI value chain. In the case of generative AI, for example, different actors may be the compiler of training datasets, the creator of a model trained with those datasets, the developer of a chatbot or other application built on top of that model, and the deployer of that application to consumers. Liability may lie with different actors depending on factors such as the cause of the harm at issue and which actor was in the best position to have taken reasonable steps to prevent that harm.

Ultimately, **a robust enforcement and liability regime is a necessary element of AI governance**, both to create the appropriate incentives for companies to prevent and mitigate harms and to provide compensation to those who are harmed. Congress should examine the effectiveness of existing laws, consider legislative additions and updates as needed, and (through oversight and appropriations) ensure that regulatory agencies are providing the industry guidance that this moment requires.