Amba Kak

Additional Questions for the Record Response

Subcommittee on Innovation, Data, and Commerce Hearing titled,

“Safeguarding Data and Innovation:

Building the Foundation for the Use of Artificial Intelligence”

December 11, 2023

Attachment—Additional Questions for the Record

The Honorable Debbie Dingell

Privacy should be of utmost importance to this Committee and our country, and it is a fundamental aspect in the digital age. We need to act now. The status quo in this ever-changing landscape is insufficient. And relying on self-regulation has proven repeatedly to be inadequate. Privacy – as we all know – is a fundamental right, and we must protect our families. This Committee's job is to ensure that consumers have the final say over their personal data. These privacy implications also put at risk the security of our personal information that can easily be exploited by our adversaries, like China and Russia. It is crucial that Congress addresses this. Artificial intelligence is here and continuing to evolve. It has major implications on our health and safety, economy, and national security.

1. Ms. Kak, we are seeing a state-by-state patchwork of data privacy laws, can you elaborate on the importance of enacting comprehensive federal legislation in this space?

2. Ms. Kak, at the state level, have you seen any provisions that you recommend this Committee look at to keep our consumers safe? Are there state-level provisions that you've seen have unintended consequences?

Thank you for these questions. We urgently need a strong national baseline privacy standard to protect the privacy of all Americans. Ideally, such comprehensive privacy legislation sets a federal "floor" and allows the states to go even further and broader, without restriction. The bipartisan ADPPA is an example of a strong baseline and I hope it will be reintroduced, strengthened, and signed into law, at a time when AI reinvigorates the urgency around a need for privacy standards.

There are several state-level provisions that should be instructive to development of federal privacy standards, notably strong data minimization mandates as well as support for a private right of action. The recently passed updates to the California Consumer Privacy Act include a strong data minimization mandate that puts clear limits on the data that companies are permitted to collect and use, and how long they store this data for. These restrictions, based on the necessity and proportionality of data use, are in relation to specific product or service requested by the individual, rather than leaving it to companies to decide what they deem to be "necessary" in a discretionary manner. Private causes of action are particularly potent levers for meaningfully deterring companies from engaging in unlawful conduct. Illinois' BIPA is a shining example of how a private right of action complements a strong data minimization mandate. BIPA draws clear lines prohibiting entities from ever profiting from biometric data, and requires a high standard of consent from collection. Combined with a private right of action, the BIPA has resulted in banning the now notorious company Clearview AI from selling its sensitive database of millions of our faces illegally scraped from the internet for profit.

3. Ms. Kak, could the information collected and used by AI companies heighten the risks to women who make these difficult, private, and personal health care decisions?

Absolutely. AI technologies heighten and transform the nature of risks from personal and sensitive information (and inferences). Firstly, algorithms can be applied to draw out sensitive inferences from seemingly routine and disconnected categories of information. The privacy risks are most acute in the case of inferences relating to healthcare decisions, including decisions relating to reproductive health care that disproportionately impact women and gender minorities. Second, as a general matter, the incentives for predatory and invasive data surveillance already exist, but the proliferation of AI pours gasoline on the problem. Finally, there are several AI-enabled reproductive healthcare applications that are collecting and retaining information that could be used to target and harass women if in the wrong hands. These apps such as menstrual tracking apps or birth control assistants that collect and analyze large amounts of directly sensitive data have been found to have lax security practices and vague consent forms.[1] Regulators are beginning to take action motivated by rampant harms that disproportionately affect women – the FTC's recent case against Kochava, a data broker collecting geolocation data from millions of mobile devices, and enabling inferences about their visits to reproductive health centers that could compromise them if in the wrong hands.

4. Ms. Kak, does the proliferation of AI systems and particularly generative AI systems incentivize companies to collect, process, and transfer user data unnecessary to provide a specific product or service? Why might this be bad?

Soon after the public release of chatGPT, questions from the public about what data these AI models had been trained on began to circulate,[2] followed by panic when people began to realize that chatGPT was sometimes leaking personal data "accidentally" in response to prompts. This example was not a one-off: there are ongoing privacy and security challenges introduced by large-scale AI systems. Regardless of the training procedure, guardrails, and use of anonymization in data inputs, certain AI systems can unpredictably produce highly sensitive outputs, including personally identifiable information, that pose foundational privacy problems.

At a more systemic level too, the incentives for predatory and invasive data surveillance already exist, but the proliferation of AI pours gasoline on the problem. The emphasis on scale with AI -and shoring up very large datasets - means we risk a race to the bottom with our privacy as collateral damage. Beyond threatening individual and collective privacy, this indefinite retention of data raises major security concerns. We already have examples of the real human costs of careless retention of data, from biometric information of Afghan citizens in American-managed databases that fell into the hands of the Taliban, to the intricate web of third-party data brokers that buy and sell sensitive information about people that can be used to target them unfairly or to hinder their access to credit, housing, and education.Information that's never collected in the first place cannot be breached, and that which is deleted after it's no longer needed, is no longer at risk. Otherwise we risk creating more and more "honey pots" or "goldmines for cyber criminals" that are an attractive target for interception by unauthorized third parties, including malicious state and non-state actors.

[1] https://foundation.mozilla.org/en/blog/in-post-roe-v-wade-era-mozilla-labels-18-of-25-popular-period-and-pregnancy-tracking-tech-with-privacy-not-included-warning/

[2] Clothilde Goujard, "Italian privacy regulator bans ChatGPT," *Politico*, March 31, 2023, https://www.politico.eu/article/italian-privacy-regulator-bans-chatgpt/.

5. Ms. Kak, can you outline the ways in which comprehensive privacy legislation would serve to shield American families from the excessive collection, known as overcollection, and potential misuse of their personal data?

Yes, taking the ADPPA as an example, these are a few of many more tools we *would have* in hand to regulate AI:

- Data minimization, which would mitigate the supercharged incentives for firms to excessively hoover up data on users
- Data rights, which could compel transparency into these increasingly opaque AI systems
- Or its civil rights provisions to address the structurally discriminatory outcomes of AI systems (boosting what federal agencies are already doing under existing laws to curb algorithmic discrimination).

6. Ms. Kak, can you elaborate on the importance of implementing a data minimization requirement to protect consumers?

Data minimization rules don't hinge on user consent: they apply regardless, overcoming the now well known deficiencies of a privacy regime that hinges exclusively on individuals being able to meaningfully exercise choices online given the structural power asymmetries between individuals and massive tech firms that abound. This is particularly important in contexts such as workplace surveillance, where the entities deploying increasingly invasive 'productivity monitoring' and other AI-enabled measures have significant power over those on whom such systems are deployed, rendering 'consent' meaningless.

Beyond the broad principle, data minimization mandate can include prohibitions on specific kinds of data use that have well known harms, such as prohibiting targeted advertising to children or the use of data about people's interior mental states in so-called "emotion recognition" systems that have been repeatedly demonstrated as being based on faulty foundations. As we, Accountable Tech and EPIC emphasize in the 'Zero Trust AI Framework', data minimization rules are essential levers at a time when AI is tipped to further exacerbate information asymmetries between individuals and communities, on the one hand, and the large corporations that create and collect data about them which has increasing power over their lives, on the other.

<u>The Honorable Lori Trahan</u>

1. Ms. Kak, in your testimony, discussed the use of audits and impact assessments to proactively identify and mitigate harms. At what stages of the AI development and deployment process would impact assessments and audits be appropriate, and how should they be implemented?

For algorithmic audits and impact assessments to be meaningful, they must be done across the lifecycle of AI development (including, and especially, before the development and development of the system) *and* must also be done on an ongoing (rather than one off) basis. Early impact assessment is important in order to encourage active deliberation on data and design choices, when modifications are possible, and before harm has transpired. As part of this initial risk assessment, developers and deployers must document planned and foreseeable modifications that they will make based on the risks or harms that have surfaced as part of this process. In fact, documentation should be a core part of any evaluation mandate – it is a mode to encourage reflexivity as well as provides enforcement agencies with information on the basis of which they can investigate harms when they do occur.

This also means that providers across the supply chain for AI development need to have tailored obligations that are appropriate to the control they exercise over key data and design choices. For example, providers of foundation models that are then made available to application developers via API are best placed to mitigate and correct risks as they maintain control over the underlying model irrespective of the end use case.

There must also be ongoing evaluation, especially post-deployment or wide availability of the system. Without this, there is a real likelihood of missing those harms that only materialize in the specific contexts of use or failing to recognize emergent behaviors from complex models that surface over time with use.