Stanford University
Human-Centered
Artificial Intelligence

Department of Commerce
National Telecommunications and Information Administration
Docket No. 230407-0093
RIN 0660-XC057
AI Accountability Policy Request for Comment

June 11, 2023

To Whom It May Concern:

Thank you for the opportunity to provide comments on the NTIA's AI Accountability Policy RFC. My comments narrowly address Questions 16, 20, 22, 25 and focus explicitly on privacy and data accountability issues. I submit these comments on behalf of myself and provide my affiliation for identification purposes only. For context, I am the Privacy and Data Policy Fellow at the Stanford Institute for Human-Centered Artificial Intelligence, where my research focuses on consumer privacy, privacy and data policy, and artificial intelligence.

The RFC asks a straightforward question in #25: Is the lack of a general federal data protection or privacy law a barrier to effective AI accountability?

My response to this question is an emphatic "yes!". The lack of a federal level law that provides privacy and data protection rights to consumers creates inconsistency with privacy issues in the technology sector generally, and also specifically makes it challenging to respond to developments in AI that threaten privacy. There are certainly data accountability measures that can be taken absent a federal level privacy and data protection law that address issues of data quality, bias, and data reliability with dataset development. However, the lack of individual rights to delete, correct, limit, prevent sale of, and opt-out of or into data collection and processing across the entire data ecosystem leaves individuals vulnerable and presages a narrow, sectoral approach to protecting individuals' data privacy similar to the present federal sectoral approach that has left consumers widely unprotected. Furthermore, a lack of restraint on businesses' collection, use, and reuse or sale of data encourages a race to the bottom with regards to practices that exploit consumer data. Regulating data only for the purposes of applications in artificial intelligence makes no sense given that the vast majority of data that powers the development of AI today was collected or generated in a multiplicity of contexts that may have no relationship to the AI products in which it is used.

Questions 16, 20, and 22 ask about accountability mechanisms for data. Question 16 asks about where in the lifecycle accountability mechanisms should be focused, while

Question 20 focuses on records, and 22 on data quality and data voids. Before addressing any of these topics, NTIA staff should first be aware of the excellent work of Dr. Mehtab Khan and Dr. Alex Hanna on dataset accountability measures Their 2022 paper, "The Subjects and Stages of AI Dataset Development: A Framework for Dataset Accountability," presently provides an excellent overview of many of the emergent mechanisms to address nascent dataset accountability concerns as well as a framework for assessing who is implicated by informational harms at every stage of the dataset creation process.

To consider Q16, the bulk of the discussion about AI harms has focused on the output side of the AI process, and proportionately little has focused on the input side. The explosion of generative AI in the first half of 2023 raised the stakes on the input side of the equation. Prior to the release and substantial public awareness of large language models such as ChatGPT, BARD, and visual AI tools such as DALL-E, MidJourney, Lensa, and others, privacy discussions were primarily centered on the use of data collected by large platforms or other forms of proprietary data; the company most publicly implicated for widespread public data scraping was Clearview AI, which as been subject to litigation the US and sanctions by EU data protection authorities. The quick emergence of generative AI tools has raised awareness that many sources of data beyond facial images are being scraped and collected to build these tools, and that the public has little insight into what these sources are. In the EU, data protection authorities are grappling with whether data that might be considered publicly accessible and not subject to regulation in the US violates data protection statutes (in the case of Clearview AI, the answer has been affirmative).

Even so, the proposed EU AI Act does not account for issues of dataset accountability outside of the boundaries of the GDPR, and while it will certainly impact how companies gather the data of European citizens, as currently written it will not fully address questions of data provenance, quality, consent, and transparency when it comes to what data is being used to power all forms of AI, not just generative AI. It is unclear today where many companies are sourcing their data from, and especially true with those using data scraped from the web. To address the data quality question noted in Q22, this uncertainty raises the question of whether a transparency measure that required addressing data provenance, including issues of consent as well as copyright, would prevent or ameliorate privacy harms, copyright violations, as well as aid in detecting bias and improving quality. For example, transparency requirements could aid in assessing the extent to which highly toxic, misogynistic, and racist sites are scraped and included as source material into datasets. While there can be valid reasons for including such material, for general applications such inclusion is questionable, and assumptions that the sheer amount of data included in these massive datasets will either average out toxic sources or that companies can control their emergence through model fine tuning simply isn't foolproof. Greater transparency around the proportion of such material in training datasets could potentially provide a clearer understanding of the impact of such material on a model's outputs.

However, the questions raised above elucidate the fact that like all questions around AI regulation, returning to Q16 and Q20 (records) there continues to be uncertainty about how best to proceed, suggesting the need for researchers to work with policymakers to ascertain which methods hold the most promise, and what types of recordkeeping would allow others to audit a system and gain sufficient knowledge of how the datasets used inform the model's outputs. Presently, most of the accountability mechanisms discussed in Khan and Hanna's paper have been developed by those with the greatest access to data: employees and researchers at large technology companies. This fact does not suggest these approaches are invalid, but rather illustrates the point that it is still early days in this space; because access to data and compute lies primarily in the hands of large tech companies, both academics and civil society groups have had little opportunity to expand upon these mechanisms and proposals with actual experience and empirical evidence. There needs to be greater access to models, data, and compute, whether it be through public sector investment such as the NAIRR, open collaborations by industry with academia and civil society, as well as regulatory sandboxes for piloting many of these proposed mechanisms to ensure they provide measurable and meaningful results.

What many in the privacy community can confidently say today is that without both individual consumer privacy rights in place as well as rules that place limits and accountability mechanisms on the systematic collection and uses of data that are beyond the capacity of individual rights to address, the development of datasets for AI applications will trample on both individual and societal-level expectations of information privacy. A right to request the deletion of one's personal data, such as which presently exists under California's Consumer Privacy Act (CCPA), is a limited tool when pitted against hundreds or even thousands of actors collecting data, and the burden of making such requests relies upon the individual. A federal level law creating a similar right will be an improvement, but as long as the burden of these requests rests on the shoulders of individuals, individual rights cannot make a sufficient dent in expansive data collection. Privacy and data protection legislation must strike a balance between individual rights, societal interests in privacy, as well as clarify what constitutes public and freely collectible data and what violates this balance between these interests.

While the focus of these comments has largely been on the input side of dataset creation, there are also substantial issues with data collected by AI models through user interactions, and whether this data is used for retraining or reinforcement learning. Research has demonstrated that models can memorize and regurgitate the personal data users provide to them. Furthermore, there are urgent questions related to explainability rights (with respect to FTC enforcement of laws such as the Fair Credit Reporting Act as well as the CCPA) and whether AI systems can provide accurate and sufficient explanations of outputs to satisfy explainability and transparency requirements. It is unclear whether having a full accounting of an AI system's training data or model documentation will be able to meet the requirement of providing consumers with clear and accurate reasons for specific decisions.

Nearly three decades into the expansion of the consumer internet, it is safe to say that industry self-regulation with regards to data and privacy practices leads to a race to the

bottom with incentives for companies to collect and monetize as much data as possible. There is no reason to believe that this next phase of technology development will be any different. The sheer amount of data required for the development of AI systems creates incentives to maximize data collection and reuse. Without specific guardrails to compel ethical and privacy-preserving practices with respect to data, we may witness the practices that have proliferated during the past decade of expansive data collection become even more intrusive.

Sincerely,

Dr. Jennifer King
Privacy & Data Policy Fellow,
Stanford Institute for Human Centered Artificial Intelligence