PREPARED TESTIMONY AND STATEMENT FOR THE RECORD
OF

AMBA KAK
EXECUTIVE DIRECTOR, AI NOW INSTITUTE

on behalf of herself and Dr. Sarah Myers West, Managing Director, AI Now Institute

"SAFEGUARDING DATA AND INNOVATION: BUILDING THE FOUNDATION FOR THE USE OF ARTIFICIAL INTELLIGENCE."

BEFORE THE

U.S. HOUSE COMMITTEE ON ENERGY, COMMERCE
SUBCOMMITTEE ON INNOVATION, DATA, AND COMMERCE

Chair Rodgers, Ranking Member Pallone, and Members of the Committee, thank you for inviting me to appear before you and provide testimony on this important issue. My name is Amba Kak and I am the Executive Director of the AI Now Institute and Senior Affiliate Fellow at the Khoury College of Computer Sciences, Northeastern University. AI Now is a policy research organization, founded in 2017, committed to actionable research on artificial intelligence ("AI"). I have over a decade of experience working on technology policy and research in the United States and across multiple other jurisdictions, with a special focus on AI and data privacy law. This testimony is offered on behalf of myself and my colleague Dr. Sarah Myers West and our remarks are based on research we have conducted at AI Now.[1]

I applaud the Committee for taking the initiative to advance this conversation and in particular for recognizing that privacy and innovation are mutually reinforcing goals that can, and must, be advanced in concert. As excitement and trepidation about large scale AI systems continues to fill headlines and hearings, it's important to remember that there is nothing about the current trajectory of these privately developed technologies that is inevitable. It goes without saying that in a democracy, the trajectory of powerful technologies should be shaped in the public interest through public deliberation, not solely by a handful of corporate actors driven, ultimately, by commercial incentives: regulation can play a crucial role in ensuring such democratic shaping of technological systems.

Which brings me to the one overarching point I want to make in today's testimony: *We already have many of the regulatory tools we need to govern and regulate AI effectively, including privacy, consumer protection, and competition frameworks. Now is the time to extend what we have in pursuit of ensuring that our legal regime meets the moment. Specifically, I encourage this Committee to prioritize data privacy—and in particular, the passage of strong, legally enforceable data minimization mandates, already included in legislative proposals such as the ADPPA which has already received this Committee's resounding support. Data protection is a core mechanism that can help mitigate the serious privacy and competition implications of large scale AI.*[2]

In fact, the notion that we need to wipe away years of regulation and policy and create new frameworks from scratch serves large industry players more than it does the rest of us: it serves to delay, and to provide current actors with significant influence on the scope and direction of such policymaking. AI systems are not wholly novel. Far from it. And rather than view them that way, to responsibly govern these technologies we must instead disaggregate these systems, or the "AI stack", into their composite inputs, recognizing the details of how they work and what they require to operate. These include close examination of data, computational infrastructure, or labor. Precise and technically-aware regulatory strategies can then be deployed at different layers of this stack, for example preventing cloud companies from using their

[1] See generally: Amba Kak and Sarah Myers West, "AI Now 2023 Landscape: Confronting Tech Power", AI Now Institute, April 11, 2023, https://ainowinstitute.org/2023-landscape.

[2] "Zero Trust AI Governance", Accountable Tech, AI Now Institute, and EPIC, August 10, 2023, https://ainowinstitute.org/publication/zero-trust-ai-governance.

dominant market position to restrict competition in the AI market, or copyright strategies against use of artistic works by image generation tools, or, as is the subject of this testimony, how data regulation can prevent AI firms from the irresponsible collection and retention of personal information.[3] Once this is done, we can explore whether new approaches to address previously unanticipated harms or to tackle specific sectoral use cases are needed. But before that, we must leverage and continue to strengthen the regulatory toolbox we have already honed over the last decade.

To illuminate this argument, we divide it into three specific points:

First, that data privacy regulation *is* AI regulation and provides many essential tools that we need to govern AI and protect the public from harm. As AI systems proliferate in our social and economic lives, a strong federal privacy law, such as the ADPPA, is an ever more urgent priority.

Second, that, as it stands today, there is no large-scale AI without Big Tech given their stronghold on access to both data and computational infrastructure. This given their combined access to both data and computational infrastructure. If we want a vibrant, innovative and competitive AI ecosystem, then privacy and competition goals must be advanced in concert.

Finally, legally binding data minimization rules that tackle unfettered first-party data surveillance as well as limit secondary uses of data for training AI are a key way forward. Without these, we risk a race to the bottom with consumer privacy and competition as the collateral damage in the AI race.

1. Data privacy regulation *is* AI regulation and provides many essential tools that we need to govern AI and protect the public from harm. As AI systems proliferate in our social and economic lives, a strong federal privacy law, such as the ADPPA, is an ever more urgent priority.

Soon after the public release of chatGPT, questions from the public about what data these AI models had been trained on began to circulate,[4] followed by panic when people began to realize that chatGPT was sometimes leaking personal data "accidentally" in response to prompts.[5] This example was not a one-off: there are ongoing privacy and security challenges introduced by

---

[3] Jai Vipra and Sarah Myers West, "Computational Power and AI", AI Now Institute, https://ainowinstitute.org/publication/policy/compute-and-ai; Tejas Narechania and Ganesh Sitaraman, "An Antimonopoly Approach To Governing Artificial Intelligence", Vanderbilt University, October 6, 2023, https://cdn.vanderbilt.edu/vu-URL/wp-content/uploads/sites/412/2023/10/06212048/Narechania-Sitaraman-Antimonopoly-AI-2023.10.6.pdf.pdf; Jennifer Cobbe, Michael Veale and Jatinder Singh, "Understanding Accountability in Algorithmic Supply Chains," 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23), April 7, 2023, https://ssrn.com/abstract=4430778.

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

[5] Nicholas Carlini et al, "Extracting Training Data from Large Language Models", 30th USENIX security Symposium, December 2020, https://arxiv.org/abs/2012.07805; Nicholas Carlini et al, "Extracting Training Data from Diffusion Models", January 2023, https://arxiv.org/abs/2301.13188; OpenAI, "March 20 ChatGPT outage: Here's what happened," March 24, 2023, https://openai.com/blog/march-20-chatgpt-outage.

large-scale AI systems.[6] 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.[7]

Regulators in countries with data privacy laws were able to act quickly in response: Italy issued a temporary ban on chatGPT to its citizens based on concerns that the mass data collection and storage of data to create the systems had been done in violation of privacy laws.[8] This ban was lifted after OpenAI verified compliance with requests for greater transparency and privacy protective measures be implemented in its systems, including granting users certain opt-out rights, such as being able to toggle off the option for conversations to be used for training ChatGPT algorithms.[9] The Japanese[10], Swiss[11] and Spanish[12] data protection authorities also issued notices following similar enquiries. Others raised concerns that AI systems like ChatGPT were not in compliance with the established "right to be forgotten" guaranteed in the GDPR and other international privacy laws, which guards against inaccurate or misleading information, and provides remedies of erasure.[13]**In stark contrast, the absence of a similar legal framework in the US has limited its ability to swiftly and nimbly respond to this moment, though enforcement agencies are doing all they can using existing authorities and limited resources**.[14] We must make these tools more robust.

But like AI systems themselves, the use of data privacy law to regulate AI far predates this current hype moment. Regulating the collection of certain kinds of sensitive data, like biometrics for example, effectively limits the unchecked proliferation of AI systems that require this data as a necessary input for training and deployment. The Illinois Biometric Information Privacy Act (BIPA) and several other similar state laws, for example, place a strict consent requirement for the collection of face and other biometric data. BIPA has already been used to challenge several

---

[6] James Vincent, "Apple restricts employees from using ChatGPT over fear of data leaks," *The Verge*, May 19, 2023, https://www.theverge.com/2023/5/19/23729619/apple-bans-chatgpt-openai-fears-data-leak; Jeffrey Dastin and Anna Tong, "Focus: Google, one of AI's biggest backers, warns own staff about chatbots," *Reuters*, June 15, 2023, https://www.reuters.com/technology/google-one-ais-biggest-backers-warns-own-staff-about-chatbots-2023-06-15/.

[7] El-Mahdi El-Mhamdi, Sadegh Farhadkhani, Rachid Guerraoui, Nirupam Gupta, Lê-Nguyên Hoang, Rafaël Pinot, Sebastien Rouault, and John Stephan, "On the Impossible Safety of Large AI Models," *ArXiv*, May 9, 2023, https://arxiv.org/pdf/2209.15259.

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

[9] "Italy lifts ban on ChatGPT after data privacy improvements," DW, April 29, 2023, https://www.dw.com/en/ai-italy-lifts-ban-on-chatgpt-after-data-privacy-improvements/a-65469742.

[10] Kantaro Komiya and Sam Nussey, "Japan privacy watchdog warns ChatGPT-maker OpenAI on user data," *Reuters*, June 2, 2023, https://www reuters.com/technology/japan-privacy-watchdog-warns-chatgpt-maker-openai-data-collection-2023-06-02/.

[11] Nicole Beranek Zanon and Monika Abt, "Data protection complaints about the use of ChatGPT in Italy and Switzerland," *Lexology*, May 11, 2023, https://www.lexology.com/library/detail.aspx?g=a9ae0e7f-cf0f-4f58-9836-13861dbdeb96

[12] Reuters Staff, "Spanish data watchdog to investigate potential data breaches by ChatGPT," *Reuters*, April 13, 2023, https://www.reuters.com/article/eu-chatgpt-spain/spanish-data-watchdog-to-investigate-potential-data-breaches-by-chatgpt-idUKL8N36F4GL.

[13] Jess Weatherbed, "OpenAI's regulatory troubles are only just beginning", *The Verge*, May 5 2023 https://www.theverge.com/2023/5/5/23709833/openai-chatgpt-gdpr-ai-regulation-europe-eu-italy.

[14] Natasha Lomas, "ChatGPT resumes service in Italy after adding privacy disclosures and controls," *TechCrunch*, April 28, 2023, https://techcrunch.com/2023/04/28/chatgpt-resumes-in-italy/.

concerning AI systems,[15] including Clearview AI: a company notorious for claiming to have captured more than 10 billion faceprints from peoples' online photos is now permanently banned from making its face database available to most businesses and other private actors because of the settlement in *ACLU v. Clearview AI*.[16] Purpose limitation, i.e. that data use should also be limited or related to the purpose for which it was collected, is another key part of the data minimization standard. The FTC enforced this standard in its recent Amazon Alexa case, where Amazon was fined for violating children's privacy by indefinitely retaining their data and then leveraging such data for improving its Alexa algorithm.[17] In another context, automated hiring and firing of workers, an increasing concern globally with the proliferation of platform-based gig work, has also recently been successfully challenged in the Amsterdam Court of Appeals using the right to demand access to their personal data processed by any organization and to receive meaningful information about the processing of such data, as guaranteed by the GDPR.[18]

So what specific levers do data protection frameworks offer to regulate AI? Taking the American Data Privacy and Protection Act (ADPPA) as an example of a strong baseline standard, this is an illustrative list of what beneficial interventions could be possible if we had a strong federal data privacy law:

a. **Data minimization:** Such a law could enable a proactive obligation on entities to put reasonable limits on the collection, use, and retention of personal data in the interest of the individual and group data holders. These 'data minimization' rules, which are described in the ADPPA as central to the 'duty of loyalty'[19] to individuals, are a core part of global data protection laws globally. These rules cut against prevailing incentives that promote indiscriminate surveillance and data mining and privilege commercial benefits even when they run counter to individual interests.

   Another reason we urgently need a data minimization rule is data security,[20] and this concern is heightened in the age of large scale AI which, barring regulation, will further incentivize reckless collection and retention of sensitive information. We already have examples of the real human costs of careless retention of data, from biometric

---

[15] Woodrow Hartzog, "BIPA: The Most Important Biometric Privacy Law in the US?," in "Regulating Biometrics: Global Approaches and Urgent Questions," AI Now Institute, September 1, 2020, https://ainowinstitute.org/wp-content/uploads/2023/09/regulatingbiometrics-hartzog.pdf.

[16] ACLU, "In Big Win, Settlement Ensures Clearview AI Complies With Groundbreaking Illinois Biometric Privacy Law," press release, May 9, 2022, https://www.aclu.org/press-releases/big-win-settlement-ensures-clearview-ai-complies-with-groundbreaking-illinois. It's worth noting that this Clearview settlement does not limit law enforcement use outside of Illinois. Even in Illinois, law enforcement would be able to use it after 5 years.

[17] Federal Trade Commission, U.S. v. Amazon.com (Alexa), cases and proceedings, July 21, 2023, https://www.ftc.gov/legal-library/browse/cases-proceedings/192-3128-amazoncom-alexa-us-v.

[18] Worker Info Exchange, "Historic digital rights win for WIE and the ADCU over Uber and Ola at Amsterdam Court of Appeal," blog, April 4, 2023, https://www.workerinfoexchange.org/post/historic-digital-rights-win-for-wie-and-the-adcu-over-uber-and-ola-at-amsterdam-court-of-appeal.

[19] *See generally* Neil Richards and Woodrow Hartzog, "A Duty of Loyalty for Privacy Law," *Washington University Law Review* 961 (2021), July 3, 2020, http://dx.doi.org/10.2139/ssrn.3642217.

[20] *See generally:* Federal Trade Commission, "Commission Statement Marking the FTC's 50th Data Security Settlement," statement, January 31, 2014, https://www.ftc.gov/system/files/documents/cases/140131gmrstatement.pdf; Daniel Solove and Woodrow Hartzog, Breached!: Why Data Security Law Fails and How to Improve it, Oxford University Press (2022).

information of Afghan citizens in American-managed databases that fell into the hands of the Taliban,[21] 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.[22] 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"[23] that are an attractive target for interception by unauthorized third parties,[24] including malicious state and non-state actors.

Crucially, 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.[25] 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.[26]

Beyond the broad principle, data privacy laws can include prohibitions on specific kinds of data use that have well known harms, such as prohibiting targeted advertising to children[27] 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

---

[21] Eileen Guo & Hikmat Noori, "This is the real story of the Afghan biometric databases abandoned to the Taliban", MIT Tech Review, August 30 2021, https://www.technologyreview.com/2021/08/30/1033941/afghanistan-biometric-databases-us-military-40-data-points/.

[22] See for example: Federal Trade Commission, "FTC Sues Kochava for Selling Data that Tracks People at Reproductive Health Clinics, Places of Worship, and Other Sensitive Locations", August 29, 2022, https://www.ftc.gov/news-events/news/press-releases/2022/08/ftc-sues-kochava-selling-data-tracks-people-reproductive-health-clinics-places-worship-other.

[23] Dimitri Sirota, "The Art Of Letting Go: How Data Minimization Can Improve Cybersecurity And Reduce Cost," *Forbes*, March 29, 2023, https://www.forbes.com/sites/forbestechcouncil/2023/03/29/the-art-of-letting-go-how-data-minimization-can-improve-cybersecurity-and-reduce-cost/?sh=641958c75340.

[24] For examples of "leaky" data from IoT devices and mobile phones, leaving personal information of users vulnerable to interception, see: Anna Maria Mandalari, Daniel J. Dubois, Roman Kolcun, Muhammad Talha Paracha, Hamed Haddadi, David Choffnes, "Blocking without Breaking: Identification and Mitigation of Non-Essential IoT Traffic," In Proceedings of Privacy Enhancing Technologies Symposium (PETS), 2021. https://doi.org/10.48550/arXiv.2105.05162.

[25] Federal Trade Commission, "Commercial Surveillance and Data Security Rulemaking," notice, August 11, 2022, https://www.ftc.gov/legal-library/browse/federal-register-notices/commercial-surveillance-data-security-rulemaking; David Medine and Gayatri Murthy, "Companies, not people, should bear the burden of protecting data," Brookings, December 18, 2029, https://www.brookings.edu/articles/companies-not-people-should-bear-the-burden-of-protecting-data/.

[26] Wilneida Negrón, "Little Tech Is Coming for Workers: A Framework for Reclaiming and Building Worker Power," Coworker.org, November, 2021, https://home.coworker.org/wp-content/uploads/2021/11/Little-Tech-Is-Coming-for-Workers.pdf

[27] See: American Data Privacy and Protection Act, H.R. 8152, https://docs.house.gov/meetings/IF/IF00/20220720/115041/BILLS-1178152rh.pdf

foundations.[28] 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.[29] We will come back to this in point 3.

b. **Prohibition against use of data in ways that discriminate:** ADPPA includes a prohibition on the use of personal data in ways that discriminate. It is now well documented that AI systems are routinely, and often structurally, biased in ways that entrench and embed historical inequities[30] in sensitive social domains like healthcare,[31] hiring,[32] education,[33] housing,[34] and criminal justice.[35] This should not come as a

---

[28] Access Now, European Digital Rights (EDRi), Bits of Freedom, Article 19, and IT-Pol, May 2022, https://www.accessnow.org/wp-content/uploads/2022/05/Prohibit-emotion-recognition-in-the-Artificial-Intelligence-Act.pdf.

[29] "Zero Trust AI Governance", Accountable Tech, AI Now Institute, and EPIC, August 10, 2023, https://ainowinstitute.org/publication/zero-trust-ai-governance.

[30] Federal Trade Commission, "Joint Statement on Enforcement Efforts Against Discrimination and Bias in Automated Systems," public statement, April 25, 2023, https://www.ftc.gov/legal-library/browse/cases-proceedings/public-statements/joint-statement-enforcement-efforts-against-discrimination-bias-automated-systems; The White House, "Blueprint for an AI Bill of Rights," August, 2022, https://www.whitehouse.gov/ostp/ai-bill-of-rights/; Samir Jain, "CDT and Coalition Urge White House to Ensure Forthcoming AI Executive Order Advances Civil Rights & Civil Liberties," Center for Democracy & Technology, September 5, 2023, https://cdt.org/insights/cdt-and-coalition-urge-white-house-to-ensure-forthcoming-ai-executive-order-advances-civil-rights-civil-liberties/.

[31] Ziad Obermeyer et al., "Dissecting racial bias in an algorithm used to manage the health of populations," Science 366, 447-453 (2019), October 25, 2019, https://www.science.org/doi/10.1126/science.aax2342.

[32] U.S. Equal Employment Opportunity Commission, "Artificial Intelligence and Algorithmic Fairness Initiative," January 23, 2023, https://www.eeoc.gov/ai; Pauline T. Kim, "Data-Driven Discrimination at Work," 58 *William & Mary Law Review* 857 (2017), February 1, 2017, https://scholarship.law.wm.edu/wmlr/vol58/iss3/4; Ifeoma Ajunwa, "Algorithms at Work: Productivity Monitoring Applications and Wearable Technology as the New Data-Centric Research Agenda for Employment and Labor Law," 63 *St. Louis University Law Journal* 21 (2019), September 10, 2018, https://ssrn.com/abstract=3247286; Aaron Rieke and Miranda Bogen, "Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias," Upturn, December 10, 2018, https://www.upturn.org/work/help-wanted/.

[33] Kristin Woelfel, Elizabeth Laird and Maddy Dwyer, "Letter to ED and the White House from Tech Policy, Civil Rights, and Civil Liberties Advocates Calling for Civil Rights Guidance and Enforcement Regarding EdTech and AI," Center for Democracy & Technology, September 20, 2023, https://cdt.org/insights/letter-to-ed-and-the-white-house-from-tech-policy-civil-rights-and-civil-liberties-advocates-calling-for-civil-rights-guidance-and-enforcement-regarding-edtech-and-ai/; Andre M. Perry and Nicol Turner Lee, " AI is coming to schools, and if we're not careful, so will its biases," Brookings, September 26, 2019, https://www.brookings.edu/articles/ai-is-coming-to-schools-and-if-were-not-careful-so-will-its-biases/.

[34] U.S. Justice Department, "Justice Department Secures Groundbreaking Settlement Agreement with Meta Platforms, Formerly Known as Facebook, to Resolve Allegations of Discriminatory Advertising," press release, June 21, 2022, https://www.justice.gov/opa/pr/justice-department-secures-groundbreaking-settlement-agreement-meta-platforms-formerly-known; Lauren Kirchner and Matthew Goldstein, " Access Denied: Faulty Automated Background Checks Freeze Out Renters," *The Markup*, May 28, 2020, https://themarkup.org/locked-out/2020/05/28/access-denied-faulty-automated-background-checks-freeze-out-renters; Ridhi Shetty, " CDT Comments to Federal Agencies Highlight Risks of Data Used in Tenant Screening," Center for Democracy & Technology, June 2, 2023, https://cdt.org/insights/cdt-comments-to-federal-agencies-highlight-risks-of-data-used-in-tenant-screening/.

[35] Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, "Machine Bias," ProPublica, May 23, 2016, https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

surprise, given that these systems necessarily draw their map of "the world" from data that reflects discriminatory histories and sentiments. As recently highlighted in the factsheet accompanying the Biden Administration's Blueprint for an AI Bill of Rights, several federal agencies are already applying existing laws and mechanisms to address algorithmic discrimination in housing, employment, and other opportunities.[36] The ADPPA's civil rights provision would provide an additional means of redress contra AI systems that perpetuate discrimination.

c. **Impact Assessments:** ADPPA also includes a mandate for impact assessments or audits of AI systems in order to proactively identify and mitigate harms, including relating to discrimination, privacy, and security. These evaluations go beyond conventional privacy impact assessments that assess systems against relatively narrow privacy and security criteria, in favor of a more expansive stocktaking that require companies to evaluate whether particular groups will be harmed as a result of the design or use of the AI system. Researchers such as Dr. Alex Hanna and Dr. Mehtab Khan, for example, have put forward a multi-layered framework to scrutinize the multiple complex layers of large scale AI models.[37]

   While such evaluations are positive in theory, we must proceed with a note of caution: there is a significant risk that any audit or evaluation standard can devolve into a superficial checkbox exercise,[38] more useful in offloading liability than in protecting the public. This, unless it is structured deliberately to avoid such a trap:

   - Meaningful assessments that mandate evaluation happen *before* products are made available or in use in the public domain, and are subject to evaluation on an ongoing basis while in operation. It is essential that the criteria for such evaluations are not limited to narrow technical parameters or be tested only under so-called "laboratory-like conditions".[39]

---

[36] The White House, "Blueprint for an AI Bill of Rights," August, 2022, https://www.whitehouse.gov/ostp/ai-bill-of-rights/.

[37] Mehtab Khan and Alex Hanna, "The Subjects and Stages of AI Dataset Development: A Framework for Dataset Accountability," Forthcoming 19 *Ohio State Technology Law Journal* (2023), September 13, 2022, http://dx.doi.org/10.2139/ssrn.4217148.

[38] Amba Kak and Sarah Myers West, *Algorithmic Accountability: Moving Beyond Audits*, AI Now Institute, April 11, 2023, https://ainowinstitute.org/publication/algorithmic-accountability; Sasha Costanza-Chock, Inioluwa Deborah Raji and Joy Buolamwini, "Who Audits the Auditors? Recommendations from a field scan of the algorithmic auditing ecosystem," FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, Pages 1571–1583, June 2022, https://doi.org/10.1145/3531146.3533213. Professor Woody Hartzog refers to audits and similar procedural interventions as "privacy half measures" that are necessary but wholly insufficient in protecting users, see *Hearing On "Oversight Of A.I.: Legislating On Artificial Intelligence" Before the Subcommittee On Privacy, Technology, And The Law,* U.S. Senate Committee On The Judiciary, September 12, 2023, (Statement of Woodrow Hartzog), https://techpolicy.press/wp-content/uploads/2023/09/2023-09-12_pm_-_testimony_-_hartzog.pdf.

[39] Ben Green and Lily Hu, "The Myth in the Methodology: Towards a Recontextualization of Fairness in Machine Learning," 35th International Conference on Machine Learning, 2018, https://econcs.seas.harvard.edu/files/econcs/files/green_icml18.pdf; Shira Mitchell, Eric Potash, Solon Barocas, Alexander D'Amour, and Kristian Lum, "Algorithmic Fairness: Choices, Assumptions, and Definitions," *Annual Review of Statistics and Its Application* 8 (2021): 141–163, https://doi.org/10.1146/annurev-statistics-042720-125902; and Rodrigo Ochigame, "The Long History of Algorithmic Fairness," *Phenomenal World,* January 30, 2020, https://www.phenomenalworld.org/analysis/long-history-algorithmic-fairness/.

- Evaluations must be conducted by independent, disinterested and adequately resourced and protected third parties such as researchers, civil society, or the appropriate federal agencies, by charging that such evaluations are subject to both regulatory and public scrutiny.
- There must be real consequences for a failure to mitigate or prevent harms that are identified. This includes strict penalties but also, crucially, abandoning systems that are designed in ways that make such harms inevitable.

d. **Individual data rights:** Finally, ADPPA also includes a suite of data rights that allow individuals the right to access, correct, port, and delete their information (along with a private right of action). Data rights are a crucial complement to the proactive obligations of data minimization, as they empower individuals to ascertain the nature and scale of commercial surveillance, and to act on such information to correct, order deletion, or otherwise seek redress if they believe any other obligations owed to them under the Act have not been fulfilled. In California, under the California Consumer Privacy Act, individuals are empowered to require businesses to share what information they hold about them, opt-out of the sale of their information, to ask for the deletion of such information, and even sue a business directly if it fails to implement reasonable security measures and their data is compromised in a breach.[40] Companies like Walmart have already reported 55,351 requests under CCPA to stop the sale of personal information, 16,375 to access, and 2,542 to delete personal information– a large majority of these requests have been fulfilled.[41]

2. As it stands today, there is no large-scale AI without Big Tech given their stronghold on access to both data and computational resources. To prevent further concentration of power in the AI industry, privacy and competition goals must be advanced in concert.

Large-scale AI depends principally on data and compute resources (this includes both hardware components such as chips, as well as cloud computing) as essential inputs. Big Tech companies are already positioned at a considerable advantage at many points in the AI stack. Currently, the largest consumer technology companies such as Google, Microsoft, and Amazon dominate access to such compute resources (and other companies, as a rule, depend on them for these resources).[42] This is closely related to these companies' pre-existing data advantage, which enables them to collect and store large amounts of good-quality data about millions and billions of people via their vast market penetration. This data advantage can give models developed by Big Tech companies an edge over those developed without the benefit of such data. Even if alternative models do avail themselves of similar computational power. Indeed, access to high quality data can result in smaller models (those trained on less data and requiring less

[40] EPIC, California Consumer Privacy Act (CCPA) https://epic.org/california-consumer-privacy-act-ccpa/#:~:text=Sample%20form%3A,here%5D%20has%20collected%20about%20me.
[41] Walmart, "How Many California Consumer Privacy Act Requests Did We Fulfill Last Year?," https://corporate.walmart.com/privacy-security/california-privacy-rights/metrics.
[42] Jai Vipra and Sarah Myers West, "Computational Power and AI", AI Now Institute, https://ainowinstitute.org/publication/policy/compute-and-ai.

computational power for training) that perform better than larger models trained without such quality data. OpenAI has reportedly already used YouTube data to train its models, which leaves the door open for Google to use data not only from YouTube, but from Gmail, Google Drive, and all its other services.[43] Similarly, Microsoft can potentially use data from its enterprise services, and AWS from its cloud services. Each of these companies has also forged partnerships and acquisitions in specific sectors that give them access to troves of sensitive data, such as in the electronic health records space.[44] Repositories of publicly available data currently available online, is also likely to soon dwindle or become less valuable in comparison to proprietary datasets held by these companies. This is because the publicly available data will already have been used, and because newly produced data on the internet is starting to be protected more by platforms who recognize its value and want exclusive access. We're already seeing this happen – Reddit, Stack Overflow and X have already implemented some protections against free use of data from their platforms.[45]

In fact, today's AI boom should be understood as driven at its core by commercial data surveillance, that led to the infrastructural dominance of a small handful of firms across our digital lives.[46] Unlike other actors that must largely rely on third-party intermediaries to access data, large firms are exploiting the fact that they directly control the vast majority of the environment in which data is collected: they are able to take advantage of the network effects associated with the scale at which they operate by collecting, analyzing, and using data within platforms they wholly own and control.[47] This is a product of a self reinforcing feedback loop, which over time has led to these firms being so dominant and pervasive that it is virtually impossible *not* to use their systems.[48]

The push to build AI at larger and larger scale only increases the demand for the very same resources that these firms have steadily accumulated and are best positioned to further

---

[43] Jon Victor, 'Why YouTube Could Give Google an Edge in AI', *The Information*, 14 June 2023, https://www.theinformation.com/articles/why-youtube-could-give-google-an-edge-in-ai.

[44] Karen Weise, "Amazon to acquire One Medical clinics in latest push into health care," *New York Times*, July 21, 2022, https://www.nytimes.com/2022/07/21/business/amazon-one-medical-deal.html; Tina Reed, "Google Cloud announces Epic partnership," *Axios*, November 14, 2022, https://www.axios.com/2022/11/14/google-cloud-announces-epic-partnership; Epic, "Epic and Microsoft Bring GPT-4 to EHRs," May 5, 2023, https://www.epic.com/epic/post/epic-and-microsoft-bring-gpt-4-to-ehrs/.

[45] Mike Isaac, "Reddit Wants to Get Paid for Helping to Teach Big A.I. Systems," *New York Times*, April 18, 2023, https://www.nytimes.com/2023/04/18/technology/reddit-ai-openai-google.html; see @XDevelopers, March 29, 2023, https://x.com/XDevelopers/status/1641222782594990080?s=20; Paresh Dave, "Stack Overflow Will Charge AI Giants for Training Data", Wired, April 20, 2023, https://www.wired.com/story/stack-overflow-will-charge-ai-giants-for-training-data/.

[46] Meredith Whittaker, "The Steep Cost of Capture," Interactions 28, no. 6 (November–December 2021): 51, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4135581.

[47] Lina M. Khan, "Sources of Tech Platform Power," Georgetown Law Technology Review 2, no.2, (2018): 325–334, https://georgetownlawtechreview.org/wp-content/uploads/2018/07/2.2-Khan-pp-225-34.pdf; Lina M. Khan, "The Separation of Platforms and Commerce," Columbia Law Review 119, no. 4 (May 2019): 973–1098, https://columbialawreview.org/content/the-separation-of-platforms-and-commerce.

[48] Kashmir Hill, "I Tried to Live without the Tech Giants. It Was Impossible," *New York Times*, July 31, 2020, https://www.nytimes.com/2020/07/31/technology/blocking-the-tech-giants.html.

consolidate.[49] While all others are increasingly relegated to the status of clients and renters, buying and leasing resources from the handful of Big Tech companies that control them. This market reality must inform any privacy and AI-specific regulatory efforts. Privacy and competition law are too often siloed from one another,[50] leading to interventions that could easily compromise the objectives of one issue over the other.[51] And firms are, in turn, taking advantage of this to amass information asymmetries that contribute to further concentration of their power.[52]

3. **Legally binding data minimization rules that tackle unfettered first-party data collection as well as limit secondary uses of data for training AI are a key way forward. Without these, we risk a race to the bottom with consumer privacy and competition as the biggest losers.**

With these points in hand, I would conclude by highlighting that data minimization rules are particularly potent levers to address both privacy and competition harms that are likely to be exacerbated as AI systems proliferate.[53] This includes both the general mandates that limit excessive or unanticipated collection, use, and retention of data as well as more specific restrictions such as regulating secondary uses of data collected from consumers in one context for the purpose of training AI models.

The FTC has already outlined this principle in its recent Amazon Alexa case,[54] and the Commission's Advanced Notice of Proposed Rulemaking (ANPRM) on Commercial Surveillance also contemplates similar data minimization rules.[55] Several civil society organizations including EPIC, Accountable Tech, and the Center for Democracy & Technology have endorsed legislative proposals that would encode data minimization mandates, including

---

[49] Sarah Myers West, "Competition authorities need to move fast and break up AI," Financial Times, April 17, 2023, https://www.ft.com/content/638b5be7-fab7-4fe6-a0cf-7dabefcdd722.

[50] Udbhav Tiwari, "Competition should not be weaponized to hobble privacy protections on the open web, April 2022, Mozilla, https://blog mozilla.org/netpolicy/2022/04/12/competition-should-not-be-weaponized-to-hobble-privacy-protections-on-the-open-web/.

[51] Maurice E. Stucke, "The Relationship between Privacy and Antitrust," Notre Dame Law Review Reflection 97, no. 5 (2022): 400–417, https://ndlawreview.org/wp-content/uploads/2022/07/Stucke_97-Notre-Dame-L.-Rev.-Reflection-400-C.pdf.

[52] For example Article 5 of the European Union's Digital Markets Act which prohibits large "gatekeeper" platforms from the cross-use of personal data between its various service offerings, without explicit user consent; European Commission, "The Digital Markets Act: ensuring fair and open digital markets," https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en.

[53] Amba Kak and Sarah Myers West, *Data Minimization*, AI Now Institute, April 11, 2023, https://ainowinstitute.org/spotlight/data-minimization.

[54] Federal Trade Commission, U.S. v. Amazon.com (Alexa), cases and proceedings, July 21, 2023, https://www.ftc.gov/legal-library/browse/cases-proceedings/192-3128-amazoncom-alexa-us-v.

[55] Federal Trade Commission, "FTC and DOJ Charge Amazon with Violating Children's Privacy Law by Keeping Kids' Alexa Voice Recordings Forever and Undermining Parents' Deletion Requests," May 31, 2023, https://www.ftc.gov/news-events/news/press-releases/2023/05/ftc-doj-charge-amazon-violating-childrens-privacy-law-keeping-kids-alexa-voice-recordings-forever.

restricting the use of data for targeted advertising,[56] or a narrower version that limits the use of sensitive data for all secondary purposes, including advertising;[57] restricting the collection and use of biometric information for particular groups such as children;[58] and in certain contexts such as workplaces,[59] and schools.[60]

The key lesson of the last decade has been understanding that control over data is about power asymmetries, and since companies have clear commercial benefit from widening this asymmetry, regulation is essential to protect the public from harms. As we recently argued, alongside Accountable Tech and EPIC, if we want the future of AI to protect civil rights, advance democracy, and improve people's lives, we must fundamentally change the incentive structure that shapes AI development. Passing strong federal privacy legislation is a critical and overdue step in that direction.[61]

---

[56] Accountable Tech, "Accountable Tech Petitions FTC to Ban Surveillance Advertising as an 'Unfair Method of Competition'," press release, September 28, 2021, https://accountabletech.org/media/accountable-tech-petitions-ftc-to-ban-surveillance-advertising-as-an-unfair-method-of-competition; Congresswoman Anna G. Eshoo, "Eshoo, Schakowsky, Wyden, Booker Introduce Bill to Ban Surveillance Advertising," press release, September 18, 2023, https://eshoo.house.gov/media/press-releases/eshoo-schakowsky-wyden-booker-introduce-bill-ban-surveillance-advertising.

[57] Electronic Privacy Information Center (EPIC) and Consumer Reports, How the FTC Can Mandate Data Minimization through a Section 5 Unfairness Rulemaking, January 2022, https://epic.org/documents/how-the-ftc-can-mandate-data-minimization-through-a-section-5-unfairness-rulemaking.

[58] Lindsey Barrett, "Ban Facial Recognition Technologies for Children—And for Everyone Else," Boston University Journal of Science and Technology Law 26, no. 2 (2020): 223–285. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3660118.

[59] Worker Rights: Workplace Technology Accountability Act, A.B. 1651 (California Legislature, 2021–2022 Regular Session), January 13, 2022.

[60] Stefanie Coyle and Rashida Richardson, "Bottom-Up Biometric Regulation: A Community's Response to Using Face Surveillance in Schools," in "Regulating Biometrics: Global Approaches and Urgent Questions," AI Now Institute, September 1, 2020, https://ainowinstitute.org/wp-content/uploads/2023/09/regulatingbiometrics_Bottom-Up-Biometric-Regulation.pdf.

[61] "Zero Trust AI Governance", Accountable Tech, AI Now Institute, and EPIC, August 10, 2023, https://ainowinstitute.org/publication/zero-trust-ai-governance.