PREPARED TESTIMONY AND STATEMENT FOR THE RECORD OF

Amba Kak
Co-Executive Director, AI Now Institute

on behalf of herself and Dr. Sarah Myers West, Co-Executive Director, AI Now Institute

"The Need to Protect Americans' Privacy and the AI Accelerant"

Before the

U.S. Senate Committee on Commerce, Science, and Transportation

Chair Cantwell, Ranking Member Cruz, and esteemed Members of the Committee, thank you for inviting me to testify on this important set of issues. I deeply appreciate this Committee for taking the initiative to spotlight this urgently needed conversation, and in particular for recognizing that privacy and AI innovation are mutually reinforcing goals that can, and must, be advanced in concert. My name is Amba Kak, and I co-lead the AI Now Institute, a leading policy research institute founded in 2016 that focuses on the social and economic impacts of artificial intelligence technologies. I have spent over fifteen years as a global policy expert designing and advocating for technology policy in the public interest, examining topics ranging from privacy to competition to algorithmic accountability, across roles in government, industry, and civil society. I recently served as a senior advisor on artificial intelligence at the Federal Trade Commission, where my role was to provide technological expertise in support of the agency's enforcement and policy work, focused on how to mitigate and redress harms from data-driven systems like AI. 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]

As excitement and trepidation about large-scale AI systems continues to fill headlines and hearings, it's important to remember that nothing about the current trajectory of these privately developed technologies is inevitable. 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: the trajectory of AI is at a crucial inflection point. Without regulatory intervention, we are doomed to replicate the extractive, invasive, and often harmful data practices and business models that have characterized the past decade of the tech industry. A federal data privacy law, especially one with strong data minimization, could act as a foundational intervention to break this cycle and challenge the culture of impunity and recklessness that is hurting both consumers and competition.

In fact, the notion that we need to wipe away years of regulation and policy and create new frameworks from scratch for AI 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, and labor. Precise and technically aware regulatory strategies can then be

[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.

deployed at different layers of this stack, preventing cloud companies from using their dominant market position to restrict competition in the AI market, for example; or copyright strategies against use of artistic works by image-generation tools; or, as is the subject of this testimony, AI firms from the irresponsible collection and retention of personal information.[2] Once this is done, we can explore whether new approaches to address previously unanticipated harms or to tackle specific sectoral use cases are needed. Before that, though, we must leverage and continue to strengthen the regulatory toolbox we have already honed over the past decade.

To illuminate my argument, I will divide it into three specific points:

First**,** privacy risks are implicated across the AI life cycle. The generative AI boom further unleashes new forms of familiar privacy harms, supercharges the incentives for irresponsible data surveillance, and creates conditions ripe for extractive and exploitative business models.

***Second**,* the turn toward large-scale AI further consolidates Big Tech's already staggering control over consumer data, which deepens power asymmetries and allows these companies to act recklessly and with impunity. A strong data minimization rule would ensure not only the advancement of privacy, but would also act as a powerful curb on the concentration of power we've seen in this sector.

Finally, a legally binding data privacy mandate, including strong data minimization, individual data rights, algorithmic impact assessments, and protections against algorithmic discrimination, offers a foundational toolkit for demanding accountability from AI companies.

I. **Privacy risks are implicated across the AI life cycle.** The generative AI boom further unleashes new forms of familiar privacy harms, supercharges the incentives for irresponsible data surveillance, and creates conditions ripe for extractive and exploitative business models.

In the wake of a highly charged AI race with companies rushing to release new products and features to market before competitors, we're seeing a sharp uptick in privacy lapses, from unexpected leaks of personal information in chatbot outputs[3] to features that threaten to

[2] See Jai Vipra and Sarah Myers West, "Computational Power and AI," AI Now Institute, September 27, 2023, https://ainowinstitute.org/publication/policy/compute-and-ai; Tejas Narechania and Ganesh Sitaraman, "An Antimonopoly Approach to Governing Artificial Intelligence," Vanderbilt Policy Accelerator for Political Economy and Regulation, 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; and 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.

[3] Jordan Pearson, "ChatGPT Can Reveal Personal Information from Real People, Google Researchers Show," *Vice*, November 29, 2023, www.vice.com/en/article/88xe75/chatgpt-can-reveal-personal-information-from-real-people-google-researchers-show.

fundamentally compromise the privacy and security of our personal devices.[4] While the list of egregious and obvious privacy failures is already long, we also need a systematic approach that brings into view every stage of the AI data life cycle as well as captures the more structural pathologies set in motion by the AI boom.

In this vein, we look into privacy harms that (1) emanate at the training and development stage, (2) emanate at the application and output stage, and those that (3) emanate from the business models and incentives shaping the AI market.

(I) ***Training and development stage***. AI models are trained on large amounts of data, and the early stages of training and then fine-tuning models can set in motion some of the most harmful and far-reaching data practices.

While there is a lack of basic transparency about the datasets used to train many commercially available models today, we know that at least some have taken advantage of publicly available data, scraping the web to create massive datasets of images and text, as well as voice and video data. In 2009, Meta changed its settings so that much previously private user data became public by default while users scrambled to revert to their original settings.[5] A month later, Mark Zuckerberg disingenuously argued that privacy was no longer the "social norm."[6] Statements like this subvert the most basic privacy expectations of citizens whose digital lives are hoovered up by firms for profit, shielded by broad and inscrutable terms of service and settings that can be changed without people's consent. Only through public scandals like Cambridge Analytica in 2018 did the public become aware that it had to contend with the dangers of this kind of centralized data power in the hands of a few companies.[7]

Developments in generative AI have brought into sharp focus the stakes of this free-for-all approach to mining the public sphere. Soon after the public release of ChatGPT, questions from the public about what data these AI models had been trained on began to circulate,[8] followed by panic when people began to realize that ChatGPT was sometimes leaking personal data "accidentally" in response to prompts.[9]

We're also seeing Big Tech firms store and use data collected in one context for other unanticipated purposes, using AI as a catchall justification. Companies haven't given clear

[4] Zak Doffman, "Google Confirms Serious AI Risks for iPhone and Android Users," *Forbes,* February 15, 2024, www.forbes.com/sites/zakdoffman/2024/02/12/google-warns-as-free-ai-upgrade-for-iphone-android-and-samsung-users.

[5] Nick Bilton, "'He Doesn't Believe in It': Mark Zuckerberg Has Never Cared about Your Privacy, and He's Not Going to Change," *Vanity Fair*, November 20, 2018, https://www.vanityfair.com/news/2018/11/mark-zuckerberg-has-never-cared-about-your-privacy.

[6] Bobbie Johnson, "Privacy No Longer a Social Norm, Says Facebook Founder," *Guardian*, January 10, 2010, https://www.theguardian.com/technology/2010/jan/11/facebook-privacy.

[7] Ibid.

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

[9] See 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; and OpenAI, "March 20 ChatGPT Outage: Here's What Happened," March 24, 2023, https://openai.com/blog/march-20-chatgpt-outage.

answers to the question of whether or not they're using internal data to train new AI models,[10] and Meta and Google recently announced an update to their terms that explicitly allows training of AI from user data.[11] These fundamental choices of what data to use to train AI models determine the likelihood of inaccurate and discriminatory outputs. Recent research demonstrates that as AI models scale, using larger and larger datasets like the ones used in current LLMs, the tendency to produce inaccurate and harmful stereotypes also scales.[12] Meanwhile, a decade of evidence on predictive AI systems now bears out the "garbage in, garbage-out" thesis, which holds that inaccurate, incomplete, and discriminatory training datasets go on to produce decisions or recommendations in high-stakes domains with harmful consequences for people's lives.[13]

(2) ***Applications, outputs, and decision-making stage.*** Downstream, we see a new range of privacy threats culminate as AI models are applied to consumer-facing applications, or used in systems that aid or make decisions, recommendations, or inferences that impact people's lives in material ways.

Generative AI systems currently on the market have been unexpectedly and routinely leaking personal information that is traced back to training datasets, including sensitive or even confidential data.[14] While generative AI companies advise their users not to include personal information in their prompts, many still do;[15] more concerningly, research suggests that LLM-powered systems like ChatGPT are capable of making detailed and sensitive inferences even from apparently anonymized prompts.[16] Mindful of these unresolved and persistent privacy challenges, many of the largest technology firms have banned their employees from using services like ChatGPT.

With a scramble to rush to market and a lack of regulatory friction, we're seeing multiple AI companies announce untested and potentially harmful applications of AI that rely on people's sensitive information—including biometrics—to make questionable inferences. For example, on the occasion of OpenAI's recent rollout of Sora, a chatbot that provides multimedia output in response to prompts, CEO Sam Altman claimed that the software could detect emotional states from people's voice recordings—even as there is mounting evidence (acknowledged by

[10] Cordilia James, "Are Instagram and Facebook Really Using Your Posts to Train AI? What to Know," *Wall Street Journal*, June 21, 2024, https://www.wsj.com/tech/ai/meta-ai-training-instagram-facebook-explained-a3d36cdb.

[11] Eli Tan, "When the Terms of Service Change to Make Way for A.I. Training," *New York Times*, June 26, 2024, https://www.nytimes.com/2024/06/26/technology/terms-service-ai-training.html.

[12] Abeba Birhane et al, "On Hate Scaling Laws For Data-Swamps", June 28 2023, https://arxiv.org/abs/2306.13141.

[13] See Heather Rodroguez, "Garbage In, Garbage Out: The Potential Pitfalls of Artificial Intelligence," Texas A&M University College of Arts and Sciences, January 19, 2023, artsci.tamu.edu/news/2023/01/garbage-in-garbage-out-the-potential-pitfalls-of-artificial-intelligence.html; and Joan M. Teno, "Garbage In, Garbage Out—Words of Caution on Big Data and Machine Learning in Medical Practice," JAMA Forum, February 16, 2023, https://jamanetwork.com/journals/jama-health-forum/fullarticle/2801776.

[14] Lily Hay Newman, "ChatGPT Spit Out Sensitive Data When Told to Repeat 'Poem' Forever," *Wired*, December 2, 2023, https://www.wired.com/story/chatgpt-poem-forever-security-roundup.

[15] Heidi Mitchell, "Is It Safe to Share Personal Information With a Chatbot?" *Wall Street Journal*, January 18, 2024, https://www.wsj.com/tech/ai/ai-chatbot-sharing-personal-information-229d41a0.

[16] Mack DeGeurin, "ChatGPT Can 'Infer' Personal Details from Anonymous Text," *Gizmodo,* October 17, 2023, gizmodo.com/chatgpt-llm-infers-identifying-traits-in-anonymous-text-1850934318.

regulators globally[17]) that such inferences have dubious scientific validity, and potentially reinforce inaccurate and discriminatory stereotypes. Data privacy laws around the world are already being used to put in place strict limitations on specific kinds of data use that have well-known harms, including such "emotion recognition" systems[18] as well as targeted advertising to children.[19]

(3) ***Business model harms.*** As the past decade illuminates, tech firms already have strong incentives for irresponsible and invasive data collection, fueled primarily by a business model that relies on personalized behavioral targeting of consumers with advertising. The AI boom exacerbates this, fueling a race to the bottom. In fact, a key feature of the current market for large-scale AI is that it is not only computationally, ecologically, and data intensive, it is also very, *very* expensive to develop and run these systems.[20] These eye-watering costs will need a path to profit. By all accounts, though, a viable business model remains elusive.[21] It is precisely in this kind of environment, with a few incumbent firms feeling the pressure to turn a profit, that predatory business models tend to emerge.

Finally, there is also a broader harm that cuts across the indiscriminate collection and retention of data at these various stages of the AI life cycle: the more data that is collected and stored indefinitely, the more we are creating "honeypots" or "goldmines for cyber criminals"[22] that are an attractive target for interception by unauthorized third parties,[23] including malicious state and non-state actors. 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,[24] to the intricate web of third-party data brokers that buy and sell sensitive

[17] Information Commissioner's Office, "'Immature Biometric Technologies Could Be Discriminating against People' Says ICO in Warning to Organisations," October 26, 2022, https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2022/10/immature-biometric-technologies-could-be-discriminating-against-people-says-ico-in-warning-to-organisations.

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

[19] See the American Data Privacy and Protection Act, H.R. 8152, 117th Congress, June 21, 2022, https://docs.house.gov/meetings/IF/IF00/20220720/115041/BILLS-1178152rh.pdf.

[20] Seth Fiegerman and Matt Day, "Why AI Is So Expensive," Bloomberg, April 30, 2024, https://www.bloomberg.com/news/articles/2024-04-30/why-artificial-intelligence-is-so-expensive.

[21] See David Cahn, "AI's $600B Question," Sequoia Capital, June 20, 2024, https://www.sequoiacap.com/article/ais-600b-question; and Benj Edwards, "So Far, AI Hasn't Been Profitable for Big Tech," *Ars Technica*, October 10, 2023, https://arstechnica.com/information-technology/2023/10/so-far-ai-hasnt-been-profitable-for-big-tech.

[22] 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.

[23] For examples of "leaky" data from internet of things (IoT) devices and mobile phones, leaving personal information of users vulnerable to interception, see Anna Maria Mandalari et al., "Blocking without Breaking: Identification and Mitigation of Non-Essential IoT Traffic," Proceedings of Privacy Enhancing Technologies Symposium (PETS), May 11, 2021, https://doi.org/10.48550/arXiv.2105.05162.

[24] Eileen Guo and Hikmat Noori, "This Is the Real Story of the Afghan Biometric Databases Abandoned to the Taliban," *MIT Technology Review*, August 30, 2021,

information about people that can be used to target them unfairly or to hinder their access to credit, housing, and education.[25] Data minimization is based on the premise that 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.

II. The turn toward large-scale AI further consolidates Big Tech's already staggering control over consumer data, which deepens power asymmetries and allows these companies to act recklessly and with impunity.

Large-scale AI depends principally on data and compute resources (this includes both cloud computing and hardware components like chips) 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).[26] 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 billions of people via their vast market penetration.

The idea that "data is everywhere" and therefore not a scarce resource is intuitively appealing but misses the point: quality data *is* scarce. Datasets with high levels of human curation and human feedback; niche datasets especially in high-impact sectors like finance or healthcare; datasets that come with assurances of accuracy, legitimacy, and diversity at scale are becoming a key source of competitive advantage for Big Tech companies, especially in the hypercompetitive generative AI market. This data advantage can give models developed by Big Tech companies an edge over those developed without the benefit of such data. Indeed, access to high-quality data can result in smaller models (those trained on less data and requiring less 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 also from Gmail, Google Drive, and all its other properties.[27] 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.[28]

https://www.technologyreview.com/2021/08/30/1033941/afghanistan-biometric-databases-us-military-40-data-points.

[25] 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.

[26] Jai Vipra and Sarah Myers West, "Computational Power and AI," AI Now Institute, September 27, 2023, https://ainowinstitute.org/publication/policy/compute-and-ai.

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

[28] See 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; and 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.

Repositories of publicly available data currently available online are also likely to dwindle or become less valuable soon in comparison to proprietary datasets held by these companies. We're already seeing a trend toward more restrictions on publicly available data[29] expensive content deals between large AI firms and big publishers like *The Atlantic* and *Axel Springer*[30]and websites like *Reddit* and *Stack Overflow*;[31] and a general lack of transparency around what datasets are being used to train AI.[32]

In this environment, 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.[33] 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.[34]

This market reality must inform any privacy and AI-specific regulatory efforts. Privacy and competition law are too often siloed from each other,[35] leading to interventions that could easily compromise the objectives of one issue over the other.[36] And firms are, in turn, taking advantage of this to amass information asymmetries that contribute to further concentration of their power.[37]

[29] Isabelle Basquette, "AI Startups Have Tons of Cash, but Not Enough Data. That's a Problem," *Wall Street Journal*, June 15, 2023, https://www.wsj.com/articles/ai-startups-have-tons-of-cash-but-not-enough-data-thats-a-problem-d69de120.

[30] Damon Beres, "A Devil's Bargain with OpenAI," *Atlantic*, May 29 2024, https://www.theatlantic.com/technology/archive/2024/05/a-devils-bargain-with-openai/678537.

[31] Associated Press, "As AI Learns from Stack Overflow, Reddit, and More Platforms, Companies Are Adapting While Users Protest," *Fast Company*, July 3, 2024, https://www.fastcompany.com/91150665/ai-learns-stack-overflow-reddit-facebook-adapting-users-protest.

[32] See 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; @XDevelopers, March 29, 2023, https://x.com/XDevelopers/status/1641222782594990080; and 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.

[33] 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.

[34] 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.

[35] Udbhav Tiwari, "Competition Should Not Be Weaponized to Hobble Privacy Protections on the Open Web," Mozilla (blog), April 12, 2022, https://blog.mozilla.org/netpolicy/2022/04/12/competition-should-not-be-weaponized-to-hobble-privacy-protections-on-the-open-web.

[36] 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.

[37] For example, Article 5 of the European Union's Digital Markets Act prohibits large "gatekeeper" platforms from the cross-use of personal data between its various service offerings, without explicit user consent. See European Commission, "The Digital Markets Act: Ensuring Fair and Open Digital Markets," accessed July 9, 2024, https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en.

This concentration of power enabled by control over data isn't just a problem for potential competitors of Big Tech. Too much centralized economic power in the hands of too few harms our democracy—especially when these very same actors have proven themselves to be reckless and far from dependable custodians of this power. Amid the hype surrounding AI, companies are rushing to market with technologies that are far from ready to be broadly accessible. Google recently rolled out its AI Overviews feature in its search engine results; within days it was producing inaccurate, nonsensical, and even dangerous answers to people's queries.[38] Meta's new AI agents, which the company integrated into millions of Instagram and Facebook accounts, have generated misinformation and misled people into believing they were interacting with real human beings.[39] And Microsoft has been broadly panned for proposing a new AI-enabled feature named Recall that raises numerous privacy-related red flags.

These are the very same companies that resist regulatory guardrails in the name of "innovation." It's time to question the premise: Is this scramble for reckless growth by a handful of surveillance monopolies really innovation? It's not surprising that these companies, free from regulatory constraints or competitive market pressures, are acting out. A data privacy mandate that embeds transparency and accountability around how these companies build AI, and when they determine they are fit for market release, isn't curbing innovation: it's a long overdue check on these companies.

III. Data privacy law—in particular strong data minimization, impact assessments, data rights, and protections against algorithmic discrimination—provides a foundational toolkit for demanding accountability from AI companies.

Taking stock of some of the myriad, diffused ways in which AI is poised to heighten threats to our individual and collective privacy, and worsen the concentrations of power in Big Tech, we can now ask: *How might the AI market develop differently in the presence of a strong, broad-ranging federal privacy law?*

a. Data minimization:

Data minimization rules impose 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 recent proposals such as the APRA,[40] are a core part of global data protection laws. As AI Now, Accountable Tech, and EPIC emphasize in our "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

[38] Ellie Stevens, "The 7 Most Shocking Google AI Answers We've Seen So Far," *Fast Company*, May 30, 2024, https://www.fastcompany.com/91132974/shocking-google-ai-overview-answers.

[39] "Meta's New AI Agents Confuse Facebook Users," Associated Press, April 20, 2024, https://www.voanews.com/a/meta-s-new-ai-agents-confuse-facebook-users-/7576420.html.

[40] US Senate Committee on Commerce, Science, and Transportation, "Committee Chairs Cantwell, McMorris Rodgers Unveil Historic Draft Comprehensive Data Privacy Legislation," April 7, 2024, https://www.commerce.senate.gov/2024/4/committee-chairs-cantwell-mcmorris-rodgers-unveil-historic-draft-comprehensive-data-privacy-legislation.

communities on one hand, and the large corporations that create and collect data about them on the other.[41]

Collection limitation, for example, would force firms to adhere to limits when it comes to data surveillance and acquisition—to build within the constraints of necessity and proportionality. This would replace reckless organizational data cultures with reflexivity that forces engineers to calibrate decisions about data, keeping in mind the privacy and security vulnerabilities these create. In response to Microsoft's announcement of its Recall feature that continuously screenshots our activities on the computer, the lawmakers and the public would be empowered to demand basic accountability: is such data surveillance, that creates a honey pot for bad actors and unauthorized access, at all proportionate? It is likely, in fact, that this feature may have never been announced in the first place had the company done even a rudimentary impact assessment that evaluated risks to privacy and security. Put simply, a strong data minimization mandate would have disincentivized the development of these patently unsafe features to begin with.

A purpose limitation rule, on the other hand, could restrain Big Tech monopolies from endlessly combining data collected for distinct purposes in pursuit of consolidating their data advantage against competitors. We already have examples of these rules being applied to protect consumers from harm. The FTC has also penalized Amazon for storing children's voiceprints—highly sensitive data—and shot down the company's justification that it would be used to improve its Alexa algorithm.[42]

Most 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 that abound between individuals and massive tech firms.[43] Just this week, Open AI CEO Sam Altman, in collaboration with another company, announced Thrive AI Health, pitched as a "hyper-personalized AI health coach."[44] They set out a pervasive vision of data surveillance, ranging from highly intimate sleeping, eating, and exercise behaviors combined with medical data. The premise is that because individuals can "choose" whether to share this data, any privacy concerns are put to rest. This flies in the face of a decade of evidence that indicates we cannot rely solely on people's choices to protect them, when the long-term implications of unauthorized access, out-of-context sharing, or malicious use are difficult if not impossible for the average consumer to meaningfully comprehend. This is the outcome of a regulatory environment that has failed to place limits on the unrestrained collection, storage, and use of sensitive data, allowing companies to fall back on consent as a catchall defense.

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

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

[43] See 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; and David Medine and Gayatri Murthy, "Companies, Not People, Should Bear the Burden of Protecting Data," Brookings, December 18, 2019, https://www.brookings.edu/articles/companies-not-people-should-bear-the-burden-of-protecting-data.

[44] Sam Altman and Ariana Huffington, "AI-Driven Behavior Change Could Transform Health Care," *Time*, July 7, 2024, https://time.com/6994739/ai-behavior-change-health-care.

Beyond the general principle of data minimization, a data privacy law could include prohibitions on specific kinds of data use that have well-known harms, such as targeted advertising to children,[45] or uses based on sensitive data categories,[46] or the use of data about people's interior mental states in so-called "emotion recognition" systems that have been repeatedly found to be based on faulty foundations.[47] Perhaps, Open AI CEO Sam Altman would have likely been stopped in his tracks before claiming the recent multimedia chatbot Sora would be able to "detect people's emotional states" from people's voice recordings.

b. Data rights:

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 legislation have not been fulfilled.

Currently, the only constraint on usage of any consumer data for training of proprietary models comes from the terms of service of those products, which can be changed at will, as Google and Meta recently did. Notable too that while European users were alerted by Meta that it would use publicly available posts to train its AI, American users received no such notification.[48] With a comprehensive data privacy law, these individuals would have, at minimum, the ability to demand transparency around the use of their data.

Under the latest text of APRA, consumers would also have a broad right to opt out of algorithmic decision-making that comprises "consequential decisions"—defined as decisions, including ads, that may impact an individual's equal access to housing, employment, healthcare, and so on.[49] Such algorithmic decision-making is ubiquitous today, with limited oversight. As just one example, an IBM survey in 2023 showed that out of 8,500 participants in the survey, 42 percent were already using AI screening to filter out candidates, and another 40 percent were in the process of integrating with such technology.[50] These screening softwares have been known to filter out qualified candidates based on their age, gender, or even hobbies, with marginalized candidates bearing the brunt of maximum harm.[51] A right to opt out of consequential decisions would allow these candidates a fairer review of their applications and would force companies to

[45] See American Data Privacy and Protection Act, H.R. 8152, 117th Congress, https://docs.house.gov/meetings/IF/IF00/20220720/115041/BILLS-1178152rh.pdf.

[46] Ibid.

[47] Access Now, European Digital Rights (EDRi), Bits of Freedom, Article 19, and IT-Pol, "Prohibit Emotion Recognition in the Artificial Intelligence Act."

[48] Eli Tan, "When the Terms of Service Change to Make Way for A.I. Training."

[49] US Senate Committee on Commerce, Science, and Transportation, "Committee Chairs Cantwell, McMorris Rodgers Unveil Historic Draft Comprehensive Data Privacy Legislation," April 7, 2024, https://www.commerce.senate.gov/2024/4/committee-chairs-cantwell-mcmorris-rodgers-unveil-historic-draft-comprehensive-data-privacy-legislation.

[50] "Data Suggests Growth in Enterprise Adoption of AI is Due to 'Widespread Deployment by Early Adopters, But Barriers Keep 40% in the Exploration and Experimentation Phase,'" IBM Newsroom, January 10, 2024, https://newsroom.ibm.com/2024-01-10-Data-Suggests-Growth-in-Enterprise-Adoption-of-AI-is-Due-to-Widespread-Deployment-by-Early-Adopters.

[51] 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.

put better pipelines to employment in place that do not exacerbate entrenched inequalities in the workplace.

c. Impact Assessments:

A data privacy law should also include 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 requires companies to evaluate whether particular groups will be harmed as a result of the design or use of the AI system. Researchers like Dr. Alex Hanna and Dr. Mehtab Khan, for example, have put forward a multilayered framework to scrutinize the multiple complex layers of large-scale AI models.[52]

One could imagine that some of the most concerning recent AI features and products, from Microsoft's Recall to Google AI Overviews, would perhaps never have been announced or brought to market had firms been required to comprehensively evaluate the privacy and security implications of their systems before release.

A note of caution on impact assessments: while such evaluations are positive in theory, these obligations must be drafted to ensure meaningful accountability. There is a significant risk that any audit or evaluation standard can devolve into a superficial checkbox exercise,[53] more useful in offloading liability than in protecting the public. With that in mind, we recommend the following :

- Meaningful assessments that mandate evaluation should happen *before* products are made available for use in the public domain, and should be subject to evaluation on an ongoing basis while in operation. It is essential that the criteria for such evaluations not be limited to narrow technical parameters or be tested only under so-called "laboratory-like conditions."[54]

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

[53] See Amba Kak and Sarah Myers West, *Algorithmic Accountability: Moving Beyond Audits*, AI Now Institute, April 11, 2023, https://ainowinstitute.org/publication/algorithmic-accountability; and 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, 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.

[54] See 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

- 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. Prohibition against discrimination:

A range of privacy proposals, both in the United States and globally, include protections against using personal data in AI in ways that discriminate.[55] It is now well documented that AI systems are routinely, and often structurally, biased in ways that entrench and embed historical inequities[56] in sensitive social domains like healthcare,[57] hiring,[58] education,[59] housing,[60] and

Algorithmic Fairness," *Phenomenal World,* January 30, 2020, https://www.phenomenalworld.org/analysis/long-history-algorithmic-fairness.

[55] US Senate Committee on Commerce, Science, and Transportation, "Committee Chairs Cantwell, McMorris Rodgers Unveil Historic Draft Comprehensive Data Privacy Legislation."

[56] See 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; 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.

[57] Ziad Obermeyer et al., "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations," *Science* 366 (October 25, 2019): 447-453, https://www.science.org/doi/10.1126/science.aax2342.

[58] See 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, 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; and Aaron Rieke and Miranda Bogen, "Help Wanted."

[59] See 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; and 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.

[60] See 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-renter; Ridhi Shetty, "CDT Comments to Federal Agencies Highlight Risks of Data Used in Tenant Screening," Center for

criminal justice.[61] This should not come as a 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 fact sheet 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 realms.[62] A civil rights provision in a federal privacy law would provide an overarching means of redress against AI systems that perpetuate discrimination.

To conclude, the key lesson of the past decade has been understanding that control over data is about power asymmetries, and since companies derive clear commercial benefit from widening this asymmetry, regulation is essential to protect the public from harm. Passing strong federal privacy legislation is a critical and overdue step in that direction.[63] And while it is true that the United States is already behind in terms of enacting a comprehensive data privacy law, in those countries that have these legislative mandates in place, there have been major gaps and ambiguities in implementation. An opportunity exists, therefore, to enact and creatively apply foundational privacy principles to the emergent landscape of AI systems, setting the gold standard of enforcement for the rest of the world.

Democracy & Technology, June 2, 2023, https://cdt.org/insights/cdt-comments-to-federal-agencies-highlight-risks-of-data-used-in-tenant-screening.

[61] 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.

[62] White House, "Blueprint for an AI Bill of Rights."

[63] Accountable Tech, AI Now Institute, and EPIC, "Zero Trust AI Governance."