October 16, 2023

TO: Members, Subcommittee on Innovation, Data, and Commerce
FROM: Committee Majority Staff
RE: Hearing Entitled "Safeguarding Data and Innovation: Building the Foundation for the Use of Artificial Intelligence"

I. INTRODUCTION

The Subcommittee on Innovation, Data, and Commerce will hold a hearing on Wednesday, October 18, 2023, at 10:00 a.m. in 2123 Rayburn House Office Building. The hearing is entitled "Safeguarding Data and Innovation: Building the Foundation for the Use of Artificial Intelligence."

II. WITNESSES

- Victoria Espinel – President and Chief Executive Officer, BSA | The Software Alliance
- Raffi Krikorian – Chief Technology Officer, Emerson Collective
- Amba Kak – Executive Director, AI Now Institute
- Clark Gregg – Actor and Screenwriter, SAG-AFTRA
- Jon Leibowitz – Former Chair and Commissioner, Federal Trade Commission

III. BACKGROUND

Artificial intelligence (AI) is a class of technologies that perform tasks commonly believed to require human intelligence, like decision making and inferences. Though collectively known as AI, these technologies encompass a wide range of methodologies and modalities, like machine learning and image processing.[1] With such a broad definition, our understanding of AI has and will continue to shift over time as consumers adjust their expectations to new and more advanced AI. For example, in 1995, DeepBlue, a chess algorithm which defeated Grandmaster Garry Kasparov, was considered cutting edge AI. Today's most advanced AI pilot self-driving vehicles, inspect the content we see on social media, and combat cybersecurity threats.

[1] Laurie A. Harris, *Overview of Artificial Intelligence*, Congressional Research Service (October 24, 2017), https://www.crs.gov/Reports/IF10608?source=search.

At its core, AI relies on three components: data, computing power, and learning algorithms.[2] Vast amounts of data train algorithms in large data centers. These trained algorithms are then applied to make or aid in decision making. This data, in a number of cases, can be comprised of personal and sensitive information about individuals, that may or may not have been collected and used without their consent. As Congress examines AI and questions if regulation is needed, the foundational step is to enact a comprehensive federal data privacy and security law that sets protections for American's sensitive information.

IV. NEED FOR COMPREHENSIVE DATA PRIVACY AND SECURITY

Currently, there is no comprehensive U.S. federal data privacy and security law. To provide this protection to citizens, twelve U.S. states have enacted their own state privacy laws,[3] which has created a patchwork of rules and regulations that can cause confusion for consumers, gaps in privacy protections, compliance burdens and uncertainty for businesses, while at the same time hampering innovators and creators. States with comprehensive data privacy and security laws, like California, have regulators and legislators that have already signaled they will regulate AI.[4] Similarly, following the implementation of the General Data Protection Regulation (GDPR), the European Union has proposed the AI Act, the world's first comprehensive AI law, which will regulate artificial intelligence.[5] To stay ahead of legislators and regulators in Sacramento and Brussels from setting de facto rules for how business across the U.S. can collect and process information, and deploy AI, Congress must enact a comprehensive federal data privacy and security law.

V. AI FOUNDATIONS

According to research published by Microsoft in 2001, the performance of algorithms benefits significantly from larger training sets.[6] In the decades since, the amount of quality data available for training algorithms has increased dramatically. As data available for training has grown, the complexity and performance of AI has also grown. An AI's parameters, defined as the weights and biases that a model learns during training,[7] can be used to measure the complexity and performance. For example, in 2001 a top vision model contained 180 thousand

[2] Ben Buchanan, *Prepared Testimony House Oversight Committee* (April 18, 2018), https://www.belfercenter.org/publication/prepared-testimony-house-oversight-committee.
[3] Andrew Folks, *US State Privacy Legislation Tracker* (September 15, 2023), International Association of Privacy Professionals. https://iapp.org/resources/article/us-state-privacy-legislation-tracker/.
[4] Billy Perrigo, *California Bill Proposes Regulating AI at State Level*, Time Magazine (September 13, 2023), https://time.com/6313588/california-ai-regulation-bill/.
[5] European Parliament, *EU AI Act: first regulation on artificial intelligence* (June 14, 2023), https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.
[6] Michele Banko and Eric Brill, *Scaling to Very Very Large Corpora for Natural Language Disambiguation*, https://aclanthology.org/P01-1005.pdf.
[7] Google, *Machine Learning Glossary* (September 26, 2023), https://developers.google.com/machine-learning/glossary#p.

parameters.[8] When ChatGPT was publicly released in November 2022, it had more than 175 billion parameters.[9]

Data used to train AI models is collected in a variety of ways. Some companies use a combination of internal and external data.[10] Other firms, such as those deploying large scale models, mostly rely on data harvested from the internet. Data harvesting is a process by which data, often publicly available information, is copied from the internet. There have been many cases where personal information or works registered as intellectual property have been copied into training data sets.[11] Some companies package this data and sell it to firms while others create open-source data sets for developers, like LAION (Large-scale AI Open Network).[12] This large-scale data collection has spurred conversations regarding data ownership, with some tech companies establishing systems to minimize and deter scraping, arguing that scraping diminishes user experience.[13] In contrast, others argue that such freely available information on the internet should be "fair game."[14] Alternatively, some rely on synthetic data, which is a secondary set of data generated by taking a relational database and creating a generative machine learning model for it.[15] The result is a data set that has the same mathematical properties as the original data set but does not compromise privacy.[16] Still, the use of synthetic data in specific inferential circumstances has been criticized for sacrificing accuracy to improve privacy.[17]

Along with data, hardware and algorithms are the pillars of AI,[18] with algorithms that define the logic of the models, hardware that runs the algorithms, and data that is used by algorithms to train the models.[19] Increased amounts of training data due to the lack of a U.S.

---

[8] Paul Viola and Michael Jones, *Rapid Object Detection using a Boosted Cascade of Simple Features,* Accepted Conference on Computer Vision and Pattern Recognition (December 8, 2001) https://www.cs.cmu.edu/~efros/courses/LBMV07/Papers/viola-cvpr-01.pdf.
[9] Eric Griffith, *GPT-4 vs. ChatGPT-3.5: What's the Difference*, PC Magazine (March 16, 2023), https://www.pcmag.com/news/the-new-chatgpt-what-you-get-with-gpt-4-vs-gpt-35.
[10] Sara Brown, *Why external data should be part of your data strategy*, MIT Sloan School of Management (February 18, 2021), https://mitsloan mit.edu/ideas-made-to-matter/why-external-data-should-be-part-your-data-strategy.
[11] Kevin Schaul, Szu Yu Chen and Nitasha Tiku, *Inside the secret list of websites that make AI like ChatGPT sound smart*, Washington Post (April 19, 2023), https://wapo.st/3S071CL.
[12] Marissa Newman and Aggi Cantrill, *The Future of AI Relies on a High School Teacher's Free Database*, Bloomberg (April 23, 2023), https://www.bloomberg.com/news/features/2023-04-24/a-high-school-teacher-s-free-image-database-powers-ai-unicorns#xj4y7vzkg.
[13] Reuters*, Twitter now requires users to sign in to view tweets* (June 30, 2023), https://www.reuters.com/technology/twitter-now-needs-users-sign-view-tweets-2023-06-30/.
[14] *Supra* note 12.
[15] Brian Eastwood, *What is synthetic data — and how can it help you competitively?*, MIT Sloan School of Management (January, 23, 2023). https://mitsloan mit.edu/ideas-made-to-matter/what-synthetic-data-and-how-can-it-help-you-competitively.
[16] *Id*.
[17] Sabine Neschke, *Synthetic Data*, Bipartisan Policy Center (October 01, 2021), https://bipartisanpolicy.org/blog/synthetic-data/.
[18] Olga Megorskaya, *Training Data: The Overlooked Problem Of Modern AI*, Forbes (June 27, 2022), forbes.com/sites/forbestechcouncil/2022/06/27/training-data-the-overlooked-problem-of-modern-ai/?sh=1828c17a218b.
[19] Toloka, *Evolution of The Data Production Paradigm in AI* (September 17, 2021), https://hackernoon.com/evolution-of-the-data-production-paradigm-in-ai.

federal data privacy standard, combined with increased computing power, has unlocked many of the new AI use-cases we see today.[20]

VI. USE CASES

"Generative AI" describes a class of algorithms that can generate high-quality text, images, and other content based on training data.[21] Examples include chatbots like OpenAI's ChatGPT, image generators like Stable Diffusion, and code generators like GitHub's copilot. Applications of chains are instrumental in increasing productivity, reducing constraints on the economy, as we see with supply chains, and offer new opportunities for jobs. One such application we have seen grow from AI are self-driving vehicles, commonly referred to as autonomous vehicles (AV), which promise to improve roadway safety drastically,[22] fill voids that we see within labor shortages,[23] and contribute billions of dollars to the American economy.[24]

These benefits, however, will only occur in America if we establish a path forward for America to lead in the development and deployment of the technology. That is why in 2020, the Committee developed legislation to establish strategies to lead on emerging technologies, specifically AI, in the American COMPETE Act.[25] By following the principles outlined in that statute, the U.S. can implement strategies necessary to lead on AI and related technologies, rather than China determining the future of its usage and adoption. We have already seen bad things happen when data protections are not in place. For instance, "deep fake" describes a false but realistic photo, audio, video, or other forgery created by AI.[26] While there are many beneficial use cases for the AI techniques underlying deep fakes, like instances where medical researchers have synthesized fake medical images to train disease detection algorithms for rare diseases and to minimize patient privacy concerns,[27] the deceptive uses have drawn much attention. For example, some users have inserted actors into productions that they did not appear in, or agree to have their likeness included.[28] The results are increasingly realistic, and these models often do not disclose if the product is created with artificial intelligence, causing concern.

---

[20] Karl Rupp, *Microprocessor Trend Data (2022)*, Our World In Data, https://ourworldindata.org/grapher/transistors-per-microprocessor.
[21] Kim Martineau, *What is generative AI?,* IBM Research (April 20, 2023), https://research.ibm.com/blog/what-is-generative-AI.
[22] John Bailey, *Autonomous Vehicles: A Safer Road Ahead*, American Enterprise Institute (September 20, 2023), https://www.aei.org/technology-and-innovation/autonomous-vehicles-a-safer-road-ahead/.
[23] Melanie Stone, *How Industrial Automation Helps Ease Labor Challenges., Cyngn (June 4, 2023), https://www.cyngn.com/blog/how-industrial-automation-helps-ease-labor-challenges.*
[24] Skip Descant*, Autonomous Vehicles to Have Huge Impact on Economy, Tech Sector* (June 26, 2018), https://www.govtech.com/fs/automation/autonomous-vehicles-to-have-huge-impact-on-economy-tech-sector.html.
[25] H.R. 8132, American Competitiveness of a More Productive Emerging Tech Economy Act (116th Congress), https://www.congress.gov/bill/116th-congress/house-bill/8132.
[26] Kelley M. Sayler and Laurie A. Harris, *Deep Fakes and National Security*, Congressional Research Service (April 17, 2023), https://www.crs.gov/Reports/IF11333?source=search.
[27] Ibid.
[28] Dawn Chmielewski, *Actors decry 'existential crisis' over AI-generated 'synthetic' actors*, Reuters (July 21, 2023),https://www.reuters.com/technology/actors-decry-existential-crisis-over-ai-generated-synthetic-actors-2023-07-21/.

Like deep fake technology, the wide range of chatbot use cases are raising concern. The way one interacts in a conversational nature with the technology can lead people to share information they would not otherwise volunteer to a traditional search engine. Further, some companies list in their policies that chatbot input data can be used for targeted advertising.[29] Additionally, some have turned to these systems for medical advice, highlighting issues with how these systems relate to existing data protection requirements, along with the uncertainty about the accuracy of the information they provide, and the bias ingrained in the algorithms they are drawn from.[30]

VII. FEDERAL TRADE COMMISSION AUTHORITY

While the field of AI is still developing, the Federal Trade Commission (FTC) has identified where they see they have authority in this space:[31]

- **Section 5 of the FTC Act.** The FTC Act prohibits unfair or deceptive practices. That would include the sale or use of – for example – racially biased algorithms.
- **Fair Credit Reporting Act.** The FCRA comes into play in certain circumstances where an algorithm is used to deny people employment, housing, credit, insurance, or other benefits.
- **Equal Credit Opportunity Act.** The ECOA makes it illegal for a company to use a biased algorithm that results in credit discrimination on the basis of race, color, religion, national origin, sex, marital status, age, or because a person receives public assistance.

VIII. PREVIOUS LEGISLATION

In 2021, Chair Rodgers' bipartisan legislation with former-Rep. Bobby Rush, known as the American COMPETE Act,[32] was signed into law as part of the "Consolidated Appropriations Act, 2021."[33] The COMPETE Act required the Department of Commerce to outline strategies and policies to promote emerging technologies such as artificial intelligence.

Last Congress, the Committee passed the American Data Privacy and Protection Act (ADPPA),[34] which included the ground rules around collection, transfer, use and storage of

---

[29] Kate O'Flaherty, *Cybercrime: be careful what you tell your chatbot helper…*, The Guardian (April 9, 2023), https://www.theguardian.com/technology/2023/apr/09/cybercrime-chatbot-privacy-security-helper-chatgpt-google-bard-microsoft-bing-chat.

[30] Sara Reardon, *AI Chatbots Can Diagnose Medical Conditions at Home. How Good Are They?*, Scientific American (March 31, 2023), https://www.scientificamerican.com/article/ai-chatbots-can-diagnose-medical-conditions-at-home-how-good-are-they/.

[31] Elisa Jillson, *Aiming for truth, fairness, and equity in your company's use of AI*, Federal Trade Commission (April 19, 2021), https://www ftc.gov/business-guidance/blog/2021/04/aiming-truth-fairness-equity-your-companys-use-ai.

[32] H.R. 8132, American Competitiveness of a More Productive Emerging Tech Economy Act (116th Congress), https://www.congress.gov/bill/116th-congress/house-bill/8132.

[33] Consolidated Appropriations Act, 2021, Public Law 116-260 (Dec. 27, 2020), https://www.congress.gov/116/plaws/publ260/PLAW-116publ260.pdf.

[34] H.R. 8152, American Data Privacy and Protection Act (117th Congress), https://www.congress.gov/bill/117th-congress/house-bill/8152.

personal data, as well as including requirements for certain companies to conduct algorithmic impact assessments.

IX. ISSUES

- What privacy concerns arise from the widespread adoption of large language models and other forms of artificial intelligence?
- How can increasing user control over their data help alleviate concerns about the future of AI?

X. STAFF CONTACTS

- Tim Kurth, Chief Counsel
- Teddy Tanzer, Senior Counsel
- Brannon Rains, Professional Staff Member
- Michael Cameron, Professional Staff Member
- Jessica Herron, Clerk