The Senate Homeland Security and Government Affairs Committee Hearing on AI: Risks and Opportunities

Mar 8, 2023

Remarks delivered by Suresh Venkatasubramanian.

Chairman Peters, Ranking Member Paul, and members of the Homeland Security and Government Affairs Committee. I thank you for inviting me to testify at this important hearing on the risks and opportunities of AI. I'm a professor of computer science and director of the Center for Technological Responsibility at Brown University. I recently completed a stint as a White House tech policy advisor in the Biden Administration and included in my portfolio was developing the recently released Blueprint for an AI Bill of Rights.[1] I have spent the last decade studying and researching the impact of automated systems (and AI) on people's rights, opportunities, and ability to access services. I've also spent time working with civil society groups and advising state and local governments on sound approaches to governing the use of technology that impacts people's lives.

What is AI?

We are here today to talk about AI – artificial intelligence. As an academic discipline, AI seeks to design automated systems that can sense, interact, reason, and behave in the way that humans do, and in some cases even surpass us.

We learn from the data we receive. And thus, one sub-area of AI that is dominant right now, fueled by the collection of vast amounts of data, is *machine learning*[2] – the design of systems that can incorporate historical data into the predictions that they produce, and in some cases keep adapting as more data appears. Machine learning grew in part out of decades of work in statistics: this is important to bear in mind since many systems that say they are using AI are really using statistical techniques that were invented decades ago and that are now supercharged by data.

Virtually every sector of society is now touched by machine learning. Algorithms created via machine learning are used to incarcerate individuals before trial[3], decide what sentence they should get if convicted[4], and decide whether they should get parole, and under what conditions.[5] Algorithms created via machine learning are used to determine a detected

---

[1] https://www.whitehouse.gov/ostp/ai-bill-of-rights/
[2] Hal Daumé III. A course in machine learning. http://ciml.info/
[3] David G. Robinson and Logan Koepke. Civil Rights and Pretrial Risk Assessments. Upturn, Inc., Dec. 2019. https://www.upturn.org/static/files/Robinson-Koepke-Civil-Rights-Critical-Issue-Brief.pdf
[4] John Villasenor and Virginia Foggo. Algorithms and sentencing: what does due process require? Brookings Institute, Mar. 2019. https://www.brookings.edu/blog/techtank/2019/03/21/algorithms-and-sentencing-what-does-due-process-require/
[5] Casey et al., Using offender risk and needs assessment information at sentencing. Nat'l Center for State Courts, 2011. https://www.ncsc.org/__data/assets/pdf_file/0019/25174/rna-guide-final.pdf

sound could be a gunshot,[6] or whether a blurred partial picture of an individual matches a known suspect.[7] Algorithms are used to monitor children in school for risk of suicide;[8] they are used to predict learning outcomes, and likelihood of success in educational settings.[9]

Algorithms created via machine learning screen candidate resumes for employers, analyze the results of video interviews or online interactive tests, and provide "fit" scores when employers are making hiring decisions.[10]

Machine learning algorithms are used to determine whether applicants for benefits are legitimate or fraudulent, what kinds of benefits they are eligible for, and how much they should receive.[11] These same algorithms are used to assess whether children are at risk for neglect or abuse, and whether social workers should intervene in a family.[12] These algorithms decide whether individuals should get health care, and what kind of care.[13] They interpret the results of medical tests. They decide whether individuals should get insurance coverage, and what price they should pay for this coverage.[14] Algorithms decide whether a potential renter should be considered by a landlord,[15] and what price this tenant should pay.[16] They are used to estimate the market value for a house, and what mortgage rate an individual can be asked to pay.[17] Algorithms are used to decide whether someone is a good credit risk for a loan.[18]

---

[6] Ferguson et al. The Chicago police department's use of shotspotter technology. Office of the Inspector General, Chicago, Aug. 2021. https://igchicago.org/wp-content/uploads/2021/08/Chicago-Police-Departments-Use-of-ShotSpotter-Technology.pdf
[7] https://www.clearview.ai/law-enforcement
[8] https://www.gaggle.net/
[9] https://www.civitaslearning.com/
[10] Bogen and Rieke. Help Wanted: An examination of hiring algorithms, equity, and bias. Upturn, Dec 2018. https://apo.org.au/sites/default/files/resource-files/2018-12/apo-nid210071.pdf
[11] Angwin. The Seven-Year Struggle to Hold an Out-of-Control Algorithm to Account. The Markup, Oct. 2022. https://themarkup.org/newsletter/hello-world/the-seven-year-struggle-to-hold-an-out-of-control-algorithm-to-account
[12] Samant et al. Family Surveillance by Algorithm: The Rapidly Spreading Tools Few Have Heard Of. ACLU, Sep. 2021. https://www.aclu.org/news/womens-rights/family-surveillance-by-algorithm-the-rapidly-spreading-tools-few-have-heard-of
[13] Ziad Obermeyer, et al., Dissecting racial bias in an algorithm used to manage the health of populations, 366 Science (2019), https://www.science.org/doi/10.1126/science.aax2342.
[14] I. E. Kumar. Colorado DOI weighs in on how to prevent algorithmic discrimination in life insurance. Center for Tech Responsibility, Brown University, Mar 2023. https://cntr.substack.com/p/colorado-doi-weighs-in-on-how-to
[15] K. Waddell. How Tenant Screening Reports Make It Hard for People to Bounce Back From Tough Times. Consumer Reports, Mar 2021. https://www.consumerreports.org/algorithmic-bias/tenant-screening-reports-make-it-hard-to-bounce-back-from-tough-times-a2331058426/
[16] Vogell. Rent Going Up? One Company's Algorithm Could Be Why. ProPublica, Oct. 2022. https://www.propublica.org/article/yieldstar-rent-increase-realpage-rent
[17] Consumer Financial Protection Bureau. Consumer Financial Protection Bureau Outlines Options To Prevent Algorithmic Bias In Home Valuations. Feb. 2022. https://www.consumerfinance.gov/about-us/newsroom/cfpb-outlines-options-to-prevent-algorithmic-bias-in-home-valuations/
[18] I.E Kumar et al. Equalizing Credit Opportunity in Algorithms: Aligning Algorithmic Fairness Research with U.S. Fair Lending Regulation. Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES '22). https://doi.org/10.1145/3514094.3534154

The list goes on and on.

There are many ways to build such "learning" systems. One specific kind of system that has risen to great prominence, in part driven by the availability of cheap and powerful computing power, is deep learning.[19] Deep learning has proven to be most powerful when analyzing images, text, audio, or video. Deep learning algorithms are used in facial recognition systems, in systems that analyze brain scans for neurological disorders, in the cameras installed on cars with driver assist or some other form of autonomous driving, in systems that translate from one language to another, and in systems that convert speech to text and vice versa. This list has grown rapidly and will continue to grow as we develop the underlying technology.

A *transformer*[20] is a particular kind of deep learning system, and as the name suggests, learns how to transform inputs and generate new kinds of output. Transformers are most useful for generating new kinds of content, whether it be deepfakes, plausibly realistic video segments, and of course text dialogue systems like GPT3[21], ChatGPT[22], Bard[23], and many others. Transformers need to ingest extremely large amounts of data, and require huge compute power, to do what they do.

The Failures of AI

Whether the system being used is a standard machine learning system, or one using more specialized architectures like deep learning, or even a transformer, all these systems share some common features that are important for how we might govern them. These are not algorithms or computer programs like the software of the 80s and 90s, or even the 00s. They are "algorithms for making algorithms":[24] the distinctive feature of a machine learning system is that the output of the learning algorithm that is fed vast amounts of data *is itself an algorithm* that purports to solve the underlying problem, whether a prediction task, an image analysis, or a text-based interaction with a user.

As a consequence of the above, *we don't actually know for sure whether and how these algorithms work and why they produce the output that they do.* This might come as a surprise, given how much we hear every day about the amazing and miraculous successes of AI. And yet, we also hear every day about the failures of AI systems.

---

[19] The "deep" refers to a specific aspect of the design of these systems and is not a statement about the quality of the results produced.
[20] A. Vaswani et al. Attention is all you need. Advances in neural information processing systems 30 (2017). https://papers.nips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html
[21] https://openai.com/blog/gpt-3-apps
[22] https://chat.openai.com/
[23] https://blog.google/technology/ai/bard-google-ai-search-updates/
[24] Venkatasubramanian. When an algorithm isn't. Medium, Oct 2015. https://medium.com/@geomblog/when-an-algorithm-isn-t-2b9fe01b9bb5

Princeton professor Arvind Narayanan has likened AI systems to Snake oil: "Much of what's being sold today as AI is snake oil: it does not, and cannot work".[25] Researchers Deb Raji, Lizzie Kumar, Aaron Horowitz, and Andrew Selbst have referred to this same problem as "The Fallacy of AI Functionality", asserting that "Deployed AI systems often do not work" and laying out a series of case studies illustrating the myriad, and different, ways in which AI systems fail.[26]

AI systems fail when the algorithms draw incorrect conclusions from data and penalize individuals subject to those conclusions. A company installed AI-powered cameras in its delivery vans to evaluate the road safety habits of its drivers, but the system incorrectly penalized drivers when other cars cut them off or when other events beyond their control took place on the road. As a result, drivers were incorrectly ineligible to receive a bonus.[27]

AI systems fail when they seek to make predictions based on faulty data: a system that tried to predict effectiveness of health interventions used historical data on the cost of health care that was racially biased and produced racially biased outcomes.[13] Another system ended up causing the IRS to audit Black taxpayers more often than other taxpayers, for no apparent reason.[28]

AI systems fail when they are built using data from one group of people, and then are applied to a different group of individuals. The National Disabled Law Students Association expressed concerns that individuals with disabilities were more likely to be flagged as potentially suspicious by remote proctoring AI systems because of their disability-specific access needs such as needing longer breaks or using screen readers or dictation software .[29]

AI systems fail when the results of one automated decision system are fed into another (or even the same one), causing any errors in the original system to be amplified. An algorithm used to deploy police was found to repeatedly send police to neighborhoods they regularly visit, even if those neighborhoods were not the ones with the highest crime rates. These

---

[25] A. Narayanan. How to recognize AI snake oil. Nov. 2019. https://www.cs.princeton.edu/~arvindn/talks/MIT-STS-AI-snakeoil.pdf

[26] I. D. Raji et al. The Fallacy of AI Functionality. In 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22). https://dl.acm.org/doi/fullHtml/10.1145/3531146.3533158

[27] Lauren Kaori Gurley. Amazon's AI Cameras Are Punishing Drivers for Mistakes They Didn't Make. Motherboard. Sep. 20, 2021. https://www.vice.com/en/article/88npjv/amazons-ai-cameras-are-punishing-drivers-for-mistakes-they-didnt-make

[28] Jim Tankersley. Black Americans Are Much More Likely to Face Tax Audits, Study Finds. New York Times, Jan. 31, 2023. https://www.nytimes.com/2023/01/31/us/politics/black-americans-irs-tax-audits.html

[29] See, e.g., National Disabled Law Students Association. Report on Concerns Regarding Online Administration of Bar Exams. Jul. 29, 2020. https://ndlsa.org/wp-content/uploads/2020/08/NDLSA_Online-Exam-Concerns-Report1.pdf; Lydia X. Z. Brown. How Automated Test Proctoring Software Discriminates Against Disabled Students. Center for Democracy and Technology. Nov. 16, 2020. https://cdt.org/insights/how-automated-test-proctoring-software-discriminates-against-disabled-students/

incorrect crime predictions were the result of a feedback loop generated from the reuse of data from previous arrests and algorithm predictions.[30]

AI systems fail when they are so opaque that errors in how they function cannot be detected. In one example, a system awarding benefits changed its criteria invisibly. Individuals were denied benefits due to data entry errors and other system flaws. These flaws were only revealed when an explanation of the system was demanded and produced.[31]

The truth is that AI systems are not magic, and nor are they, as some would have us believe, about to bring about the downfall of humanity. AI is technology, like so many others that have entered society before it. And like any other piece of "magical" technology – drugs, cars, planes – AI need guardrails so that we can be protected from the worst failures of the technology while still benefiting from the progress AI offers.

What we should be doing

Many proposals for guardrails exist. These include

- The Blueprint for an AI Bill of Rights[32] issued by the White House in October 2022, which lists five key principles that protect us when automated systems are deployed in ways that affect our rights, opportunities, and access to critical services. The Blueprint also provides a detailed set of expectations that systems should comply with in order to satisfy these principles;
- The AI Risk Management Framework[33] developed by the National Institute of Standards and Technology that will help those deploying and using AI systems to properly estimate risks associated with the use of the systems; and
- The AI accountability framework for Federal agencies and other entities[34] published by the General Accounting Office in 2021.

And Congress has already acted to provide some guidance, including passing

- The National AI Initiative Act and the AI in Government Act in the 116th Congress; and
- The AI Training Act and The Advancing American AI Act in the 117th Congress.

---

[30] Kristian Lum and William Isaac. To Predict and Serve? Significance. Vol. 13, No. 5, p. 14-19. Oct. 7, 2016. https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x; Aaron Sankin, Dhruv Mehrotra, Surya Mattu, and Annie Gilbertson. Crime Prediction Software Promised to Be Free of Biases. New Data Shows It Perpetuates Them. The Markup and Gizmodo. Dec. 2, 2021. https://themarkup.org/prediction-bias/2021/12/02/crime-prediction-software-promised-to-be-free-of-biases-new-data-shows-it-perpetuates-them

[31] Jay Stanley. Pitfalls of Artificial Intelligence Decisionmaking Highlighted In Idaho ACLU Case. ACLU. Jun. 2, 2017. https://www.aclu.org/blog/privacy-technology/pitfalls-artificial-intelligence-decisionmaking-highlighted-idaho-aclu-case

[32] https://www.whitehouse.gov/ostp/ai-bill-of-rights/

[33] https://www.nist.gov/itl/ai-risk-management-framework

[34] https://www.gao.gov/products/gao-21-519sp

But we need to do more if we want a society where we can enjoy the benefits of modern technology without so many of the harms. All the frameworks that I described have at their core a collection of ideas that should be the basis of legislation that places guardrails on the deployment of AI in society. These ideas are as follows:

- We should do rigorous and independent testing of automated systems to evaluate their safety, effectiveness, potential discriminatory outcomes, and other forms of impact. Evaluation should be performed before deployment, after deployment, and in an ongoing manner.
- There should be clear governance frameworks for any AI deployments that impact people, and there should be clear lines of responsibility and authority for overseeing these systems.
- Any deployment must come with a clear articulation of harms and risks in context, and a concrete focus on mitigation strategies.
- It should be very clear when algorithms are being used, and why individual decisions were made in the way they were, because without that none of the above is even possible.
- There should wherever reasonable be human alternative approaches to using automated systems, and ways for individuals to obtain human recourse when systems fail (because they will).[35]
- There should be clear and mandated reporting on all the above.

Some of these ideas have appeared in executive orders in both the Biden and Trump Administrations. In particular, the Biden Administration recently issued Executive Order 14091[36] that emphasizes a focus on equity in agencies when "designing, developing, acquiring, and using artificial intelligence", and asks agencies to remedy discrimination by "protecting the public from algorithmic discrimination".

Congress should enshrine these ideas in legislation and extend the scope of legislation not just to government uses of AI, but to private-sector uses of AI that have people-facing impact as well.

All the above examples of harms associated with the deployment of AI society were uncovered through civil advocacy, journalism, and *sociotechnical* research that brought scholars from technical disciplines, the social sciences, and the humanities together to study these "collisions" between technology and society. Such research is extremely

---

[35] This became a crucial issue recently. Individuals trying to obtain benefits from the government were required to use a third-party identity verification system. This system (based partially on facial recognition) failed to work (especially on individuals with darker skins) and there were no alternative pathways provided: in fact, people often had to wait for hours and hours on hold to reach a human operator because the system did not have appropriate means for human recourse. https://www.bloomberg.com/news/features/2022-01-20/cybersecurity-company-id-me-is-becoming-government-s-digital-gatekeeper

[36] https://www.federalregister.gov/documents/2023/02/22/2023-03779/further-advancing-racial-equity-and-support-for-underserved-communities-through-the-federal

important and has been the most effective way to identify problems and propose concrete solutions, including all the ideas I mention above.

Congress should invest in innovative sociotechnical research that will continue to uncover and mitigate the harms that accrue as our "algorithmic society" expands.

Conclusion

I'm a computer scientist, and part of my work is to imagine technological futures. There's a future in which automated technology is an assistant: it enables human freedom, liberty, and flourishing. Where the technology we build is inclusive and helps us *all* achieve our dreams and maximize our potential.

But there's another future, in which we are at the mercy of technology, which the world is shaped by algorithms and we are forced to conform. In which those who have access to resources and power control that world and the rest of us are left behind.

I know which future I want to imagine and work towards. I believe that it is our job to lay down the rules of the road – the guardrails and protections – so that we can achieve that future. And I know we can do it if we try.

Thank you for giving me the opportunity to speak.