Testimony on the New York AI Act (S1169A)

Sorelle Friedler, Ph.D.
Shibulal Family Professor of Computer Science, Haverford College
Chair, U.S. Technology Policy Committee, Association for Computing Machinery

Thanks for having me here today to discuss artificial intelligence (AI).

My name is Sorelle Friedler, and I'm the Shibulal Family Professor of Computer Science at Haverford College. I also Chair U.S. tech policy for the Association for Computing Machinery. ACM is the main professional organization for computer scientists. I formerly served as the Assistant Director for Data and Democracy at the White House Office of Science and Technology Policy. During my time at the White House, I co-authored the AI Bill of Rights and helped develop policy governing AI use across the federal government. I have done research on Responsible AI techniques for more than a decade and am a co-founder of the Conference on Fairness, Accountability, and Transparency, the leading publication venue focused on the ways that AI impacts people. Before becoming a professor, I was a software engineer at Google.

I'd like to start by talking to you today about what we can and can't expect from AI. Broadly, we can think of AI as taking data about people and historical events and finding patterns that predict and allow the replication of past actions. These patterns of the past do not perfectly predict the future. If there's one thing you take away from my testimony here today, let it be this:

AI is not designed to work all the time.

The guarantees that we make about AI as computer scientists are statistical. We might say, "this system achieves 98% accuracy." 98% sounds pretty good! It also means we got it wrong 2% of the time. A lot of AI governance is about making sure there are systems in place to handle that 2%. Given New York's population, if an AI system with a 2% error rate makes a decision about everyone in the state, that's about 400,000 people who will get the wrong result. We need plans in place to determine who will suffer from these errors and staffing to help people fix them.

Overlapping Safety Nets

Just as the AI system will sometimes fail, so too will any safeguard meant to fix these errors. Thus, good AI governance approaches layer these safeguards to create an overlapping safety net, so that as few people as possible will slip through the cracks. I'll motivate these layers with a few examples.

In my own work and work of colleagues we have found that predictive policing systems are regularly incorrect in discriminatory ways.[1] Instead of identifying locations where crime will take place in the future, they repeatedly send police back to the same neighborhoods where they've made arrests in the past. If historical policing and arrests were more likely to occur in predominantly Black neighborhoods, this will ensure future

[1] https://proceedings.mlr.press/v81/ensign18a.html; https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x

policing and arrests match that biased pattern, *regardless of the actual locations of high crime areas*.

Safeguards needed to help protect the public against these errors include:

1. An audit and transparency into the **safety and efficacy** of the AI system. This is necessary for understanding the expected error rate, and thus for setting up the rest of the AI governance effort.
2. An assessment of the system for **algorithmic discrimination**. AI replicates the biases of the past that appear in its training data, so system errors are likely to reproduce discrimination.

Next, consider a real-life example (reported by Wired[2]) about the opioid crisis. A woman in serious pain was in the hospital. Her doctor prescribed opioids, yet after a few days, she was cut off from pain medication, discharged from the hospital, and her doctor terminated their relationship in a letter referencing a "report from the NarxCare database," but no further explanation was given. NarxCare is a database and AI system that is supposed to flag patients at high risk of an opioid addiction or overdose, but she couldn't figure out why it was flagging her, nor could her doctors. Concerns have been raised by health policy experts about the effectiveness of NarxCare and other opioid overdose risk algorithms that are in wide use.[3] This is a question we should know the answer to, but investigating such systems without direct access to the AI systems or associated data themselves is quite hard. Many patients don’t even know such systems are being used, or what data they’re basing decisions on. In the case of the woman in the hospital, it turns out her dogs' medications had been entered into the database under her name!

In addition to testing for safety and efficacy, it would have helped to have the following safeguards:

3. A process for **human intervention** to manually identify and fix system problems, including human review and appeal processes.
4. **Transparency** into the AI system, assessment findings, and governance processes. Audits are needed to determine the efficacy of a system and whether it’s likely to be discriminatory. Notice and explanation are needed for impacted individuals to exercise their rights and receive redress via human intervention. If you don’t know an AI system has been used on you, how can you ask someone to fix its results?

5. Finally, for all these cases a **governance and risk mitigation plan** is necessary for clear lines of responsibility and oversight of these AI governance processes, and
6. **enforcement** is key in ensuring these important steps are actually followed.

[2] https://www.wired.com/story/opioid-drug-addiction-algorithm-chronic-pain/
[3] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10531142/

Prohibiting AI Use

AI systems will fail on some inputs. If those failures can't be managed, AI shouldn't be used. There are times when the right answer will be to *prohibit AI use*. In cases where errors would be damaging or hard to catch, AI can't be the answer. For example, we should be cautious about AI tutors in our classrooms — they may confidently assert falsehoods, spreading lies in our schools.

In other cases, even systems working perfectly may be damaging to society. In the bill, you prohibit **social scoring systems**. Such systems have the potential to degrade societal trust, limit opportunities, and limit the freedom of speech. Prohibiting them is thus a reasonable response to the potential societal danger they pose. I encourage you to consider whether other AI systems should also be prohibited. For example, some states have **limited facial recognition use by law enforcement** in investigations to specific crimes and required a warrant and notice.[4] Others have argued that the use of **affective computing by law enforcement** – especially the purported ability to determine emotions and lies from a photo or video of someone's face – should be banned entirely.[5] It doesn't work, is unlikely to ever work, and even if it *did* work would be damaging to freedom of expression.

Similarity to Other AI Governance Efforts

Don't let others tell you that the AI governance steps you propose in this bill are unusual, unworkable, or otherwise outside of the norm. The safeguards outlined in this bill have become the standard set of safeguards in AI governance bills across the country and many are already requirements for any AI used by the federal government. Both the Biden and Trump Administrations issued federal guidance for the use of AI – whether developed in house or procured. I was involved in the Biden Administration development of the guidance. Despite the differences between the administrations, their AI guidance documents are remarkably similar.[6]

Both the Biden and Trump Administration AI guidance documents require:

- Scoping of included AI systems based on high-impact use cases
- An AI impact assessment, including safety and efficacy testing
- An independent review of the system
- Human oversight and intervention
- Timely human review and appeal processes
- Consultation with impacted users
- Public transparency into the AI use cases, findings of the AI impact assessment, and in-place AI governance processes

[4] https://www.techpolicy.press/status-of-state-laws-on-facial-recognition-surveillance-continued-progress-and-smart-innovations/

[5] https://www.brookings.edu/articles/why-president-biden-should-ban-affective-computing-in-federal-law-enforcement/

[6] https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5346150; https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-21-Accelerating-Federal-Use-of-AI-through-Innovation-Governance-and-Public-Trust.pdf; https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf

The Biden Administration guidance additionally included:

- Assessment and mitigation of algorithmic discrimination

Any companies that have AI systems used by the federal government are required to have the Trump Administration AI governance processes in place by April of this year, and the previous Biden Administration processes were required by December of 2024. Many companies should have already implemented these AI governance requirements, which are similar to the ones in this bill.

Thanks again for inviting me to speak, and for this important work you're doing for New York.