Testimony on Artificial Intelligence for Philadelphia City Council

Sorelle Friedler

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). As a Philly resident, I'm especially glad to see that you're taking seriously the way that AI can be used to make decisions that directly impact all of our lives.

My name is Sorelle Friedler, and I'm the Shibulal Family Professor of Computer Science at Haverford College. I'm also the Chair of the U.S. Technology Policy Committee for the Association for Computing Machinery — the main professional organization for computer scientists — and 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[1] and helped develop policy governing AI use across the federal government.[2] 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, one of the main publication venues focusing on the ways that AI impacts people. Before becoming a professor, I was a software engineer at Google.

I've been asked to talk to you today about what AI is. As policymakers, you don't need to understand the detailed formula or specific algorithms behind these systems — what you do need are analogies that you can rely on, and that build on your existing expertise as you make decisions for our city.[3] I want to give you a few of those analogies today.

AI is like the past. 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.

---

[1] White House Office of Science and Technology Policy, AI Bill of Rights. https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/

[2] OMB Guidance M-24-10, Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence 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

[3] The analogies described in this testimony are explained more fully in this article: https://equitablegrowth.org/analogies-for-ai-policymaking/

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 our city's population, if an AI system with a 2% error rate makes a decision about everyone in the city, that's more than 31,000 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.

This also means 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 order for you all to have control and make sure AI systems used by the Administration actually do the job, we need careful guardrails[4] including:

- testing requirements,
- human review processes, and
- transparency into how AI is being used in our city.

AI is like redlining. Since AI is like the past, it will replicate the biases of the past. In my own work and work of colleagues we have found that predictive policing systems have this problem.[5] 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 policing and arrests match that biased pattern, *regardless of the actual locations of high crime areas*.

AI is like a line of best fit. To better understand how an AI system works behind the scenes, I think it's useful to think of a line of best fit. You take data with two attributes —

---

[4] For more details on these and other guardrails, see the Biden Administration's guidance to the federal government on guardrails for AI use from OMB: 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

[5] To Predict and Serve? by Kristian Lum and William Isaac. *Significance*, 2016. https://academic.oup.com/jrssig/article/13/5/14/7029190 ; Runaway Feedback Loops in Predictive Policing by Danielle Ensign, Sorelle A. Friedler, Scott Neville, Carlos Scheidegger, and Suresh Venkatasubramanian. *FAccT*, 2018. https://proceedings.mlr.press/v81/ensign18a.html

x and y — and plot it. Then you draw a line through that data. This is a way to identify the relationship between two attributes, perhaps home values 5 years ago and home values now. It can then be used to predict resulting data in the future, say, home values in 5 years. Modern AI systems are built on a large number of building blocks that work essentially like this.

AI is like a flow chart. To handle more complex patterns, AI systems operate like a giant flow chart with the results of a computation, like a line of best fit, determining which flow chart step comes next. This could allow us to make different home value predictions for different neighborhoods. The specific computations performed at each step are determined based on "training," which is the word for taking data from the past and finding patterns in it.

AI is like the web remixed. Often, when we're talking about AI these days, we're talking about generative AI and chatbots like ChatGPT. I think it's useful to focus on the data used to create these systems. AI companies gathered all the data from the web they could get their hands on — old newspaper articles, reddit posts, YouTube comment sections, porn sites, conspiracy theories — everything. We can think of the results as the web remixed, with prompts to the chatbot providing information about which parts of the web to remix and show to the user.

AI is like a factory. One final analogy: AI is like the modernization of a factory. It replaces people's jobs with automation. The data centers that run AI take a huge physical footprint. They spew pollution into the surrounding neighborhoods, and — as we have already seen in Philly — take so much energy that they increase all of our electricity costs. Yet there's one important way AI is *not* like a factory: it doesn't create long-term jobs *here*. After construction, data centers are mostly empty of people, with just a few security guards.

I hope as you provide needed oversight to city use of AI, you remember that *this technology is not magic* — you *can* expect to be able to have your questions about it answered, have transparency into how these systems are being used, and take control over whether AI is working for us.

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