TechXplore

White House science adviser calls for more safeguards against artificial intelligence risks

August 21 2023, by Matt O'brien

Credit: AP Illustration/Peter Hamlin

When President Joe Biden has questions about artificial intelligence, one expert he turns to is his science adviser Arati Prabhakar, director of the White House Office of Science and Technology Policy.

Prabhakar is helping to guide the U.S. approach to safeguarding AI technology, relying in part on cooperation from big American tech firms like Amazon, Google, Microsoft and Meta.

The India-born, Texas-raised engineer and applied physicist is coming at the problem from a career bridging work in government—including leading the Defense Department's advanced technology research arm—and the private sector as a former Silicon Valley executive and venture capitalist.

She spoke with The Associated Press earlier this month ahead of a White House-organized test of AI systems at the DefCon hacker convention in Las Vegas. The interview has been edited for length and clarity.

Q: Does the president come to ask you about AI?

A: I've had the great privilege of talking with him several times about artificial intelligence. Those are great conversations because he's laser-focused on understanding what it is and how people are using it. Then immediately he just goes to the consequences and deep implications. Those have been some very good conversations. Very exploratory, but also very focused on action.

Q: Senate Majority Leader Chuck Schumer (who's pushing for AI regulations) says making AI models explainable is a priority. How realistic is that?

A: It's a technical feature of these deep-learning, machine-learning systems, that they are opaque. They are black box in nature. But most of the risks we deal with as human beings come from things that are not explainable. As an example, I take a medicine every single day. While I can't actually predict exactly how it's going to interact with the cells in my body, we have found ways to make pharmaceuticals safe enough. Think about drugs before we had clinical trials. You could hawk some powder or syrup and it might make you better or it might kill you. But when we have clinical trials and a process in place, we started having the technical means to know enough to start harnessing the value of pharmaceuticals. This is the journey we have to be on now for artificial intelligence. We're not going to have perfect measures, but I think we can get to the point where we know enough about the safety and effectiveness of these systems to really use them and to get the value that they can offer.

Q: What are some specific AI applications you're concerned about?

A: Some of the things we see are big and obvious. If you break the guardrails of a chatbot, which people do routinely, and coax it to tell you how to build a weapon, well, clearly that's concerning. Some of the harms are much more subtle. When these systems are trained off human data, they sometimes distill the bias that's in that data. There's now a fairly substantial, distressing history of facial recognition systems being used inappropriately and leading to wrongful arrests of Black people. And then privacy concerns. All of our data that's out in the world, each individual piece may not reveal much about us, but when you put it all together it tells you rather a lot about each of us.

Q: Seven companies (including Google, Microsoft and ChatGPT-maker OpenAI) agreed in July to meet voluntary AI safety standards set by the White House. Were any of those commitments harder to get? Where's the friction?

A: I want to start by just acknowledging how fortunate we are that so many of the companies that are driving AI technology today are American companies. It reflects a long history of valuing innovation in this country. That's a tremendous advantage. We also just have to be very, very clear that with every good intention in the world, the realities of operating in the market are, by definition, a constraint on how far these individual companies can go. We hope many more will join them and voluntary commitments will grow. We just have to be clear that's only one part of it. That's companies stepping up to their responsibilities. But we in government need to step up to ours, both in the executive branch and for the legislative branch.

Q: Do you have a timeline for future actions (such as a planned Biden executive order)? Will it include enforceable accountability measures for AI developers?

A: Many measures are under consideration. I don't have a timeline for you. I will just say fast. And that comes directly from the top. The president has been clear that this is an urgent issue.

Citation: White House science adviser calls for more safeguards against artificial intelligence risks (2023, August 21) retrieved 9 May 2024 from https://techxplore.com/news/2023-08-white-house-science-safeguards-artificial.html