Regulating AI Is a Good Start. Next, Tackle the Secretive Way Government Uses It

Brandon L. Garrett and Cynthia Rudin

Published in The Messenger November 2, 2023

Safe, secure and trustworthy: Those are the terms the White House used to title its just-released executive order on artificial intelligence (AI), which marks an important milestone in U.S. engagement with the rapid spread of AI technology in law and society. The title is appropriate: Until now, AI has threatened safety, security and trust, with very little law to regulate it.

We are scholars who study AI technology and how courts engage with science. We view this order as a welcome and impressive broad roadmap for change, particularly since it calls for federal agencies to begin detailed assessments and reviews. Yet, regulating the industry will not be enough. Some of AI’s most sinister capabilities have been manifested in law enforcement and in our courts, where the primary user has been the government itself.

As we have recently written, one of the most troubling trends in criminal investigations is the growing use of “black box” technology, in which law enforcement rely on AI models or algorithms that are either too complex for people to understand or they simply conceal how it functions.

These black box systems can be biased, error-prone and unfair, yet the federal response so far has given great cause for concern. The federal government uses a range of AI systems that are “black box” and of unknown reliability and fairness in criminal proceedings.

Take one example, a criminal risk assessment used to predict the likelihood that a person will reoffend. The stakes may be enormous for the person involved if they are jailed or imprisoned because the system labels them as high risk.

Some risk assessment instruments are not black box; instead, they are simple and interpretable, meaning people know what they did and why. These can be depicted in a simple worksheet that assigns points based on certain factors, like the person’s age, prior offenses and current offense. A social worker, lawyer or judge can easily see how much weight each factor has, even if they may not understand how the data were used to generate the scheme or how accurate it is.

A mature body of computer science research has shown how this type of “glass box” AI — designed to be interpretable — can be more accurate than black box alternatives. Indeed, black box AI performs predictably worse in settings like the criminal system. For a black box system, there is no way to know whether the system relied on erroneous data. In fact, criminal history information can even be entered inaccurately, which can result in serious errors.

Yet, President Biden’s executive order does not address the critical issue of interpretability. It does not explicitly discuss the need for people to actually understand and have sound information about how AI functions in general or in a given case. The executive order discusses more broadly the need for “notice” that AI is being used at all, which is a start, to be sure, but far from sufficient.

This lack of interpretability is all the more important when AI systems are used without guidance or training. In the most recent report on the topic, the U.S. Government Accountability Office noted that
federal agencies have not consistently trained officials on how to use facial recognition, and that the Department of Justice (DOJ) has faced delays in issuing a policy on protecting people’s civil rights when using facial recognition.

It raises serious constitutional questions if black box AI is used in criminal investigations in ways that no one, whether it is law enforcement, lawyers, jurors or judges can understand – even if the AI is used properly. Judges are beginning to grapple with these questions, and some have ordered disclosures, but mostly in state courts. The federal government should be a leader in protecting constitutional rights when AI is used in criminal cases, and not a high-profile example of use of AI without training, policy or oversight.

Hopefully, the attorney general and DOJ will grapple with these questions and provide clear, forceful guidance to federal law enforcement and prosecutors. The order asks the attorney general to investigate how law enforcement and the DOJ are using AI, including for forensic analysis, police surveillance, risk assessments and more, to produce a year-end report with recommendations for best practices and safeguards. This will be important work, and there is much that the DOJ can do.

And there are good models for reform. Take discovery as an example. In court proceedings, federal prosecutors already share, in discovery, information about forensic expert evidence and electronic communications. While those guidelines do not address AI, these federal criminal discovery rules could be amended to reflect the need for AI discovery.

Criminal cases provide a test of any society’s evenhanded protection of constitutional rights. This order sensibly calls on the attorney general and the DOJ to take stock of how AI is being used and to report on next steps. There is much more to be done, but in the next year, we hope that a comprehensive and forceful plan takes shape that, among its contributions, requires interpretable AI.

Brandon Garrett is the L. Neil Williams professor of law at Duke University School of Law and the faculty director of the Wilson Center for Science and Justice at Duke Law.

Cynthia Rudin is the Earl D. McLean, Jr. professor of computer science and engineering at Duke University and directs the Interpretable Machine Learning Lab.