Research group releases white papers on governance of AI

December 11 2023, by Peter Dizikes

An MIT ad hoc committee has released a new set of policy papers about the governance of artificial intelligence. Credit: Jake Belcher

Providing a resource for U.S. policymakers, a committee of MIT leaders and scholars has released a set of policy briefs that outlines a framework

for the governance of artificial intelligence. The approach includes extending current regulatory and liability approaches in pursuit of a practical way to oversee AI.

The aim of the papers is to help enhance U.S. leadership in the area of artificial intelligence broadly, while limiting harm that could result from the new technologies and encouraging exploration of how AI deployment could be beneficial to society.

The main policy paper, "A Framework for U.S. AI Governance: Creating a Safe and Thriving AI Sector," suggests AI tools can often be regulated by existing U.S. government entities that already oversee the relevant domains. The recommendations also underscore the importance of identifying the purpose of AI tools, which would enable regulations to fit those applications.

"As a country we're already regulating a lot of relatively high-risk things and providing governance there," says Dan Huttenlocher, dean of the MIT Schwarzman College of Computing, who helped steer the project, which stemmed from the work of an ad hoc MIT committee. "We're not saying that's sufficient, but let's start with things where human activity is already being regulated, and which society, over time, has decided are high risk. Looking at AI that way is the practical approach."

"The framework we put together gives a concrete way of thinking about these things," says Asu Ozdaglar, the deputy dean of academics in the MIT Schwarzman College of Computing and head of MIT's Department of Electrical Engineering and Computer Science (EECS), who also helped oversee the effort.

The project includes multiple additional policy papers and comes amid heightened interest in AI over last year as well as considerable new industry investment in the field. The European Union is currently trying

to finalize AI regulations using its own approach, one that assigns broad levels of risk to certain types of applications. In that process, general-purpose AI technologies such as language models have become a new sticking point. Any governance effort faces the challenges of regulating both general and specific AI tools, as well as an array of potential problems including misinformation, deepfakes, surveillance, and more.

"We felt it was important for MIT to get involved in this because we have expertise," says David Goldston, director of the MIT Washington Office. "MIT is one of the leaders in AI research, one of the places where AI first got started. Since we are among those creating technology that is raising these important issues, we feel an obligation to help address them."

Purpose, intent, and guardrails

The main policy brief outlines how current policy could be extended to cover AI, using existing regulatory agencies and legal liability frameworks where possible. The U.S. has strict licensing laws in the field of medicine, for example. It is already illegal to impersonate a doctor; if AI were to be used to prescribe medicine or make a diagnosis under the guise of being a doctor, it should be clear that would violate the law just as strictly human malfeasance would. As the policy brief notes, this is not just a theoretical approach; autonomous vehicles, which deploy AI systems, are subject to regulation in the same manner as other vehicles.

An important step in making these regulatory and liability regimes, the policy brief emphasizes, is having AI providers define the purpose and intent of AI applications in advance. Examining new technologies on this basis would then make clear which existing sets of regulations, and regulators, are germane to any given AI tool.

However, it is also the case that AI systems may exist at multiple levels, in what technologists call a "stack" of systems that together deliver a particular service. For example, a general-purpose language model may underlie a specific new tool. In general, the brief notes, the provider of a specific service might be primarily liable for problems with it. However, "when a component system of a stack does not perform as promised, it may be reasonable for the provider of that component to share responsibility," as the first brief states. The builders of general-purpose tools should thus also be accountable should their technologies be implicated in specific problems.

"That makes governance more challenging to think about, but the foundation models should not be completely left out of consideration," Ozdaglar says. "In a lot of cases, the models are from providers, and you develop an application on top, but they are part of the stack. What is the responsibility there? If systems are not on top of the stack, it doesn't mean they should not be considered."

Having AI providers clearly define the purpose and intent of AI tools, and requiring guardrails to prevent misuse, could also help determine the extent to which either companies or end users are accountable for specific problems. The policy brief states that a good regulatory regime should be able to identify what it calls a "fork in the toaster" situation—when an end user could reasonably be held responsible for knowing the problems that misuse of a tool could produce.

Responsive and flexible

While the policy framework involves existing agencies, it includes the addition of some new oversight capacity as well. For one thing, the policy brief calls for advances in auditing of new AI tools, which could move forward along a variety of paths, whether government-initiated, user-driven, or deriving from legal liability proceedings. There would

need to be public standards for auditing, the paper notes, whether established by a nonprofit entity along the lines of the Public Company Accounting Oversight Board (PCAOB), or through a federal entity similar to the National Institute of Standards and Technology (NIST).

And the paper does call for the consideration of creating a new, government-approved "self-regulatory organization" (SRO) agency along the functional lines of FINRA, the government-created Financial Industry Regulatory Authority. Such an agency, focused on AI, could accumulate domain-specific knowledge that would allow it to be responsive and flexible when engaging with a rapidly changing AI industry.

"These things are very complex, the interactions of humans and machines, so you need responsiveness," says Huttenlocher, who is also the Henry Ellis Warren Professor in Computer Science and Artificial Intelligence and Decision-Making in EECS. "We think that if government considers new agencies, it should really look at this SRO structure. They are not handing over the keys to the store, as it's still something that's government-chartered and overseen."

As the policy papers make clear, there are several additional particular legal matters that will need addressing in the realm of AI. Copyright and other intellectual property issues related to AI generally are already the subject of litigation.

And then there are what Ozdaglar calls "human plus" legal issues, where AI has capacities that go beyond what humans are capable of doing. These include things like mass-surveillance tools, and the committee recognizes they may require special legal consideration.

"AI enables things humans cannot do, such as surveillance or fake news at scale, which may need special consideration beyond what is applicable

for humans," Ozdaglar says. "But our starting point still enables you to think about the risks, and then how that risk gets amplified because of the tools."

The set of policy papers addresses a number of regulatory issues in detail. For instance, one paper, "Labeling AI-Generated Content: Promises, Perils, and Future Directions," by Chloe Wittenberg, Ziv Epstein, Adam J. Berinsky, and David G. Rand, builds on prior research experiments about media and audience engagement to assess specific approaches for denoting AI-produced material. Another paper, "Large Language Models," by Yoon Kim, Jacob Andreas, and Dylan Hadfield-Menell, examines general-purpose language-based AI innovations.

'Part of doing this properly'

As the policy briefs make clear, another element of effective government engagement on the subject involves encouraging more research about how to make AI beneficial to society in general.

For instance, the policy paper, "Can We Have a Pro-Worker AI? Choosing a path of machines in service of minds," by Daron Acemoglu, David Autor, and Simon Johnson, explores the possibility that AI might augment and aid workers, rather than being deployed to replace them—a scenario that would provide better long-term economic growth distributed throughout society.

This range of analyses, from a variety of disciplinary perspectives, is something the ad hoc committee wanted to bring to bear on the issue of AI regulation from the start—broadening the lens that can be brought to policymaking, rather than narrowing it to a few technical questions.

"We do think academic institutions have an important role to play both in terms of expertise about technology, and the interplay of technology

and society," says Huttenlocher. "It reflects what's going to be important to governing this well, policymakers who think about social systems and technology together. That's what the nation's going to need."

Indeed, Goldston notes, the committee is attempting to bridge a gap between those excited and those concerned about AI, by working to advocate that adequate regulation accompanies advances in the technology.

As Goldston puts it, the committee releasing these papers is "is not a group that is antitechnology or trying to stifle AI. But it is, nonetheless, a group that is saying AI needs governance and oversight. That's part of doing this properly. These are people who know this technology, and they're saying that AI needs oversight."

Huttenlocher adds, "Working in service of the nation and the world is something MIT has taken seriously for many, many decades. This is a very important moment for that."

More information: Paper series: computing.mit.edu/ai-policy-briefs/

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology

Citation: Research group releases white papers on governance of AI (2023, December 11) retrieved 10 May 2024 from https://techxplore.com/news/2023-12-group-white-papers-ai.html