More than Meets AI—Part II: Building Trust, Managing Risk

Edited by Michael J. Keegan

The next contribution to this forum explores the benefits of AI, but also underscores the importance for government agencies to manage real and perceived risks associated with AI. What follows is excerpted from the report *More Than Meets AI—Part II: Building Trust, Managing Risk,* with a focus on significant challenges such as bias, security, transparency, employee knowledge, and federal budget and procurement processes.

Introduction

Many Americans have questions about effects AI technologies may have on aspects of their lives. According to an October 2018 survey, 59 percent of respondents are "very concerned" or "somewhat concerned" with job loss and displacement worries ranking highest. They also conveyed concerns about data privacy, security, hacking, and the safety of AI systems. Although these risk factors also affected public perceptions when other technologies were introduced, leaders now need to also address these concerns to foster trust as agencies rely more on AI to carry out missions. Through an executive order, an AI summit, and the creation of a website and a White House Select Committee on AI, the Office of Management and Budget and the Office of Science and Technology Policy are leading a governmentwide effort to maximize AI's benefits, while laying the groundwork for agencies to address risks responsibly. To increase the trust the public and federal employees have in government's use of AI tools, the government's strategy deals with transparency, security, technological know-how, procurement, budgeting, and risk management.

Understanding and Addressing AI Risks

As agencies integrate AI into their work, they will have to pay attention to issues ranging from the ethical to the practical. Top challenges include bias, security, transparency, employee knowledge about AI technology, and federal budget and procurement processes. Each of these challenges is discussed below, along with recommendations for how agencies could address potential concerns and develop strategies to mitigate them.

It is important for federal organizations to move forward with implementing AI technologies as they address AI risks. Their approach to lessening AI risks also must evolve rapidly if they hope to use AI to address government's most pressing challenges.

- **Bias.** Bias in AI outcomes can stem from a number of issues, including poor-quality data, limited amounts of data, or data that doesn't fully represent all aspects of a matter. Knowing that biased data may lead to biased results, agencies need to pay special attention to what information is being used with these new technologies. To address AI bias, federal organizations need employees with technical acumen and data analysis and interpretation skills who can detect data bias and inaccuracies. Experts in government need to understand the theory behind AI, how the algorithms work, and how conclusions are reached. Under the White House's February 2019 AI executive order, the National Institute of Standards and Technology (NIST), researchers are exploring ways to test and measure AI security and trustworthiness. As part of its task, the agency is working with international partners to explore the potential for global AI standards. These and similar efforts should include creating a framework for assessing bias.

- **Security.** AI is vulnerable in several ways if designed without proper security measures. AI's potentially widespread impact amplifies cybersecurity concerns. If AI systems are driving cars, fighting wars, and the like, hackers who can compromise these systems have greater capacity to do enormous damage more quickly. Attacks could alter AI training data or introduce corrupted or incorrect data that changes the conclusions of the AI tool. Hackers also could act to reveal personally identifiable information in the data on which an AI tool was trained. With security paramount, the Defense Department is investigating how to safeguard AI technology from attacks. In a 2018 strategy, the department committed to fund research and development of reliable and secure AI systems, but more work is needed to evaluate the security of AI technologies. Our government and governments in other countries could share knowledge and lessons learned, as security concerns are global in nature. Given these interconnected security implications, government has to ensure data safety and spend some time reassuring people that our cybersecurity is very much up to scratch.
- **Transparency.** With AI, agencies have the ability to accomplish activities more quickly and accurately. By making AI transparent, users can learn how and why the tool arrived at a conclusion and what data the AI technology used. Lack of transparency can pose issues when people want an explanation for why decisions were made. Some AI algorithms are proprietary; others are so complex that it is hard to explain, or for people to understand, how conclusions were reached. Without clarity about how AI produces its recommendations and conclusions or understanding from employees as to how to explain results derived from AI technology, governments may risk losing the public's trust. The AI research and development community recognizes that transparency will promote trust in AI systems. Researchers are looking into explainable AI and making AI algorithms and results less of a black box. This will enable governments and others that incorporate AI into their processes to respond to questions about the decisions involving AI technology.
- **Employee knowledge.** Maximizing AI benefits while managing AI risks hinges on hiring or training employees who understand and use the technology responsibly. Getting enough of the workforce up to speed is critical, but government often faces funding and other challenges—and often falls short on AI training and education. The federal government should emphasize expertise in technical, digital, and data skills. It should provide extensive and ongoing training to employees so they can create, understand, manage, and work with AI technology.
- **Federal budget and procurement processes.** Outdated federal acquisition and budget processes prevent agencies from buying and deploying new technology quickly and efficiently. Since most agencies start budgeting two years in advance, they often do not have the flexibility or "clairvoyance" to buy the newest technologies. The typical acquisition process involves purchasing a finished product or service, yet many AI applications are iterative, improving over time through experience. The rapid pace of AI development and improvement can leave government lagging behind. AI is moving fast, so should governments. Agencies should obtain what they need for AI by taking full advantage of the tools and flexibilities available in the budget and procurement processes. For example, agencies could use "try before you buy" acquisitions that allow them to experiment with new tools on a small scale, or staged contracts to evaluate proposals and pilot tools before investing in full.

Lessons from Canada on Maximizing AI Benefits and Managing Risks

The AI research and development community considers Canada to be at the forefront among governments at managing AI risks. The Canadian government has taken steps to ensure its departments and agencies have tools, rules and people to use AI responsibly. Based on the Canadian government's experiences, U.S. government agencies will need to balance regulation and oversight with support for private sector research, development and innovation. Canada's example outlines potential tools, rules, and people issues for consideration.

- **Tools: Simplify buying credible AI products.** In September 2018, in order to procure AI faster and more efficiently, the Canadian government released a list of more than 70 suppliers proficient in AI and AI ethics. The government deemed these qualified vendors to have delivered a successful AI product or service.
- **Rules: Create a framework to assess the risk of using AI in government.** According to an April 2019 Canadian government directive, if a department or agency is

using automated decision-making in support of service delivery, it is required to assess the associated risks. The government developed four levels of impact an AI tool might have on society and government, ranging from little to no impact that could be "reversible and brief" to very high impact, which might lead to "irreversible" and "perpetual" changes.

For use of AI with little or no impact on service programs, the directive allows for the possibility of automated end-to-end decision-making—in other words, making decisions without human involvement. However, it states that program officials must be able to explain how conclusions were reached. Requirements for AI used by high-impact programs, on the other hand, include a peer review by government academics, nongovernment organizations or other advisory boards; repeated training for employees using the AI tool; and documentation posted on relevant websites describing how the tool works. In addition, a person must make any final decisions based on an AI tool's recommendation. Depending on the impact level, programs also must disclose to the citizen whether a decision affecting them is made partly or wholly by an AI tool. The directive also addresses AI transparency and the Canadian government's right to access and test proprietary AI systems if necessary for a specific audit, investigation, inspection, examination, enforcement action, or judicial proceeding.

- **People: Train public servants on how to use AI tools.** In January 2019, to address a skills gap and ensure government programs use AI tools responsibly, the Canada School of Public Service launched a pilot cohort of its public sector Digital Academy. It is seeking to improve the digital acumen of public servants at all levels and eventually expanding training to all public employees. Elevating the digital literacy of employees can help them get more comfortable with new technologies. Aside from providing digital, data, and AI skills, the government hopes the training eases concerns by raising awareness among public servants about the current state of AI and other digital technologies, and how they could affect their jobs and even private lives.

Conclusion

AI tools also are expected to impact the federal government substantially, with implications for federal systems and structures. To capture the benefits of AI, federal agencies must be prepared to address related risks. The Office of Management and Budget and Office of Science and Technology Policy should continue to lead efforts to manage those risks, given the technology's potential to transform work governmentwide.