Can Government Manage Risks Associated with Artificial Intelligence?

Daniel J. Chenok, IBM Center for The Business of Government

Artificial intelligence can help government agencies deliver better results, but there are underlying risks and ethical issues with its implementation that need to be resolved before AI becomes part of the fabric of government. Based on insights from an expert roundtable led by the IBM Center for The Business of Government and the Partnership for Public Service, agencies will need to address multiple risks and ethical imperatives in order to realize the opportunity that AI technology brings. These include:

- **Creating Explainable Algorithms.** Machine Learning (ML) algorithms are only as good as the data provided for training. Users of these systems can take data quality for granted and can come to over-trust the algorithm's predictions. Additionally, some ML models such as deep neural networks are difficult to interpret, making it hard to understand how a decision was made (often referred to as a "black box" decision). Another issue arises when low-quality data (i.e., data that embeds bias or stereotypes or simply does not represent the population) is used in un-interpretable models, making it harder to detect bias. On the other hand, well-designed, explainable models can increase accuracy in government service delivery, such as a neural network that could correct an initial decision to deny someone benefits to which they are entitled.

Research into the interpretability of neural networks and other kinds of models will help build trust in AI.   More broadly, educating stakeholders about AI – including policymakers, educators, and even the general public -- would increase digital literacy and provide significant benefits. While universities are moving forward with AI education, the government needs a greater understanding of how data can impact AI performance. The government, industry, and academia can work together in explaining how sound data and models can both inform the ethical use of AI.

- **Applying Ethics within a Cost-Benefit Framework.** AI ethics is to AI policy as political philosophy is to law or regulation. In other words, AI ethics is less a problem to "solve" than a set of norms and frameworks that inform decisions. Therefore, AI ethics should be applied in specific contexts (e.g., for what reasons and for who is AI used?) and levels of understanding (e.g., how much AI is appropriate for a given scenario?).

One way to apply ethics involves the practice within policy-making of cost-benefit analysis. Such methodologies would allow agencies to compare the risks associated with AI (e.g., potential for human harm, discrimination, funds lost) with the benefits (e.g., lives saved, egalitarian treatment, funds saved) throughout the lifecycle of an algorithm's development and operation. Cost-benefit analyses often include scenario planning and confidence intervals, which could work well in evaluating AI systems over time—provided that the "costs" considered include not only quantifiable financial costs but more intangible, value-based risks as well (such as avoiding bias or privacy harms). Done correctly, this approach could provide a clear way to communicate risk and decisions about how and when to use AI -- including risks of leveraging AI to support a decision, relative to risks of decisions based solely on human analysis – in a way that informs public understanding and dialogue.

- **Creating a Data Governance Framework.** To ensure confidence in data used by AI and other automated decision systems, proper data governance including testing and audits will be necessary. This testing could focus on data quality, security, and data user rights, and ensure that automated decision systems are not discriminatory. Another element of data governance could set up protocols for inter-agency data sharing, which can increase efficiency but also introduce privacy risk -- privacy protection within AI has engendered divergent views, and using a risk management perspective allows agencies to assess how much personal data they need to collect and store based on the benefits to the data subjects.

Finally, governance protocols can help clarify how and when to acquire data, especially from outside parties. The government's use of third-party data to support the use of AI for regulatory decisions may differ from that used for research or analysis. And there may be inherent vulnerabilities in third-party data that call for mitigation strategies in a risk management framework.

•**Crafting A Strategic Vision for AI Use.** Government employees often lack familiarity with AI technology and are understandably skeptical about its impact on their work. Greater engagement across agencies in setting forth needs and priorities, defining factors that can promote trust in AI systems, and developing pathways to explain the technology, could enhance understanding of AI's impact across the public sector. Sharing best practices is especially important given the differential levels of maturity in AI use across agencies and even levels of government. With a greater understanding of and involvement in the technology, agencies can promote an understanding of the benefits and risks associated with AI and foster a cultural shift in supporting responsible AI adoption.

Conclusion. Overall, there is a need for more education, understanding, and skills involving AI; better data to inform AI systems; and clearer guidelines for responsible AI implementation. At the same time, a perfect AI system, one free of all ethical constraints, will never be actualized. The more practical path forward in understanding and addressing underlying risks is to engage early with developing AI prototypes, iterate on the solutions, and learn from the results. The risk of not using AI may be greater than any risk from responsible and ethically designed systems. Risk-based approaches that include public dialogue provide a starting point.

In the end, AI is about data. Agencies can build a greater understanding of the benefits and risks of AI, and help employees and citizens within an ethical framework, by reaching decisions that emerge from responsible use of the right data.