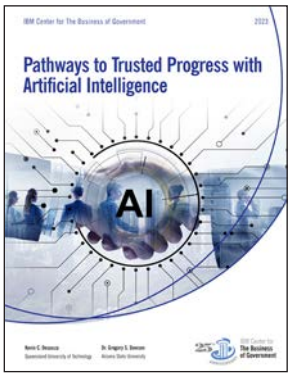

Pathways to Trusted Progress with Artificial Intelligence

Edited by Michael J. Keegan

The second contribution in this forum examines the governance and applications of AI and how governments need to develop and communicate a framework for the public to understand why AI is being used; what has been done to ensure that the AI is fair, transparent, and accurate; what experiments were done to ensure that the output is reliable; and how public value from AI is being measured and created.

What follows is an excerpt from the IBM Center report *Pathways to Trusted Progress with Artificial Intelligence* by Professors Kevin DeSouza and Greg Dawson that distills findings and recommendations from an expert roundtable of leaders in Australia that address the needs, security, and progress of delivering AI services that benefit citizens and industry. Though this report may offer perspectives from leaders in Australia, the insights present in the report and summarized here are applicable across the world. By addressing the growth and management of AI, and the governance of data aligned to AI strategies, government can take full advantage of the power of AI.

Digital Transformation Initiatives are Revolutionizing All Aspects of the Public Sector.

AI systems will play an important role in transforming government as well as the national economy. Realizing AI's potential will only occur if there is a concerted effort to ensure that citizens trust AI systems, the government, and the government use of AI.

There is a wide assortment of AI systems, and each class of AI systems has their own characteristics. However, at their core, these systems ingest vast swaths of data, employ either supervised or unsupervised learning techniques or both, and can be deployed autonomously, semiautonomously, or in an advisory capacity to augment human decision makers. Consider the following three examples of AI systems successfully deployed in the public sector:

- **Fully Autonomous:** In North Carolina, AI-powered chatbots independently manage simple, repetitive customer service tasks, such as password resets. By automating these interactions without human oversight, the system frees up human agents to handle more complex inquiries.
- **Semiautonomous:** Semiautonomous robots deliver food on college campuses, where students initiate orders via an app. Humans set initial parameters (restaurant and delivery location), but the robots independently navigate routes and avoid obstacles. This technology may expand to roles in search and rescue.
- **Augmented Decision Making:** In African wildlife conservation, AI assists rangers by predicting poaching locations and recommending patrol routes. Here, the AI provides data-driven insights, while humans make final decisions on the patrol plan.

Each of these examples illustrates how AI operates across a spectrum of autonomy, from fully independent systems to those that support human decision making.

While there have been plenty of successful deployments of AI systems, there have also been challenges. For example, an error in the AI for Britain's Universal Credit program caused underpayments for individuals paid multiple times a month, a common situation for lower-wage earners. This oversight placed affected recipients at financial risk, highlighting the need for rigorous testing and oversight in AI systems handling sensitive data. Given such challenges, how can government generate and maintain public trust when it comes to the design, development, and deployment of AI systems?

The Problem of Trust

Trust is a multidimensional concept that can be broken down into three components—ability, integrity, and benevolence. See Table. Trust Elements

Trust Elements

Trust Element Definition	Example in Government	Example in AI
Ability—Belief in the competency of the trust target	Belief that government can provide national security	Belief that the AI can correctly and consistently give the correct answer
Integrity—Belief in trustee's ability to adhere to a set of ethical principles	Belief that government will treat all people equally regardless of their gender or ethnicity	Belief that the AI will mirror society's view of ethical principles
Benevolence—Belief that the trustee wants to do good to the trustor	Belief that government will act in the best interests of the citizen	Belief that the AI has good intentions (or not negative intentions) in its functioning and outcomes

Trust in government in general has seen as steady decline over the last few years, including in Australia. According to the 2022 Edelman Trust Barometer, only 52 percent of Australians trust government to do the right thing (down 9 points from the previous year. Interestingly, 55 percent of Australians say that their default tendency is to distrust something until they see evidence that it is trustworthy. Factors attributable to the declining trust mirror trends around the world, and include decreasing interpersonal trust, perceptions of corruption, and deeply seated economic worries stemming from COVID-era policies. However, few of these distrust factors appear to directly involve the design or use of information systems, including AI systems.

Trust in AI, particularly in government applications, depends on whether citizens believe the government has the integrity and capability to deploy AI that serves the public interest. Examples illustrate successful AI implementations that foster trust:

1. **Dubai Electricity and Water Authority (DEWA):** The chatbot "Rammas," which responds to residents' queries in English and Arabic, reduced physical visits by 80 percent, showing effectiveness and responsiveness.
2. **Australian Tax Office:** This AI-enabled tax filing tool helps users verify work-related expense claims, ensuring transparency and accuracy.

Despite declining trust in government globally, studies show people regularly use AI in their private lives, with global AI adoption rising to 35 percent, according to IBM. Noteworthy private-sector AI uses include:

- Zzapp Malaria: Using AI to identify malaria risk areas to prevent outbreaks
- Vistra (U.S. power producer): An AI-powered tool improved monitoring of power plant indicators, optimizing efficiency and reducing emissions
- Wayfair: AI-supported logistics changes reduced inbound costs during the pandemic

These examples highlight how AI can boost efficiency and trustworthiness, especially when governments demonstrate responsible and beneficial AI use, potentially reversing trust decline trends.

Identifying Major Themes of AI in Government—Findings from the Workshop

The decline in trust in government directly impacts how much citizens are willing to trust it in the implementation of any powerful new technology. How can public sector leaders create trust in AI, given declining trust overall in government? To address this challenge, the IBM Center for The Business of Government hosted a forum of senior Australian government officials. This meeting provided a first-hand perspective from these officials on the status of AI, issues associated with AI, and the roadblocks and accelerators to implementing, culminating in the identification of five major themes of AI in government:

Theme 1—Government is in the business of providing services, and AI is simply a tool to facilitate that. Government should remain focused on providing government services, and not get "techno dazzled" by AI.

Theme 2—Government is held to a higher standard of performance regarding AI versus private companies, making explainability and transparency of utmost importance. Citizens expect government to get things right, and the services facilitated by AI should be sufficiently transparent and fully explained to the citizen.

Theme 3—Government needs to work holistically in terms of defining AI standard practices, operating models, etc. There is too much work and too many risks in implementing AI for standards development to happen only at the departmental level. Rather, this work needs to be coordinated at the highest level of government.

Theme 4—Adequate governance is necessary not only for AI technology, but also for the people who build AI systems and the processes used to build them. Issues emerge not only from the technology itself but also from the people and processes that implement AI.

Theme 5—There is a need to distinguish between different types of AI (fully autonomous, semiautonomous, and augmented) in establishing guidelines and approaches. Not all AI is the same, and costs, benefits, and risks differ for each type of AI. Discussing AI at a more granular can ensure optimal uses.

Recommendations Recommendations for Building Trusted AI in the Public Sector

These themes, coupled with background work done by the authors, gave rise to a list of recommendations to support building trusted AI in the public sector:

Recommendation 1—Promote AI-human collaboration when appropriate. Different kinds of AI call for different levels of human involvement, and citizens are generally more comfortable with a human being involved in providing direct services.

Recommendation 2—Focus on justifiability. Justifiability can be thought of as an outwards-facing business case, and with citizens as a primary audience. The government needs to article why an AI system needs to be developed, the amount of human involvement, and execution strategies.

Recommendation 3—Insist on explainability. Government must be able to explain why the AI came to a proposed decision, including the data that was used for the decision. This becomes increasingly important with decision making for high-stakes outcomes.

Recommendation 4—Build in contestability. Just as citizens can appeal to a person in government about the fairness of a decision, they also need to be able to contest the decisions made with AI. This feedback loop helps ensure that decisions are reasonable and not prone to bias.

Recommendation 5—Build in safety. While AI is deployed, risks can arise that make a safety feedback loop important. Government needs to either create or join an incidents-tracking database to capture and act upon feedback.

Recommendation 6—Ensure stability. The machine learning function in AI means that supporting algorithms will be constantly tweaked in response to new information. Not only does the AI system need auditing prior to implementation; regular examinations will ensure that AI provides stable results.

The emergence of AI in the world, and specifically in the public sector, makes this an exciting era. Given the frantic pace of AI development, government has a responsibility to be more proactive around the design, development, and deployment of AI systems to advance national goals. Government leaders can act now to implement fundamental recommendations to ensure successful AI delivery and that that adequate guardrails are in place to protect their citizens.