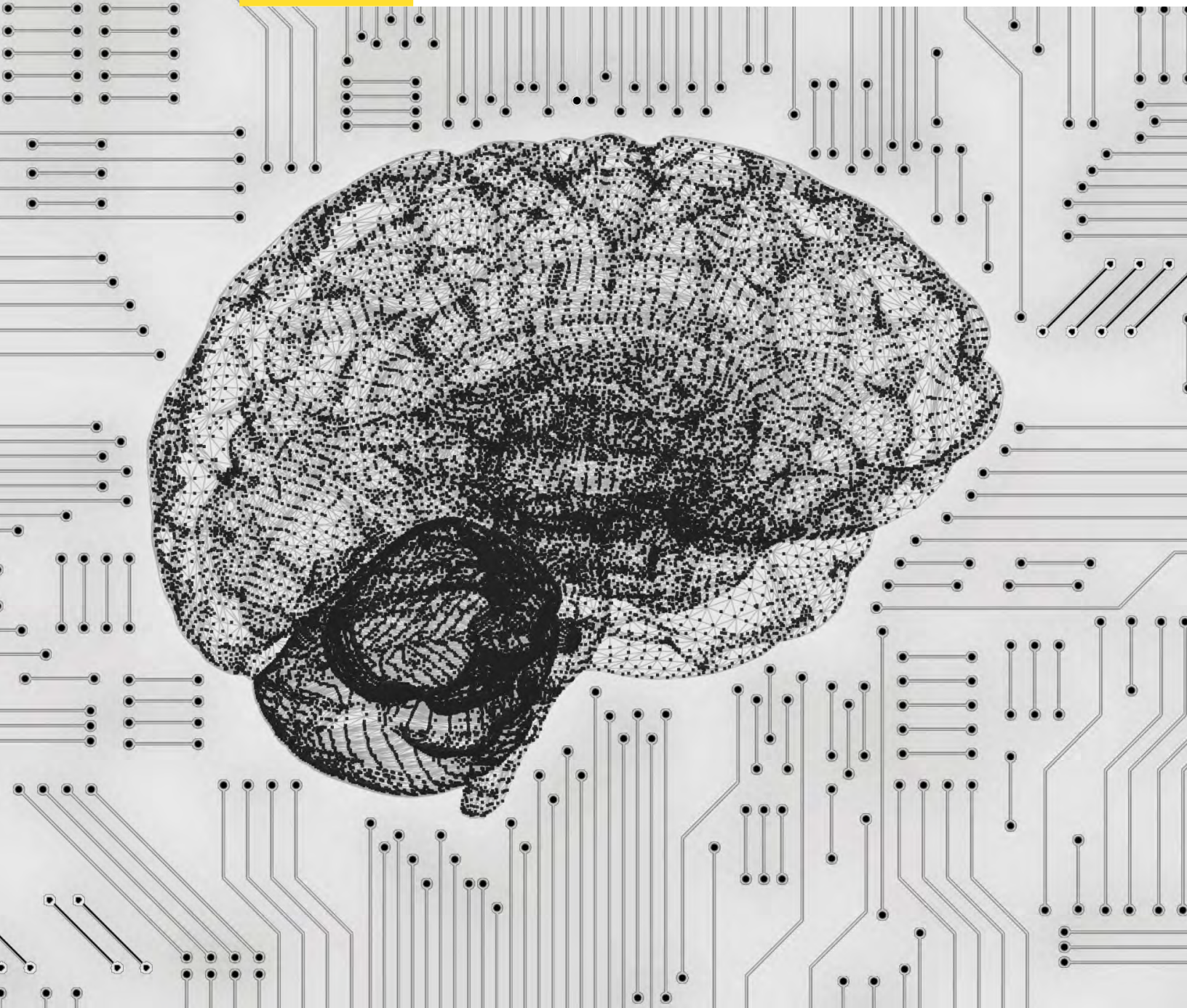
GOVERNING AI:
Navigating Risks, Rewards and Uncertainty



To encourage innovation in artificial intelligence while minimizing risks, Canada should adopt an incremental risk management approach to AI governance, supported by two new advisory institutions. By Daniel Munro

SUMMARY AND RECOMMENDATIONS

Artificial intelligence—the ability of machines to perform intelligent tasks such as sorting, analyzing, predicting and learning—promises substantial benefits for Canadians. Businesses that develop and commercialize AI have the potential to grow and create jobs, while organizations that adopt AI technologies can improve operations, enhance productivity and generate health, social and economic benefits for all.

Yet, some AI applications pose risks for individuals and communities:

- **AI-enabled automation threatens to disrupt labour markets and employment**
- **predictive analytics in finance, education, policing and other sectors can reinforce racial, gender and class biases**
- **data used in AI development and applications are often collected in ways that violate privacy and consent (see, for example, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, *Twitter and Tear Gas: The Power and Fragility of Networked Protest*, and *Data Governance in the Digital Age*)**

AI policy makers face a tension. They must establish conditions that allow AI to thrive and deliver benefits, while recognizing and responding to the harm that some AI applications can generate or reinforce. Options for addressing the tension range from a laissez-faire approach that would allow AI to develop and diffuse without limit, to a precautionary approach that would restrain the development of AI until risks are better understood and capacity to manage them is in place. Given that AI is a platform technology with many possible applications—and thus various risk profiles—it should be governed with an incremental risk-management approach that is case- and context-sensitive, rather than a blunt laissez-faire or precautionary approach. A risk-management approach allows space for AI technologies and applications to develop while monitoring and managing risks as they emerge in specific applications. To institutionalize a risk-management approach to governing AI in Canada we recommend that the Government of Canada create two new institutions:

- **an AI risk governance council**
- **an algorithm impact assessment agency**

At the heart of the AI policy challenge is a need to strike the right balance between supporting the development and diffusion of AI technologies that promise social, economic and other benefits for Canadians, and ensuring that risks to the rights and well-being of Canadians are addressed. This is not an easy task.

Because AI is an emerging technology, the exact nature and extent of potential benefits and risks are highly uncertain. Some observers favour a laissez-faire approach that places few limits on AI research and applications in order to accelerate discovery and access to benefits. The benefits may be economic—such as growth and job creation by Canadian firms that develop and commercialize AI technologies—as well as social, financial, political and health-oriented.

Image recognition and predictive analysis, for example, improve diagnosis of eye and cardiovascular diseases, breast cancer and melanoma. Predictive analysis is also being used to reduce work-

place accident risks, identify children who might be at risk of violence, predict individuals' risk of hospital readmission, and assess the credit worthiness of individuals who lack conventional credit histories, among other applications. Advances in natural language processing support increasingly common applications such as voice-activated assistants, automated customer support, translation, spam filters and interactive dialogue.

Other observers favour a precautionary approach that would limit the development and use of AI until more is known about the risks and how they can be managed. Early advances and uses of AI have already revealed serious challenges and risks, including:

1. Bias

Algorithms, and the data that feed them, have the potential to reinforce existing racial, gender, class and other biases and inequalities. The use of biased data in predictive policing models, for example, can bring additional police scrutiny to neighbourhoods with higher concentrations of minority residents—not because more crime is likely to occur, but because datasets on which the predictive models depend include more crime reports for those neighbourhoods due to past over-policing.

2. Safety

Numerous AI applications pose safety risks, ranging from algorithm-based models in the financial sector that malfunction and generate catastrophic financial losses to the development of AI-enabled lethal autonomous weapons that lack meaningful human control.

3. Privacy and consent

Access to massive datasets to support machine learning and improve analytical and decision-making capacity is essential to AI research and innovation. Yet, data is sometimes collected and used without explicit and meaningful consent from people from whom it is obtained, and often by violating privacy rights and expectations (see, for example, "At least two malls are using facial recognition technology to track shoppers' ages and genders without telling" and "Big other: surveillance capitalism and the prospects of an information civilization").

4. Explainability and accountability

There are unanswered questions about the extent to which decisions and actions must be explained to those affected by AI, and about who or what is ultimately accountable for AI-enabled decisions and actions. A key challenge is that although people expect organizations to offer explanations for the decisions that affect them—such as being denied a loan or government benefit, or receiving a fair criminal sentence—more advanced machine learning AI systems produce results based on analysis too complex for human beings to follow. For example, some systems will collect and analyze internet browsing history from loan applicants and assign a credit score (see *Weapons of Math Destruction*, 143-5) based on the extent to which their browsing history matches those of previous loan defaulters. But exactly what is problematic in the browsing history—and how it links to other data and patterns—can move beyond simple explanation. This will present challenges, especially for public sector innovation, given that justifications for AI decision-making will likely be required as a matter of political legitimacy.

A precautionary approach would tread cautiously in the face of these risks, but could also delay discovery and access to social, economic and other benefits for Canadians. How should AI governance

Although people expect organizations to offer explanations for the decisions that affect them—such as being denied a loan, a government benefit or a fair criminal sentence—more advanced AI systems produce results based on analysis too complex for human beings to follow.

proceed in light of this tension between innovation and risk?

Canada needs policies on AI ethics and governance

Canada's current approach to AI governance favours innovation over risk management. As such, it is ill-equipped to address the emerging risks associated with certain AI applications.

The federal Pan-Canadian Artificial Intelligence Strategy has little to say about AI ethics and governance and, until recently, there was little evidence that the Ministry of Innovation, Science and Economic Development (ISED)—or any other federal agency—has been thinking about a more comprehensive approach to identifying and managing the ethical, social and political risks and implications of AI. The pan-Canadian AI strategy calls for the development of "thought leadership on the economic, ethical, policy and legal implications of advances in artificial intelligence" and supports academic researchers exploring these issues through the AI & Society Program. But the "expected results" of the strategy include no mention of AI ethics and governance, focusing instead on Canada's international profile on AI research and training, developing and attracting AI talent, and enhancing innovation for socio-economic benefit.

When asked how AI will be regulated and governed, ISED says only that AI development and use must be consistent with the existing "marketplace framework," the Canadian Charter of Rights of Freedoms and the Personal Information Protection and Electronic Documents Act. Treasury Board Secretariat is leading consultations on responsible use of AI within the public sector, Global Affairs Canada coordinated a multi-university student symposium on AI and human rights issues, and some analysts within the federal government are working on approaches for algorithm bias and impact assessment. Additionally, ISED has launched National Digital and Data Consultations, which should address some of the data collection and use issues. But there is little evidence that an explicit, comprehensive federal strategy for AI ethics and governance is being developed or considered.

Some indication that Canada will pay more attention to AI ethics and governance emerged during a December 2018 meeting of G7 nations to discuss the impacts of artificial intelligence. Canada and France announced that they are seeking to create

an International Panel on Artificial Intelligence with a mission to "support and guide the responsible adoption of AI that is human-centric and grounded in human rights, inclusion, diversity, innovation and economic growth." The panel aims to engage stakeholders in science, industry, civil society, governments and international organizations on issues such as data collection and privacy; trust in AI; the future of work; responsible AI and human rights; and equity, responsibility and public good. While an important sign that AI ethics and governance are on the Canadian agenda, it is not clear what tangible effect the panel's work will have on AI governance in Canada.

OPTIONS FOR AI RISK MANAGEMENT

Canada has a government strategy to support AI research and innovation, and some provinces are making substantial investments in AI research, but it lacks strategies and institutional arrangements to identify, monitor and mitigate AI risks.

Following a case- and context-sensitive risk-management approach to governing AI, what principles and policy options might fill the gap?

The principles

To manage the tension between supporting innovation and addressing risks, Canada's approach to AI governance should do the following:

- Follow a policy on the responsible development and use of AI that prioritizes fairness, equality, safety, economic and political security, and the health and well-being of all people.

- Focus specific risk-management and regulatory actions on AI applications, not AI in general. AI risks will manifest only in the context of concrete applications and uses in specific activities and sectors, such as health diagnosis, loan assessments, predictive policing or benefits eligibility assessment. Risk assessment and management should focus on what is appropriate in those contexts.

The policies

With respect to specific policies and regulations, Canada's governments should consider the following:

- Develop and adopt a declaration on the responsible development and use of AI that

Canada's current approach to AI governance favours innovation over risk management. As such, it is ill-equipped to address the emerging risks associated with certain AI applications. A better balance is needed.

would signal to private sector developers and adopters, and public sector decision makers and civil servants, the importance of prioritizing fairness, safety, security, health and other values, principles and interests in the development and use of AI. The declaration could build on the Montreal Declaration for the Responsible Development of Artificial Intelligence.

- Develop a more comprehensive AI strategy that provides explicit guidance and funding to explore and manage the ethical, economic, legal and social dimensions of AI that are largely neglected in the current innovation-focused pan-Canadian strategy. This would bring Canada more in line with other countries working to address both the innovation and ethical dimensions of AI in their national contexts. Insights can be drawn from France's *For a Meaningful Artificial Intelligence: Towards a French and European Strategy*, Sweden's *National Approach to Artificial Intelligence*, and the U.K.'s *AI in the U.K.: Ready, willing and able?* among others.

- Require algorithm impact assessments to be completed before AI is used in sensitive areas such as healthcare, education, public safety and government benefits delivery. These would be similar to health technology assessments and environmental impact assessments but would focus on AI risks and benefits for individuals and communities, as well as the distribution of risks and benefits across demographic groups.

- Consider establishing a right to an explanation when an AI-based system produces decisions that have a significant effect on individual's financial, legal or other substantial interests. Discussion about whether and how to establish the right should be guided by the European Union's General Data Protection Regulation, which (arguably) establishes such a right. Whether such a right should exist—and whether it is technically feasible for explanations to be offered—will require public discussion. At a minimum, AI users in the private and public sectors should be alerted that they will be held accountable for outcomes that affect individuals' rights and interests.

Institutional arrangements

To realize these principles, policies and assessment activities—and to provide mechanisms for ongoing discussion about and risk management of AI—certain institutional arrangements should be established. Canada, the provinces and territories should consider creating the following:

1 **A dedicated artificial intelligence risk governance council.** This should be composed of people with technical, legal and ethical expertise to discuss, assess, report on, and provide advice to government and industry about AI innovation and risk management. Specifically, the council should:

   - lead the drafting of a declaration on the responsible development and use of AI, and a more comprehensive strategy for AI governance

   - monitor and report on trends in AI research and application, and conduct regular risk assessments of new, emerging and proliferating applications

   - provide advice to government and industry

on how to manage risks, drawing on risk-assessment results and best practices in other jurisdictions

- serve as a coordinating body for Canadian and international discussions about AI risk across sectors (e.g. health, education, innovation, economic development, law) and levels of government (federal, provincial, territorial and municipal)

The council could be created as a permanent, stand-alone arm of the federal government's existing arms-length science assessment body, the Council of Canadian Academies (CCA), and thereby benefit from the CCA's existing operational capacity and strength in convening experts from academia, industry and not-for-profit organizations. The council should support the work of the International panel on Artificial Intelligence, while also drawing from the panel's insights to articulate principles and promote practices appropriate to the Canadian context.

2 **Create an algorithm impact assessment agency.** This should be composed of technical, legal and ethics experts to conduct assessments deemed necessary or desirable by federal, provincial and territorial ministries and agencies, and to ensure that AI and algorithm applications respect the rights, interests and well-being of Canadians.

TOWARDS AI INNOVATION AND GOOD GOVERNANCE

Canada has an opportunity to be a global leader in AI research and innovation, and in effective AI governance. But while generating health, economic and social benefits from AI is already a priority among Canada's governments, managing the potential health, legal, economic and ethical risks of AI applications has largely taken a back seat. Experience with other emerging technologies should have taught us that prudent risk management is a precondition both for identifying and minimizing harms and, in turn, for generating sufficient public confidence to allow innovation to proceed. Time will tell if those lessons will be applied to AI governance or whether we face a future of unregulated AI risk and stalled AI innovation.

Daniel Munro is a Visiting Scholar and Director of Policy Projects at the Innovation Policy Lab in the Munk School of Global Affairs and Public Policy at the University of Toronto. His research interests include science, technology and innovation policy and applied ethics, including the ethics of new and emerging technologies.

For helpful suggestions and insightful conversations, the author would like to thank Tim Dutton, Sylvia Kingsmill, Maya Medeiros, Aaron Reynolds, and Mark Sutcliffe.


PUBLIC
POLICY
FORUM