Governing AI:
A Plan for Canada

Five high-impact actions needed in 2024/2025

White Paper by

AIGS GSIA

The evolving AI risk landscape

The world is in the midst of a growing AI revolution, with accelerating capabilities leading to accelerating benefits and risks. The 2010s saw the arrival of single-purpose algorithms used for everything from facial recognition to employment decisions to lethal autonomous weapons. Unfortunately, many of these algorithms were poorly designed or exhibited the biases in the human data they were trained on, leading some people to be unfairly denied jobs, mortgages and bail. The 2020s ignited public consciousness of AI with large models such as ChatGPT and Midjourney capable of writing intelligent text and generating high quality images for the first time. With it, cybercriminals have gained new tools, creative industries are being disrupted, and deepfakes and mass misinformation are putting public discourse and democracy at risk. AI is also exacerbating pre-existing concerns about digital privacy, wealth concentration in tech firms and nations, and a digital divide leaving many behind.

While these first two waves have been disruptive, there is reason to believe that even bigger ones are on the horizon. AI capabilities are on track to outperform humans at all tasks, including strategy, resource acquisition, human interaction, scientific discoveries, and boosting their own intelligence. Human-level AI is commonly called Artificial General Intelligence (AGI) and brings with it the prospect of automated technology development, mass job losses, and novel risks including global catastrophes from weaponisation or loss of control. Experts do not agree on how soon AGI will be built, with estimates ranging from beyond 2060 to as soon as 2025. However, its economic potential and near-term feasibility have many of the leading tech companies – including OpenAI, Google DeepMind, Meta and Salesforce – now making AGI their stated goal. Chart 1 and Table 1 summarise this discussion on AI capabilities and corresponding impacts over time, along with the key uncertainties moving forward:
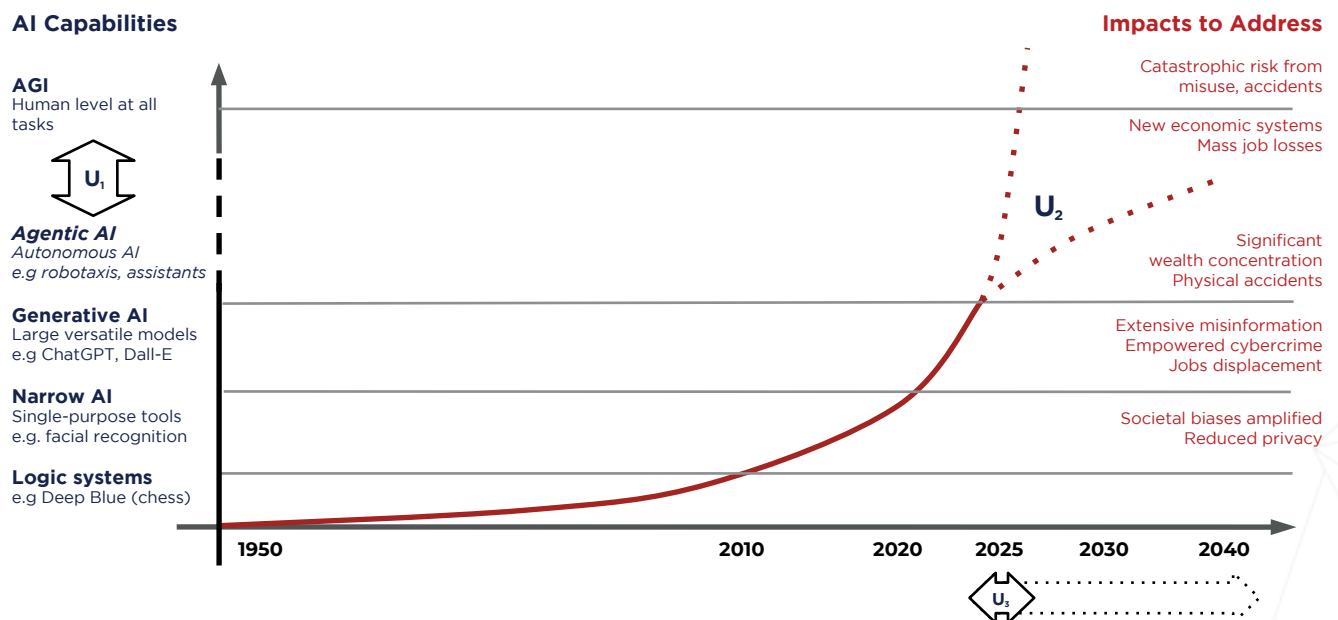

AI Capabilities

AGI
Human level at all tasks

U_1

Agentic AI
Autonomous AI
e.g robotaxis, assistants

Generative AI
Large versatile models
e.g ChatGPT, Dall-E

Narrow AI
Single-purpose tools
e.g. facial recognition

Logic systems
e.g Deep Blue (chess)

1950   2010   2020   2025   2030   2040

U_2

Impacts to Address

Catastrophic risk from misuse, accidents

New economic systems
Mass job losses

Significant wealth concentration
Physical accidents

Extensive misinformation
Empowered cybercrime
Jobs displacement

Societal biases amplified
Reduced privacy

U_3

Chart 1: High-level sketch of AI capabilities and impacts over time

Key areas of uncertainty are U_1: how many technical breakthroughs are needed to reach AGI, and U_2: whether the current exponential rate of progress will continue. This determines U_3: how much time governments have to prepare for the potential catastrophic risks and major social upheaval.

Category of AI	Narrow AI Single purpose machine learning	Generative AI General-purpose AI systems Content generation models	Towards AGI and beyond Up to and beyond human level
Timeline	Since ~2010	Since ~2020	2025? 2040? (unknown)
Key capabilities	Facial recognition Natural language processing Speech recognition Recommendations	*Narrow AI capabilities, plus:* Text, audio, image generation Information synthesis Code development Human persuasion	*Generative AI capabilities, plus:* Advanced situational awareness Long-term planning and execution Resource acquisition Reasoning and common sense Autonomous self improvement
Main impacts to address	Algorithmic bias Surveillance Transparency Info echo-chambers Carbon footprint Lethal autonomous weapons	*Narrow AI capabilities, plus:* Disruption of creative industries Accelerated job displacement Extensive misinformation Deepfake pornography Digital divide / inequality Accelerated race dynamics	*Generative AI capabilities, plus:* Mass job loss Extreme wealth concentration Social disorder from rapid change Deceptive systems Destabilised military dynamics Major accidents causing catastrophe

Table 1: Detail of AI capabilities and impacts per category

Columns detail the three broad categories of AI listed on Chart 1, which represent the recent, current and potential future capabilities, and the main corresponding impacts to address.

The response so far

Globally, measures to address AI impacts range from incomplete to nonexistent. Bias concerns are perhaps the best explored, with extensive research performed, new laws proposed, and a cottage industry of model evaluation solutions, but work remains. Copyright and liability issues are partly covered by existing laws, but much is still being decided. AI-enabled cybercrime is a rapidly growing phenomenon exacerbating existing vulnerabilities in digital infrastructure, and is only partly offset by AI-enabled cybersecurity. In regards to misinformation and deepfakes, a number of approaches such as bans and watermarking have been suggested, but the technical solutions and oversight mechanisms are nascent at best. Job displacement by Generative AI is still playing out, and if the speed and scale of disruption continue to accelerate, traditional retraining programs will likely be insufficient. And most concerning of all is the severe lack of foresight and planning by government and technologists on the economic, political, and public safety implications of AGI, and lack of public awareness and participation in elaborating desirable AGI-enabled futures.

Notable AI governance efforts include the European Union's comprehensive AI Act, tackling a range of impacts from high-risk use cases to general-purpose model monitoring. The United States issued an Executive Order outlining guidelines on frontier model safety, and is now exploring bipartisan congressional action. For its part, the UK launched the world's premier AI Safety Institute and inspired the IPCC-style scientific reports on the safety of advanced AI.

Canada's AI governance efforts started with its 2017 Pan-Canadian AI Strategy (PCAIS), 2019 Directive on Automated Decision-Making, 2020 co-founding of the Global Partnership on AI, and introduction of Bill C-27 AI & Data Act (AIDA) in June 2022. In 2023, Ottawa rolled out a guide for government use and voluntary code of practice for generative AI, and signed the Bletchley Declaration. In 2024 it announced $2.4B in budget investments to provide Canadian organisations access to AI compute and create an AI Safety Institute (AISI), which it has pledged to coordinate with partner institutions globally.

While Canada's efforts are laudable, much remains to be done to address current risks and prepare the country for the next waves of AI (Tables 1 & 2). The government's capacity and structures are insufficient to address the growing complexity of the AI risk landscape. The AIDA is still in committee and contains key limitations which will need to be resolved. The AISI's $50M in funding is insufficient given the scale of AI safety needs, and while Canada has approved key resolutions at the G7, UN, and on military use, in recent years it has mostly been a follower. Meanwhile, public understanding of AI's benefits and risks is limited, and public trust in AI is among the lowest in the world. There is therefore an urgent need for a clear vision of what success looks like and a robust set of actions to attain it.

Considerations for AI governance

Governments face a number of major challenges in addressing AI. First, the pace of AI development since 2022 has been remarkable, with traditional government processes unable to react in time to effectively guide technologists and protect citizens. Rapid advances also mean that even when laws are proposed in a timely manner, they can quickly become obsolete. Second, the high uncertainty about future AI capabilities makes it difficult to gauge what appropriate safeguards should look like. This is particularly the case for potential catastrophic risk and mass job impacts from smarter than human AI, given the massive response that would be needed to address such a scale of disruption. Third, with AI on track to play an increasing role in the economy and the military, both AI labs and nations are in an open race to be first to deploy cutting edge models, which puts ethics and safety at risk. Finally, AI is a global phenomenon, and any nation that single-handedly tries to regulate AI could stifle innovation at home and push labs to other jurisdictions, who in turn might gain a decisive advantage. To address these challenges, governments have a range of options (Table 2), the most important first step being to ensure the in-house capacity and accountability to monitor, respond to, and prepare for developments in the field.

The final consideration needs to be Canada's strategic positioning - what are we uniquely well suited to accomplish? We may not have the military or economic might of the US or EU, or host the largest AI companies, but we do have exceptional AI talent, a relatively stable and prosperous society, and a good reputation on the world stage. The world needs AI governance champions, and now is our time to leverage our strengths and deliver a robust approach to governing AI. Doing so will not only better protect Canadians, it will build Canada's leadership on what will likely become the defining issue of the 21st century.

Speed of AI development ➡	• Increase government capacity • Make legislation adaptable • Plan ahead for future AI impacts
Uncertainty on future risks ➡	• Invest in research to clarify • Adopt emergency preparedness
Economic race dynamics ➡	• Regulate and audit safety & ethics • Sync laws with other jurisdictions
Global scope ➡	• Push for global treaties, standards
Risk of stifling AI's benefit ➡	• Avoid one-size-fits-all approaches • Provide incentives for beneficial AI

Table 2: Key challenges and actionable responses

A Plan for Canada: Five High-Impact Actions to Govern AI

Given the pace of developments in AI, and the scale of potential impacts, we recommend the following five actions for Canada to significantly advance AI governance **by end of Q2 2025:**

1 **Establish a <u>Ministry of AI</u>**

◆ Ensure a voice at Cabinet and awareness of AI's impact on other top portfolios

◆ Can house AI regulatory bodies, monitor and inform leadership about AI developments, communicate with the provinces, and coordinate federal action

Rationale: AI has become too big and complex to be handled by existing ministries. It needs a direct voice at Cabinet to ensure a) its prioritisation and b) awareness of its impact on all the other top portfolios. A Ministry of AI is the best place to house monitoring and regulatory agencies and will avoid ISED's conflicting mandate of boosting development. It also clarifies who is responsible for the file, and can grow to ensure government has the capacity it needs to limit risks and harness opportunities.

2 **Champion and fund global AI governance efforts:**

◆ Support global efforts addressing frontier AI race dynamics

◆ Host talks on global AI governance during Canada's 2025 G7 leadership

Rationale: Without global coordination, domestic regulation will simply push AI development to other jurisdictions. Moreover, frontier AI development is mostly happening abroad, meaning that reducing systemic risk requires influencing the actions of other countries. Progress on global talks will require strong national champions, and Canada's hosting of the G7 in 2025 is a unique opportunity for leadership and positive impact.

3 **Improve and pass the AI & Data Act and supporting regulations**

◆ Provide a legal framework to address range of ethics and safety concerns

◆ Improve independence by moving administration to the AI ministry

Rationale: Both industry and the public need a legal framework that can provide clarity on what constitutes acceptable AI development and use. With current harms already being felt and timelines on upcoming capabilities uncertain, there is no time to drop the AIDA and reintroduce it later. Finally, to avoid conflict with ISED's mandate, the AIDC should be located in the new Ministry of AI. <u>Detailed recommendations available at AIGS.ca</u>.

4 **Increase and focus <u>Budget 2024</u> investments on safety & governance**

◆ Dedicate >$500M for research on technical AI and governance policy research, coordinated by Canada's new AI Safety Institute

◆ Allocate >20% of the proposed AI Compute Access Fund for research on frontier AI safet

Rationale: The world does not currently have the technical solutions nor the effective policy mechanisms to ensure safe and ethical AI. With hundreds of billions invested every year to boost capabilities, and timelines on the most dangerous forms of AI highly uncertain, government must make a major play to ensure that governance and safety research rapidly catch up. The $50 million for Canada's new AI Safety Institute needs to be significantly increased and the compute budget needs to be targeted towards safety research.

5 **Launch a national conversation on AI:**

◆ Nationwide open public consultations to inform policy decisions on AI in relation to the jobs transition, wealth concentration, and the right balance between pursuing rewards and limiting risks

Rationale: Even if the safety risks are managed, AI will radically transform society, potentially creating a post-jobs, post-scarcity world in which human intelligence and institutions take the backseat. Few are ready for this amount of change. Managing this transition and ensuring it reflects the will of the people will take time and needs to start early. With its relative stability, educated population, and strong AI ecosystem, Canada is in a uniquely good position to pilot such a conversation.