Governing AI:
A Plan for Canada

Five high-impact actions needed in 2023/2024

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 brought large models such as GPT-3 (later ChatGPT) and Midjourney, capable of writing intelligent text and generating high quality images. 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 coming soon. 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 through poor design, malicious use or accident. Experts do not agree on how soon AGI will be built, with estimates ranging from beyond 2060 to as soon as 2025. However, with the latest model capabilities far surpassing expectations, prediction markets have shifted dramatically towards the shorter timelines. Chart 1 and Table 1 summarise this discussion on AI capabilities and corresponding impacts over time, along with the key uncertainties moving forward:
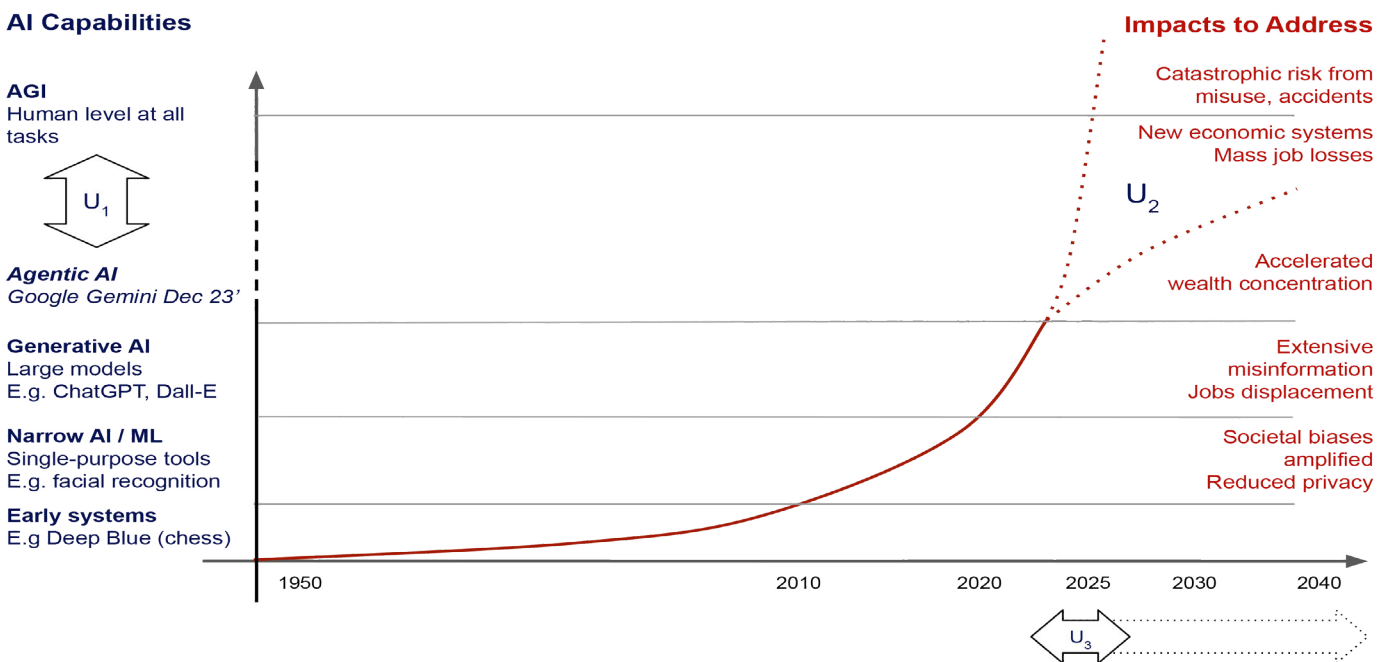

Chart 1: High-level sketch of AI capabilities and impacts over time

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 / ML Early machine learning (ML)	Large models General-purpose AI systems "Generative AI"	Towards AGI and beyond Human level and greater
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 Gene sequencing Code development	Large model capabilities, plus: Advanced situational awareness Long-term planning and execution Resource acquisition Social engineering Autonomous self improvement
Main impacts to address	Algorithmic bias Surveillance Transparency Info echo-chambers Carbon footprint Lethal autonomous weapons	Narrow AI concerns, plus: Disruption of creative industries Accelerated job displacement Offshore worker mistreatment Extensive misinformation Digital divide / inequality Accelerated race dynamics	Large model concerns, 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 key corresponding impacts to address

Considerations for AI governance

All governments face challenges in governing AI, such as the speed of its development, the difficulties of regulating non-physical products, and current economic and geopolitical incentives. 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. Any nation that single-handedly tries to regulate AI risks stifling innovation at home and pushing labs to other jurisdictions, who might gain a decisive advantage. Effective AI governance is therefore impossible without strong global collaboration.

Canada has already taken a number of initiatives with its 2017 Pan-Canadian AI Strategy (PCAIS), 2019 Directive on Automated Decision-Making, and introduction of Bill C-27 AI & Data Act in June 2022, Most recently for Generative AI, Canada created a guide for government use and a voluntary code of practice for industry.

While these efforts are laudable, most of the PCAIS investment has been to boost AI capabilities, the directives only apply to government, and the code of conduct is non-binding. If enacted, Bill C-27 would introduce some hard rules and liabilities, but it contains a number of limitations. These include, among others, the lack of clarity as to which AI systems it applies to, insufficient AI commissioner capacity and independence, and an enforcement date delayed until at least 2025. Of greatest concern, none of these efforts attempt to prepare the country for the upcoming risks from more capable AI systems (Table 1).

It is therefore urgent that a more robust approach be taken. Doing so will not only better protect Canadians, it will build Canada's leadership on what could become the defining issue of the 21st century.

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 2024:**

1 Establish a central AI agency for the Government of Canada

- Team based within the Privy Council Office
- Monitors and informs leadership about AI developments, communicates with the provinces, and coordinates federal action

Rationale: In a fast-moving environment, government needs in-house capacity to execute effectively. An expert team within a central agency will be able to ensure that the PMO and ministries have the information they need to limit risks and harness opportunities as they arise. ISED or other departments don't have the institutional incentives or mandate to navigate and balance between the full range of issues.

2 Invest >$500M in AI governance and safety research

- More than double the existing CIFAR Pan-Canadian AI Strategy funding
- Earmark the extra funds to developing better policy and technical AI safety solutions

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.

3 Champion and fund global AI governance efforts:

- Support global initiatives such as at the UN, G7, UK safety summit, and OECD
- Lead the conversation about CERN, IPCC, IAEA models for global AI governance

Rationale: Without global coordination, domestic regulation will simply push AI development to other jurisdictions. Moreover, it is likely that AGI will be built outside of Canada first, meaning that reducing catastrophic risk requires influencing the actions of other countries (esp. U.S. and China). Progress on global talks will require strong national champions, and this is Canada's opportunity to take a leading role.

4 Improve and pass the AI & Data Act

- Best mechanism for enacting policies on audits, transparency, and liability
- Establish a Canadian AI Commission to regulate high-risk models

Rationale: Both industry and the public need a legal framework that can provide clarity on what constitutes acceptable AI development and use. It is therefore urgent to accelerate the consultations and drafting of legislation. While an early election could scuttle Bill C-27, any progress made on the current AIDA will advance the conversation and be useful to future governments. Detailed recommendations available at AIGS.ca

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.

To learn more, visit AIGS.ca