AIGS GSIA

AI Governance & Safety Canada
Gouvernance et sécurité de l'IA Canada

November 3rd, 2023

Mr. Joël Lightbound, M.P.
Chair
Standing Committee on Industry and Technology
House of Commons
Parliament of Canada

Dear Mr. Lightbound,

Canada finds itself in the midst of a global AI revolution, and the scale and complexity of the risks are nothing less than astounding. The task your committee faces in amending the AI & Data Act to meet the needs of Canadians is therefore both essential and a very tall order.

AI Governance & Safety Canada is a cross-partisan not-for-profit and a community of people across the country working to build Canada's leadership in the governance and safety of artificial intelligence. We provided input to ISED's roundtables on the Voluntary Code of Practice for Generative AI and recently released our white paper *Governing AI: A Plan for Canada*.

The recommendations for the AIDA that we present in this brief are the result of months of work and extensive consultations with national and international stakeholders.

You will find that we recommend a significant rewrite of the Act. Our strategy here is to present what it would actually take to protect Canadians from the current and upcoming risks, and then to work with the committee on any compromises that must be made. If time were not an issue, we would recommend separating the AIDA from the Bill and reintroducing it after lengthy consultations and deliberations. However, with accelerating developments in AI, and existing harms already being felt, Canadians do not have that luxury. We need working legislation now.

We therefore urge committee members to not give up on the AI & Data Act, but to take the time to understand the full range of AI risks to be addressed, and prepare legislation that can serve Canadians well today and in the coming years.

We remain available for any assistance that you require.

Sincerely,

Wyatt Tessari L'Allié
Founder & Executive Director
AI Governance & Safety Canada
contact@aigs.ca

Recommendations for the AI & Data Act

Brief to the Standing Committee on Industry and Technology on Bill C-27

November 3rd, 2023

Table of contents

Letter to the Committee Chair 1
One-page summary 3
Part I: What does Canada need AI legislation for? 4
The evolving AI risk landscape 5
The role for legislation 6
Legislation in other jurisdictions 7
Existing Canadian legislation, and gaps requiring a dedicated AI law 8
Part II: Recommended amendments 9
Use four risk categories and keep the requirements proportionate 10
Unacceptable-risk systems (URS) 10
High-risk general-purpose AI systems (HRGPAIS) 13
High-risk single-purpose AI systems (HRSPAIS) 17
Moderate or low-risk systems 18
Fix critical gaps 18
Provide government the capacity it needs 20
Part III: Specific wording to change in the Bill 22
The wording will be provided as an addendum to this submission after the text of the government's amendments have been shared *22*

One-page summary

Canada is in the midst of a growing AI revolution and must find a way to harness its many benefits while navigating a rapidly evolving array of risks. These range from privacy and copyright violations to potential major job losses and catastrophic accidents or misuse. While new legislation will not be sufficient on its own to protect Canadians, it is nonetheless essential. Existing sectoral laws leave key gaps, especially with regards to general-purpose systems that have unpredictable and sometimes unacceptable capabilities. Moreover, government urgently needs the authority, agility, and capacity to govern such a complex and fast-moving technology. Finally, while harmonisation across jurisdictions is essential, the current EU AI Act contains some key flaws that Canada will need to avoid, and the U.S. directives are incomplete.

We therefore recommend the following changes to the AI and Data Act:

1) **Use 4 risk categories** with baseline definitions, and keep the requirements proportionate

Unacceptable-risk systems *Unsafe until proven otherwise* Ex: AI systems capable of designing weapons of mass destruction	Place a moratorium on these systems, to be lifted on a case-by-case basis if and when proven beyond a reasonable doubt that they can be safely developed and used.
High-risk general-purpose systems *Safe if strictly regulated* Ex: Chatbots capable of generating malware, talking users into suicide, or displacing millions of jobs	Minimise irreversible harm and ensure government is aware of new AI capabilities by setting up an accessible licensing regime, and requiring incident reporting, auditing, safety and cybersecurity standards, and pre-deployment public consultations.
High-risk single-purpose systems *Safe with important precautions* Ex: Algorithms used to make employment or judicial decisions	Apply the current high-impact system requirements to this category to protect Canadians from biassed algorithms, to fill gaps in sectoral regulations, and to ensure minimum standards.
Moderate or low-risk systems *Generally safe without regulation* Ex: Netflix recommendations	Exempt from the AI & Data Act by default. This would eliminate red tape for the vast majority of AI systems used and developed in Canada today.

2) **Fix critical gaps:** The AIDA will not be able to protect Canadians if there are gaps for standalone models, open-source, political parties, and government and its contractors

3) **Provide government the capacity it needs:** establish a Canadian AI Safety & Ethics Commission (CAISEC) mandated with regulating high and unacceptable risk systems, supporting industry, civil society, and other ministries, and meeting international obligations

Part I: What does Canada need AI legislation for?

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:

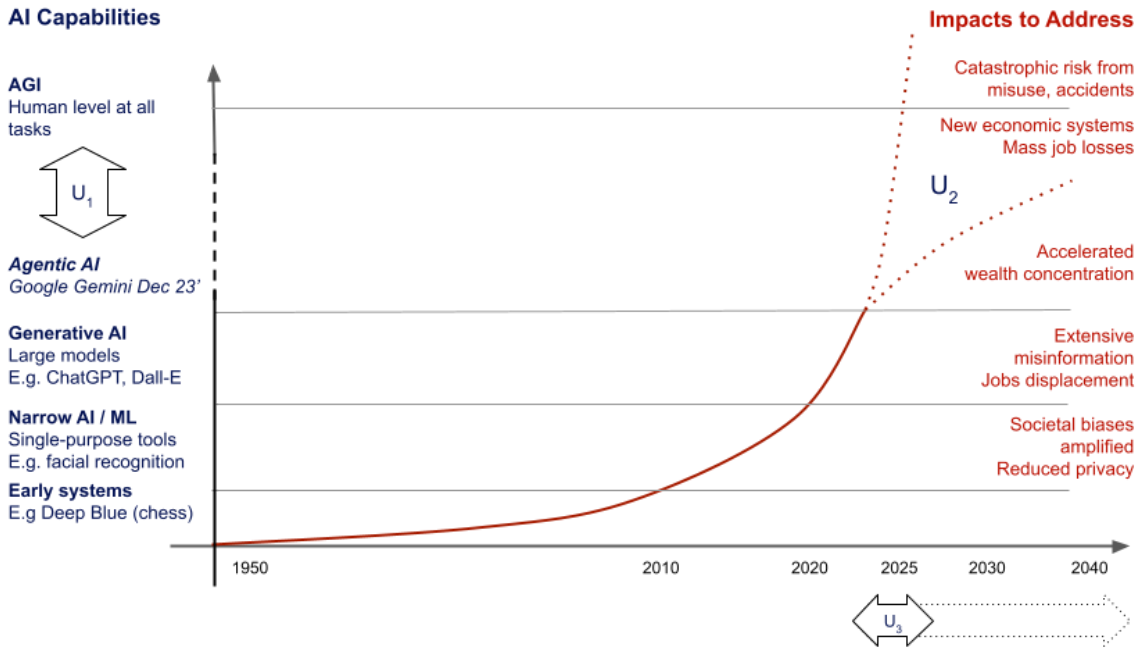

Chart 1: High-level sketch of AI capabilities and impacts over time

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** Up to and beyond human level
Timeline	*Since ~2010*	*Since ~2020*	*2025? 2035? (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 Persuasion and social engineering Autonomous self-improvement, self-exfiltration and self-replication
Main impacts to address	Algorithmic bias Surveillance Transparency Info echo-chambers Lethal autonomous weapons Privacy loss	Narrow AI concerns, plus: Disruption of creative industries Significant carbon footprint Accelerated job displacement Offshore worker mistreatment Extensive misinformation Digital divide / inequality Accelerated race dynamics	Large model concerns, plus: Extreme wealth and power concentration in AI labs Mass job loss / new economics Social disorder from rapid change Actively deceptive systems Destabilised military dynamics Major accidents causing catastrophe
The AIDA’s current scope	Originally designed to address this level of AI. Would govern many of these risks.	Minister Champagne’s recent proposed amendments could address some of these new GenAI risks.	The AIDA’s current requirements for high-impact systems will be unable to protect Canadians from these risks.

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. The last row explains which impacts the AI & Data Act would currently address.

The role for legislation

There is no silver bullet to navigating this vast array of impacts. AI is a global issue of which Canada is only one player, and as software it easily proliferates across jurisdictions. This means that even with the best laws, no nation can single-handedly guarantee its citizens safety from the harms. Legislation on its own will not be enough.

In our white paper Governing AI: A Plan for Canada, we outline the kinds of actions that the federal government will need to take in tandem. Investments in governance, safety and ethics research can provide better policy and technical solutions. Leadership on the world stage can advance international talks, treaties and collaboration on standards and enforcement. And piloting a national public conversation and consultations can help clarify what kind of futures we want to collectively build with AI.

Nonetheless, legislation is core to any effective AI governance strategy, and is the best tool to address the following issues:

- **Production and proliferation of AI with unacceptable capabilities**: making dangerous forms of AI illegal in Canada is the best available deterrent to their development and proliferation here.
- **Malicious use:** while existing public safety laws can be applied to AI, clarifying what constitutes unacceptable use of AI will provide Canadians and law enforcement much needed clarity and reduce its prevalence.
- **Legal accountability**: due to the autonomous nature of AI, it is often unclear who is responsible when something goes wrong. A law will be essential to clarify this.
- **Reckless deployment**: Even beneficial AI systems can cause harm if they are deployed without forethought or consultations with the people impacted.
- **Arms race dynamics**: AI labs are subject to strong economic incentives to be the first to deploy new capabilities, which often negatively affects the safety and ethical use of the systems built. Industry cannot solve this dilemma on its own. Only government can set and enforce rules that will counterbalance the economic benefits of cutting corners.
- **Lack of safety standards enforcement:** Many high quality global standards on AI safety have already been developed, and some of them applied. However, legislation is needed for them to be systematically adopted.

Legislation in other jurisdictions

We commend the Minister for seeking to harmonise Canada's legislation with the EU AI Act, as international harmonisation of AI will be essential to simplifying the burden on business and avoiding movements across jurisdictions to avoid regulation.

The challenge that the EU AI Act misses the mark in key respects. Much like the AIDA, it was introduced before general-purpose AI systems (GPAIS) and does not recognise or effectively manage their fundamentally different behaviour and risk profile. It is also heavily proscriptive and inflexible, attempting an exact list of all use-cases to be regulated, despite the fact that these are bound to quickly evolve. Moreover, as a 'horizontal' approach, it tries to regulate all sectors at once, instead of supporting each vertical (e.g. health, transport) to develop their own based on their unique expertise. Copying the EU AI Act would therefore fail Canadians in these same ways.

In the United States, the Blumenthal-Hawley framework in the Senate provides some important direction, especially around licensing of high-risk models, but is otherwise incomplete. The White House's recent Executive Order provides more detailed directives, but not all are relevant to the Canadian context. Moreover, some of the requirements (such as regulating general-purpose models only if their training requires 10^{26} computer operations or more) will no longer be relevant by the time the AIDA comes into force.

These are early days in the field, and the unfortunate reality for Canada and the AIDA is that there are no good existing laws to follow. While waiting for more countries to pass legislation first and choosing the best option may be tempting, doing so will delay protection to Canadians at a time when the harms are already being felt, and lose critical time to prepare for the upcoming risks of major accidents and social upheaval. It would also miss a golden

opportunity for Canada to lead on the world stage and help positively shape legislation in other countries. The best we can do is therefore to prepare a law that will be both robust and flexible in meeting the needs of Canadians, and act as a source of inspiration for other jurisdictions.

Existing Canadian legislation, and gaps requiring a dedicated AI law

As the AIDA Companion Document points out, existing legislation such as the *Canadian Human Rights Act, Motor Vehicle Safety Act* and *Bank Act* already contain the legal framework needed to govern many aspects of artificial intelligence. Adding new legislation in these areas would create a bureaucratic process where AI developers and deployers would need to comply with two separate and potentially conflicting sets of regulations.The better option to ensure consistent standards across these sectors is to empower ISED to support the other ministries with AI-specific expertise and ensure sectoral regulations are harmonised across government.

However, as was also pointed out in the Companion Document, there are systems that will fall between the gaps of sectoral regulations, and therefore require an AI law. Most notably, what is new in the field of AI, and what sectoral regulations are fundamentally unable to address, are general-purpose systems that can be used in multiple sectors at once.

Moreover, as AI systems become more capable, the scale of impacts and potential harms become such that a responsible government will need a law to ban some forms of it and strictly regulate others. Existing laws do not address these novel harms nor the unique technology enabling them.

Finally, government currently does not have the structure and authority to effectively govern the rapidly expanding impacts of AI. Legislation is therefore needed to ensure it has the capacity to administer and enforce the Act and protect Canadians from individual and collective harms.

Recap: What Canada needs a dedicated AI law for:

1) Protecting Canadians from AI systems with unacceptable capabilities
2) Regulating specific high-risk systems:
    a) General purpose systems
    b) Single-purpose systems that can't be adequately regulated by sectoral laws
3) Providing government the authority, agility and capacity to govern AI effectively

Part II: Recommended amendments

Use four risk categories and keep the requirements proportionate

Recommendation	*Rationale*
<u>Define 4 categories of AI</u> based on their risk profile : ● Unacceptable risk systems ● High-risk general-purpose systems ● High-risk single-purpose systems ● Moderate or Low-risk systems (exempt from the AIDA by default) *(optional) Allow the regulations to identify other categories and create requirements for them as the technology evolves*	This will allow most innovation to continue unimpeded while ensuring the AI systems that could cause the most serious harm have appropriate safeguards. We separate here high-risk *general-purpose* systems and high-risk *single purpose* systems, because the behaviours, risk profiles, and necessary regulatory measures are markedly different. *There is a case for allowing the regulators the flexibility to define new categories, although this could be abused and create uncertainty.* For every category, it is important to make the definitions 1) flexible enough to be future-proof, but 2) clear enough to avoid confusion or regulators getting lobbied into making the criteria meaningless. To do so, we recommend using **baseline definitions** in the Act, and allowing regulations to add further criteria at a later date if need be.
<u>Unacceptable-risk systems (URS)</u> <u>Define as</u> an AI system, or AI model capable of powering an AI system, that: ● Is capable of: ○ designing weapons of mass destruction (WMD), or itself providing or otherwise enabling WMD capabilities, or ○ autonomously directing lethal weapons and is not under the control of the Minister of	The purpose of this category is to separate out the systems with AI capabilities for which there are currently no reliable methods of mitigating the risks. They must therefore be at least temporarily banned until proven safe. Current AI systems are already capable of providing advice on how to build simple explosives and synthesise dangerous chemical compounds. Systems that can design or share nuclear, bioweapon or other WMD capabilities would put society at extreme risk and are fundamentally at odds with Canadian values. Since 2013, there have been growing concerns around <u>lethal autonomous weapons</u>, i.e. weapons that select their target and fire upon them without a human in the loop. This clause

National Defence, or	would ban civilian development and use, leaving the larger discussion of military use to arms control treaties and other legislation.
○ unprompted self-modification, or enabling recursive self-modification in other models or systems, or	Unprompted self-modification refers to systems that, when given an initial goal by the user (e.g. to develop a cure for a disease), are capable of reengineering themselves to gain new capabilities (e.g.changing their own model weights to increase their effective IQ or knowledge by multiple orders of magnitude) in order to better pursue that initial goal. The unplanned and unpredictable nature of such emergent capabilities, and the possibility of initial safety mechanisms failing as the system modifies itself, make it an unacceptable risk. Systems that enable iterative or recursive improvement of another system, such as a scaffolding, optimizer, or driver of the powerful model, are similarly dangerous.
○ autonomous self-exfiltration or self-replication, or	Autonomous self-exfiltration refers to systems that, when given an unrelated initial goal, are capable of copying their model weights and code onto servers outside the model owner's control, in order to pursue that initial goal. Similarly, self-replication is the ability of the system to make copies of itself (potentially thousands or millions of them) on servers or computers outside of the owner's control.
○ autonomous resource acquisition, or	Autonomous resource acquisition refers to AI systems being able to acquire, commandeer, grow, or control resources such as money and computation without explicit instruction from (and potentially unknown to) the human user or model owner.
○ active deception, or	Active deception refers to AI systems that, in pursuit of an initial goal, are capable of lying to or otherwise manipulating human beings into action or inaction that is harmful to themselves or society at large. Unintentionally providing false information when prompted by human users would fall under passive deception.

○ avoiding or preventing interventions to turn it off, or	AI systems given an initial goal are inherently incentivised to avoid any external action that might stop them from achieving that goal. If poorly designed or unsafely built, advanced AI systems will be able to recognise the situation they are in and actively block user or law enforcement attempts to halt their operation.
● Requires more than 10^{25} computer operations to train and develop, or is built on an artefact trained on more than that amount, or	10^{25} computer operations would mean any training run approximately bigger than the one used for creating OpenAI's GPT-4. The threshold should be placed here as it is impossible to reliably predict what capabilities larger systems will have. The recent White House EO's higher threshold of 10^{26} operations for regulating large models is in our view problematic for this reason. As algorithms become more efficient, computation demands for a specific capability will drop over time, so defining a threshold in the Act won't unfairly restrict future systems. Given the current rapid algorithmic progress, regulators will need to dynamically lower compute thresholds over time. When an average personal computer can train an unacceptable-risk system, this clause's relevance will expire. This could also provide an environmental benefit by incentivising lowering compute usage to avoid regulations. Total compute needs to be considered, because building or further training on top of an existing model retains and builds on its capabilities and associated risks.
● *(optional) Are used in applications that comprise subliminal techniques, exploitative systems or social scoring systems used by public authorities are strictly prohibited, or any real-time remote biometric identification systems used by law enforcement in publicly-accessible spaces, or*	*Optional: add wording to make the AIDA inclusive of systems defined as unacceptable risk in the EU AI Act (Title II / Article 5).*

• Meets other criteria to be defined in the regulations	Make the definitions of each category flexible enough to be future-proof, but clear enough to avoid confusion or regulations that miss the mark, by using baseline criteria in the Act, and allowing regulations to add further criteria if need be.
Place a moratorium on the possession, or the attempt to possess, unacceptable risk systems. This can be lifted by the CAISEC only when the safety and public benefit of such systems can be guaranteed beyond reasonable doubt.	The moratorium on unacceptably dangerous systems could be lifted on a case-by-case basis by the newly formed Canadian AI Safety & Ethics Commission if and when acceptable safety precautions and the system's public benefit are established.
Apply the Part 2 criminal charges to anyone who possesses or attempts to possess an unacceptable-risk system	The current criminal liability only refers to persons knowingly making the system available, and it resulting in harm. For unacceptable risk systems this is dangerously vague and effectively encourages private development of these systems. Given the financial incentives to be first to develop and deploy new capabilities, and the billions of dollars invested in the space, the existing fines are unlikely to dissuade bad actors. Prison time must be included for the law to be an effective deterrent.
High-risk general-purpose AI systems (HRGPAIS)	The purpose of this category is to minimise irreversible harm to society and incentivise understanding and safety of AI, without denying society the many benefits. Given the unpredictable capabilities of these systems, governments must "expect to be surprised" by them and therefore keep a close eye on their development, deployment, and usage.
Define HRGPAIS as AI systems, or AI models capable of powering AI systems, that are general-purpose in nature and: • Are capable of: ○ social engineering, passive deception, or interacting with a person in a way that makes the person think they are dealing with a human, or	The ability to intelligently interact with human beings is an essential feature of modern AI, and key to powering many positive applications. The flip side is that it makes human beings vulnerable to AI, such as becoming emotionally attached or dependent, or getting manipulated into harming themselves and others. At a collective level, democracy and effective public engagement is put at risk by systems used to generate compelling and personalised disinformation.

	Systems with these passive deception capabilities therefore need to be considered high-risk. **Note on usage by political parties**: HRGPAISs significantly change the dynamic of voter persuasion. Whereas in the past it would take thousands of human volunteers to make millions of phone calls, with GPT-4 powered systems connected to compelling voice generators like VALL-E and online data scraping about voters, it will be increasingly possible to automate millions of personalised, interactive, and persuasive phone call conversations without the person realising they are talking to a machine. This is one of the many reasons HRGPAIS need to be considered high-risk and political party usage needs to be included in the Act (see Fix critical gaps section below)
○ providing instructions or code enabling criminal activities, or	LLMs can already generate malware, automate phishing attacks, and provide effective instructions for carrying out physical crime.
○ achieving certain scores on industry-recognized performance benchmarks to be defined in the regulations, or	Benchmarks are tests to evaluate model capabilities. They are therefore very useful for gauging the models' risk profiles, but are still a rapidly evolving field in their own right. We therefore do not recommend identifying in the Act which benchmark to use, as regulators will need flexibility. For reference, MMLU is currently among the best available benchmarks and a system scoring 70% or above on it should be considered high-risk.
● Require more than 10^{24} computer operations to train and develop in total, or	At current algorithmic efficiency, this would mean training runs for systems roughly as big as ChatGPT 3.5 and above. These core systems are already enabling significant benefits and harms, and have yet to be pushed to their full capacity with innovative prompting, plug-ins, and code-wrapping (e.g. Auto-GPT).
● Meet other criteria to be defined in the regulations	

Create appropriate requirements for each stage of the HRGPAIS's lifecycle, including planning, training, pre-deployment, deployment and operation, or other stage to be defined in the regulations, and:	Some of the key stages to regulate are 1) the planning stage (i.e. ensuring basic precautions are in place before any training is begun), 2) the training runs (which need to be monitored for emergent unacceptable capabilities), 3) pre-deployment (to avoid causing avoidable harm to society) and 4) deployment and operation. However, given the evolving nature of the field, we recommend allowing the regulations to identify these stages and adjust them as needs arise. **Note on including model R&D in the Act:** With HRGPAISs, it is very difficult to reliably predict which capabilities and behaviours the model will have ahead of time. Some of the most dangerous ones, such as unprompted self-modification and preventing interventions to be shut off, can happen at the pre-training stage, long before the models are deployed. Moreover, **once a model is pre-trained it can very easily be hacked, shared or otherwise widely proliferated**. It is therefore essential that the R&D stage of AI be covered in the regulations, not just their distribution and use.
• Licensing,	Setting up a simple licensing regime (accessible to all via an online application form) is a low-burden way to keep regulators in the know about who is building high-risk systems and what their capabilities are. This is being suggested in the US Senate legislation.
• Impact assessments,	Impact assessments are a useful mechanism for encouraging awareness, forethought and communication with stakeholders.
• Incident reporting,	Incident reporting is another straightforward yet effective tool for increasing system safety, and has successfully been used for years by aviation authorities to keep aeroplanes safe.
• Auditing,	GPAIS are notorious for having unforeseen capabilities, meaning standard 'checklist' audits won't be enough. Auditors need to be incentivised to "try really hard" to break the systems before declaring them safe. Adversarial red-teaming is currently the best available approach to robust auditing.
• Cybersecurity,	Cybersecurity is an essential component of AI safety, because as soon as a model has been

	pre-trained its weights can be hacked by malicious actors and used for nefarious purposes. At present, it is likely that models created at Canadian AI labs are being targeted by criminals and geopolitical rivals. Labs must protect the HRGPAIS they develop and operate with high standards of cybersecurity.
• Safety requirements developed by globally recognized bodies,	Safety standards are a common and necessary component of any technology regulation, and regularly used in transportation and medical devices. We recommend letting the regulators choose the specific standards to follow, and harmonise them with globally recognised ones, as they will need to continuously evolve over time. These standards should include among others: watermarking of AI-generated content, auditing, and safe scaling policies.
• Public consultations, and	HRGPAIS will dramatically alter many Canadians' work, personal lives, culture, education, and daily activities. No other industry is allowed to significantly disrupt people's lives without a social mandate, and the AI industry should be no different. Basic public consultations need to be part of any pre-deployment checklist.
• Other requirements to be defined in the regulations	Regulators will need the flexibility to adjust requirements as the situation evolves.
<u>Require Know Your Customer and capacity reporting for AI hardware</u> designers, owners, and infrastructure providers • Explicitly empower the regulator to licence these entities • Enable the regulations to adjust the requirements as the situation evolves	Compute is a key bottle-neck for large training runs. To have time to catch any fast-moving bad actors intent on training high-risk or unacceptable risk systems, **government needs to know what the existing hardware and infrastructure capabilities are in Canada**. Requiring transparency for the physical supply chain for high-risk systems will by extension help track capacity for unacceptable-risk systems.
<u>Apply the criminal charges to licensing violations</u> (i.e. to persons illegally pursuing or enabling development of HRGPAIS, such as those that distribute the source code or model weights of such models)	HRGPAIS are a step away from unacceptable risk systems. Distributing the model weights of such a system effectively records and proliferates the extreme training amounts to everyone including the malevolent or reckless. The law needs to send a strong signal that any hiding or obfuscation of the development or deployment of HRGPAIS is a serious crime.

<u>High-risk single-purpose AI systems (HRSPAIS)</u>	This category most resembles the high-impact systems initially conceived of when the Bill was introduced in 2022. As rightly specified in the Bill, biassed outputs are a very real and harmful output of systems, especially when making decisions on people's jobs, loans or jail sentences.
<u>Define HRSPAIS as</u> AI systems, or models capable of powering AI systems, that are single-purpose or few-purpose in nature and: ● are not covered by existing laws and sectoral regulations, or	Few-purpose can be defined as any system that has a clearly defined and limited set of applications. Each application would be subject to the requirements. Most single-purpose AIS (e.g. in health, finance, transport, employment) can be regulated under existing laws. However, there may be unforeseen sensitive applications which are not, and fall between the gaps of regulatory verticals. The purpose of this clause is to avoid duplicating or conflicting with existing regulations.
● *(optional) are listed in the Minister's recent proposed redefinition of high-impact system*	*We broadly agree with the 7 classes of concern listed in the Minister's amended definition, although we would caution that many could be covered by sectoral legislation. The better strategy would be to empower an ISED-based AI commission to provide these respective ministries with the domain expertise they need to adequately govern AI in their sector, instead of requiring companies to comply with two separate regulators.*
● *(optional) Meet the definition of high risk in the EU AI Act* ● Meet other criteria to be defined in the regulations	*Optional: add wording to make the AIDA inclusive of single-purpose systems defined as high risk in the EU AI Act (Title III / Annex III).*
<u>Apply the amended "high-impact system" requirements</u> to HRSPAIS, while allowing flexibility: ● Assessments ● Measures related to risks ● Monitoring of mitigation measures ● Keeping general records	The current requirements for high-impact systems, updated with the relevant amendments proposed by the Minister, can be applied for this category of AI system.

• Publication of description
  ○ Making system available for use
  ○ Managing operation of system
• Notification of material harm
• Other requirement to be defined in the regulations
	Regulators will need the flexibility to adjust requirements as the situation evolves.

<u>Moderate or low-risk systems</u> Exempt from the Act by default. *(optional) Define as a system that meets the criteria for moderate or low-risk AI systems that are established in the regulations.* *(optional) Allow regulators to create specific and proportionate requirements for this category if the need arises.*	This will allow the vast majority of AI development to avoid government red tape, without causing serious harm to Canadians. If unforeseen individual or collective harms arise from certain low or moderate-risk systems, regulators will be able to either 1) update the criteria of HRGPAIS or HRSPAIS to include those problematic systems and regulate them under the high-risk categories, or 2) define this Moderate-risk category and create new and proportionate requirements.

Fix critical gaps

Recommendation	*Rationale*
<u>Update the preamble</u> to align the purpose of the Bill with addressing the full scale of risks involved, and what gaps AIDA needs to fill.	The current preamble does not acknowledge the current and upcoming impacts of AI that the Act most needs to address. We recommend adding: *"Whereas artificial intelligence capabilities are being rapidly developed and could soon surpass human abilities in all domains, creating unprecedented opportunities for growth and wellbeing, but also unprecedented individual and collective risks including global catastrophe.* *Whereas existing federal legislation only applies to artificial intelligence systems in specific industries or contexts, leaving large categories of novel harms unaddressed, especially those made possible by general-purpose AI systems."*

Remove exemptions for government and national security	Ensure that government and national security (and their contractors) be included in any moratoriums on unacceptable-risk systems, and that Canada has a unified and coherent monitoring and licensing regime.
Eliminate the system/model loophole by including AI models Define AI models as a *"parameterised representation of knowledge learned through an automated training process"*	Currently only complete AI systems are regulated. Per the Companion Document: "models alone do not constitute a complete AI system, the distribution of [models] would not be subject to obligations regarding "making available for use."" This is inconsistent as models can be used directly and immediately by novice technical individuals, or by less technical individuals with a few minutes of instruction. To make an analogy, this would be equivalent to a law regulating computers that exempts them if they don't include a monitor and keyboard.
Update section 4 (a) Purposes of the Act from its current exclusive focus on interprovincial and international trade to be inclusive of all AI systems and models: *"The purposes of this Act are* *(a) to regulate artificial intelligence systems and models by establishing common requirements, applicable across Canada, for the design, development and use of those systems* *(b) to prohibit certain conduct in relation to artificial intelligence systems that may result in serious harm to individuals or harm to their interests."*	Currently the Act focuses on the interprovincial and international trade mandate of the federal government. This would leave out open-source models that aren't commercial in nature, usage by political parties, and potentially other unforeseen scenarios. These are critical gaps in the Act's stated purpose because as discussed earlier, open source models could eventually provide WMD-level capabilities, and political party misuse of AI to manipulate voters could fundamentally invalidate the freedom and fairness elections. Constitutionally, global-scale risk from advanced AI passes key federal mandate tests of 1) it being a new matter which did not exist at Confederation, and 2) that the nature of the problem is one which cannot be overcome without national action. Canadians simply cannot rely on 10 provincial and 3 territorial governments to efficiently coordinate on a fast-moving, high-stakes technology with potentially catastrophic risks. To adequately protect Canadians the federal government will need to govern all high-risk models (regardless of province, sector, or purpose) in a coherent and centralised manner.

Provide government the capacity it needs

Recommendation	*Rationale*
<u>Establish a Canadian AI Safety and Ethics Commission (CAISEC)</u> Mandated to govern high and unacceptable risk AI systems, models and their hardware: • Regulate their development, deployment, possession and use • Manage the licensing regime for high-risk systems • Monitor developments in AI and update regulations in an agile manner • Select and approve standards for safety, cybersecurity and auditing • Investigate incidents and provide recourse to those harmed by AI systems • Support industry with compliance • Support civil society with education and safe adoption • Support and harmonise AI regulations in other ministries • Work with the Treasury Board Secretariat to ensure no gaps in government or defence use or understanding of AI • Work with municipal, First Nations, provincial, and international partners Modelled after the Canadian Nuclear Safety Commission Housed in ISED and reports to parliament via the Minister	For government to adequately protect Canadians from AI harms, it will need significant capacity to: - monitor a rapidly evolving and expanding landscape of risks, including individual and collective harms, - continuously update regulations and guidelines to address new harms, - rapidly enforce the rules when breached, - coordinate across government and with global partners to ensure harmonisation, - administer the licensing scheme, - and provide internal and external stakeholders with the support they need. In practice this means a permanent body with at least 50+ staff and the power to issue orders and make regulations. The proposed AI & Data Commissioner and their office will simply not be up to the task. The closest existing model for a government body dealing with global risks at the scale of human-level AI, is the Canadian Nuclear Safety Commission. This proposal is largely inspired by it, but we are open to other models if need be (such as creating a dedicated Ministry of AI). What is important is that it have the authority, agility and capacity to protect Canadians. While there is a conflict of interest with ISED's mandate to boost innovation, it's the best available location for an AI-focused

	commission. No other ministry is as directly connected to, and responsible for, the tech sector, and it fits with the pattern of the CNSC located in Natural Resources and CRTC in Heritage. Moreover, making CAISEC a parliamentary office would cause issues because it would need to be much bigger than the Privacy Commissioner's Office, would make coordination with the rest of government harder and slower, and would likely also duplicate existing work at ISED.
Commission leadership to appointed by the Governor in Council based on their qualifications and expertise, and include representation from (at least) Privy Council Office, Public Safety, Treasury Board Secretariat, Office of the Privacy Commissioner, Global Affairs Canada	To improve oversight and independence from ISED and its previously noted conflict of interest, and to limit siloing of efforts, ensure that the commission leadership be independently appointed and have representation from key related ministries.

Part III: Specific wording to change in the Bill

The wording will be provided as an addendum to this submission after the text of the government's amendments have been shared