cohere

March 1, 2024

Cohere Inc. (Cohere)
171 John St 2nd Floor
Toronto, ON M5T 1X3

Mr. Joël Lightbound, M.P.
Chair, Standing Committee on Industry and Technology (INDU)
House of Commons
Ottawa, Ontario K1A 0A6

joel.lightbound@parl.gc.ca

Re: Submission to the Committee for Study of AIDA

Dear Mr. Lightbound,

Thank you for the opportunity to make a submission to INDU regarding Part 3 of Bill C-27, the federal government's proposed *Artificial Intelligence and Data* Act ("AIDA").

Cohere is a leading global enterprise-focused generative AI company with deep Canadian roots. Cohere is in a unique position to make this submission on the regulation of AI in Canada to INDU as an innovative growth-stage business headquartered in Canada, and among a handful of model developers worldwide with global operations. Cohere's business-to-business deployment focus represents a critical and distinct voice within the AI ecosystem. As a model developer that sells to enterprises, Cohere is exposed to a wide range of application areas and use cases in AI deployment, and a diverse array of organizations implementing AI in their everyday business activities. In addition, Cohere is the only generative AI model developer to sign the White House's *Updated Voluntary Commitments*, the Canadian Federal Government's *Voluntary Artificial Intelligence Code of Conduct for the Responsible Development and Management of Advanced Generative AI Systems,* and to also endorse the G7's *Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems*.

While Cohere is supportive of the introduction of a legislative scheme to regulate AI in Canada, it is vital that the responsible and trustworthy AI regulatory approaches currently being advanced by governments and intergovernmental agencies across the world coalesce around a globally interoperable set of principles and standards. Lack of interoperability between AIDA and foreign legislative schemes and instruments will have a significant adverse economic impact on the Canadian AI ecosystem. Accordingly, it is critical that the final text of AIDA is technologically neutral and reflects a principles and risk-based approach that:

(i) is aligned with its top trading partners (e.g., US, UK, Japan, Australia, and the EU) and with ongoing global and inter-governmental initiatives focused on developing international AI-governance frameworks and standards (e.g., G7 Hiroshima AI Process, and the UN High-Level Advisory Body on Artificial Intelligence);

(ii) balances the twin objectives of promoting AI innovation and adoption while protecting individuals and society;

(iii) creates obligations between providers, downstream providers and deployers of foundational models that are proportional, contextual and role-based;

(iv) provides clarity on scope and application; and

(v) minimizes unnecessary regulatory burden that, as a matter of practice, has the effect of disproportionately impacting Canadian start-ups, 'scale-ups' and SMEs.

In Annex A to this submission, we have set out proposed amendments to the current text of AIDA which are particularly important for foundational model developers. The proposed amendments are necessary to ensure that the Federal government implements an interoperable and risk-based approach to the regulation for responsible AI in Canada that is appropriately tailored to the differing roles of various actors in the AI value chain.

Thank you for your consideration of our submission. We would be pleased to speak about any questions or comments you may have in greater detail.

Sincerely,

DocuSigned by:

Melika Carroll

403C178009844F1...

Melika Carroll
Head of Government Affairs and Public Policy

Encl.: Annex A (AIDA: Proposed Amendments)
Cc. : Miriam Burke, Clerk, INDU (INDU@parl.gc.ca)

ANNEX A: AIDA - PROPOSED AMENDMENTS

Recommendations

- **Amend the text of AIDA to ensure the scope of the proposed legislative scheme is interoperable and aligns with Canada's trading partners and emerging international AI-governance frameworks and standards.**

Specific Recommendations:

- Limit the scope of 'high-impact system' under Section 5(1) by expressly clarifying in the statutory definition that a high-impact system is one that poses a significant risk to the health, safety or fundamental rights of natural persons.

- Revise the concept and provisions related to general-purpose systems by:

    o amending the definition of a 'general-purpose system' under Section 5(1) to clarify that such AI systems: (i) are based on general-purpose machine learning models, and (ii) actually have the capability to serve a variety of purposes; and

    o ensuring the scope of the majority of obligations on general-purpose systems under Sections 7 and 8 are reserved for general-purpose systems that pose a significant risk to the health, safety or fundamental rights of natural persons.

- Replace the references to machine learning models in Section 9 with a definition that expressly refers to general-purpose machine learning models *with systemic risk.*

- Include a materiality threshold to the list of factors in Sections 8.1(2) and 10.1(2) that would give rise to statutory requirements upon a 'change' in the use of an AI system or risk mitigation measures relating to an AI system.

- Revise Section 5(2) to clarify that an AI system should be either a general-purpose system or a high-impact system, but not both.

- **Rationale for Recommendations:**

    - 'High-impact system' definition clarification

        - The current definition of a 'high-impact system' is overly broad. As proposed, the scope of AI systems regulated under AIDA as high-impact is materially broader than "high-risk" AI systems classified under the EU's AI Act. This means that far more use cases of AI in Canada will be subject to the set of statutory requirements as compared as compared with the EU.

- Under AIDA, the full scope of obligations - some of which create significant compliance burden - apply to persons that make available or manage the operations of broad enumerated classes of high-impact systems even if their activities do not give rise to any material risk of harm. For example, any AI system involved in any aspect of the employment lifecycle would be designated as a 'high-impact system' regardless of whether the ultimate HR related decision is made by a human.

- ▪ 'General-purpose systems' - Scope and Obligations

  - *Definition*: The current definition of 'general purpose system' is overly broad. In practice, <u>general purpose machine learning models</u> are the engine of a general-purpose system. The broad scope wording under AIDA's definition of 'general-purpose system' imposes requirements on *AI systems* that do not have general capabilities. Our proposed recommendation to AIDA's definition of a 'general-purpose system' more closely aligns with the similar concept under the EU AI Act.

  - *Scope and application*: Under AIDA, the full scope of obligations apply to persons that make available or manage the operations of general-purpose systems, without any consideration for the deployment context, i.e., whether the deployment is: (i) linked to activities that give rise to a risk of significant harm; or (ii) business-to-business versus business-to-consumer.

  - *International alignment*: While internationally there has been no consensus to the precise regulation of foundational models, AIDA's broad approach to general-purpose systems is out of step with its closest allies. For example:

    o As currently drafted, the text of AIDA imposes a range of obligations for general-purpose systems that exceeds the nature and scope of the requirements for general-purpose systems under the EU AI Act.

    o The UK government takes a pro-innovation regulatory approach to AI, including to general-purpose AI systems. In its *[Response to the AI Regulation White Paper Consultation](#)* released in January 2024 (the "Response Paper"), the UK government:

      ▪ Acknowledges the risks posed by general-purpose AI systems but suggests that any legislative action should take place only 'once understanding of risk has matured'.

      ▪ Outlines its immediate plan is to use voluntary measures. The Response Paper stops short of making any sort of commitment to introducing legislation in the

UK instead focusing on the role of voluntary measures in mitigating against the risks posed by these models.

- States that any future binding measures would only be introduced if 'existing mitigations were no longer adequate'. Even if it were to move to binding measures, the UK government's approach is going to be measured: any future measures would be applied to the most powerful general-purpose systems based on 'dynamic thresholds' that can quickly respond to developments in AI. This could be based on forecasts of capabilities using a combination of two proxies: compute and capability benchmarking. Any binding obligations would be principles based, focusing on the principles set out in the white paper of safety, security, transparency, fairness and accountability.

- o In contrast, in the US the *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* ("EO") sets out a standards-and-principles based approach to AI regulation. With regard to general-purpose AI systems, its approach is focused on frontier models - specifically *dual-use foundation models* - and the obligation related to such model developers is forward looking. The rationale for the EO's approach to dual use foundation models (and related obligations) is to ensure the US government better understands the capabilities of these frontier models and the EO enables this by "shining a light under the hood".

- o In January 2024, the Australian government's interim response to its AI consultation noted that will employ a risk-based approach to the regulation of AI and seeks to prevent the harms associated with AI use, by regulating the development, deployment and use of AI in high-risk contexts only, with other, lower-risk forms of AI being allowed to "flourish largely unimpeded". With respect to general purpose models, the Australian government noted the need to address frontier models.

- The common thread among the different approaches to the regulation of general-purpose systems (i.e. those that make use of frontier models that could give rise to systemic risks) is that obligations under foreign instruments is narrow and tied to materiality benchmarks.

- To the extent Canada seeks to regulate general-purpose systems more broadly - in a manner that deviates from its international partners - as is currently proposed under AIDA, such an approach would: (i) seriously adversely impact the AI innovation landscape in Canada; (ii) disproportionately impact on SMEs and 'scale-ups' that are at the

forefront of this technology; and (iii) create additional barriers for the adoption of AI by Canadian enterprises as compared to enterprises in other jurisdictions around the world.

▪ To ensure consistency with a risk-based approach throughout AIDA, a materiality threshold is required for the definition of a 'change' to the factors set out under Sections 8.1(2) and 10.1(2) that would give rise to statutory requirements upon a 'change' in the use of an AI system or risk mitigation measures relating to an AI system. The absence of a materiality threshold would create undue regulatory burden for inconsequential changes to AI systems or related measures that would not give rise to material harms.

● **Revise the requirements under AIDA imposed on persons that 'make available' or 'manage the operations' of general-purpose AI systems to ensure they are appropriately tailored to the organization's role within the AI value chain.**

> **Specific Recommendations:**
>
> - Clarify Sections 7(1)(b), 7(1)(c), and 8.2(1)(b) to specify that the organization that makes available or manages the operations of the AI system (i.e., the deploying or other "downstream" organization) is the organization that is responsible for conducting the assessment of adverse impacts that 'could result form any use of the system that is reasonably foreseeable'.
>
> - Clarify the obligations to implement risk mitigation measures set out in Sections 7(1)(c), 7(1)(d), 7(1)(h), 8.2(1)(b), 8.2(1)(c) and 8.2(1)(f) are the responsibility of that downstream person.
>
> - Clarify that the deploying or other downstream organization are the intended entities responsible for compliance with Section 7(1)(g) for persons that make available general-purpose systems to use best efforts so that members of public can identify outputs as having been generated by an AI system.
>
> - Clarify the extent of the responsibility of a foundational model provider under Sections 7 and 8 is to provide appropriate support, as appropriate in the circumstances, to such downstream actor with respect to the requirements in such provisions.

o **Rationale for Recommendations:**

▪ Many of the requirements under the current text of AIDA are simply not applicable to foundational model providers as compared with other actors in the AI chain.

▪ It is critically important to appropriately tailor the requirements under AIDA to align with the role and activities of the actor within the AI value chain, i.e., they are proportional and role-based. Such a balanced approach that considers the

nuances of the AI value chain provides flexibility, promotes responsibility across the AI value chain, reduces duplication, and is in line with an accountability and risk-centric approach.

- Currently, the requirements in Sections 7 and 8 for general-purpose systems apply broadly to any persons who make available a general-purpose system or manage its operations. By implication, as drafted, these requirements do not take into consideration that foundational models are intended to be adapted into a broad array of downstream applications and use-cases, many of which are not known to, or under the control of, the foundational model provider.

- A more appropriate, calibrated approach would take into consideration: (i) the role of the actor within the AI-value chain; (ii) the relationship between these actors; and (iii) who determines the AI system's intended purpose and how it is deployed (e.g., B2B versus B2C). In most instances, it is the deployer or downstream organization - not the person who initially makes available the AI system (i.e., a model developer) - who would be in a position to reasonably foresee risks that could result from their specific use-case. Risk identification and mitigation frameworks are ultimately dependent on the context in which an AI system is deployed. Context encompasses intended use, as well as language and culture, and many other factors local to the deployed environment and not known to the foundational model provider.

- Even though the model developer may be able to evaluate against high-level benchmarks, the deployer in most instances is the party that controls how and where it is used, including by end-users. Additionally, deployers often build and train their own proprietary models through fine-tuning workflows which can change the risk profile of a model.

- Currently, there is not a clear connection between foundational model (upstream) measurements and application harms. A foundational model is not a product or interface, but it powers many products, and so it would be seen as pragmatic to be able to measure and correct safety issues "upstream" at the model level (rather than "downstream" in the product). Even though there are ongoing efforts trying to do upstream mitigation, recent research shows that such upstream mitigation efforts may not always be successful, and that if you mitigate or measure upstream there isn't a reliable downstream effect: *Goldfarb-Tarrant et al* (2021)[1], expanded and replicated in *Cao et al (2022)*[2] and *Steed et al* (2022)[3].

- With respect to transparency requirements and requirements to ensure members of the public are able to identify AI generated outputs under Sections 7(1)(f) and 7(1)(g), the deployer or other downstream organization - not the developer of foundation models – is the best-placed actor to comply with

---

[1] Goldfarb-Tarrant et al, *Intrinsic Bias Metrics Do Not Correlate with Application Bias* (2021): https://arxiv.org/pdf/2012.15859.pdf.

[2] Cao et al, *On the Intrinsic and Extrinsic Fairness Evaluation Metrics for Contextualized Language Representations* (2022): https://arxiv.org/pdf/2203.13928.pdf.

[3] Steed et al, *Upstream Mitigation Is Not All You Need: Testing the Bias Transfer Hypothesis in Pre-Trained Language Models* (2022): https://aclanthology.org/2022.acl-long.247.pdf.

transparency requirements as: (i) the detail in any transparency measure is ultimately dependent on the context and domain in which an AI system is deployed; and (ii) the downstream organization is best situated to implement the transparency in a manner that would be meaningful to the end users. Additionally, in some instances, even if a provider were to implement a watermark or other labelling technique, this could be removed in the deployment phase without the provider's awareness and without the provider being able to intervene. Instead, the transparency obligations should focus on a collaborative approach between developer and deployers but ultimate accountability should rest with the end-point deployer who controls how an end-user interacts with the AI system.

▪ Given the role of a foundational model provider in the AI value chain, it would be commercially unreasonable to impose of these requirements on the upstream foundational model developer to make available software to detect AI generated content at no cost.

● **Amend the text of AIDA to introduce definitions for the different actors in the AI value chain.**

Specific Recommendations:

● To more clearly articulate the scope of obligations imposed on persons in the context of their role in the AI value chain, provide definitions for 'made available' and 'manages the operations' of an AI system or machine learning model.

● To ensure interoperability, it will be critical to ensure the concepts of 'made available' and 'manage the operations' are defined in a manner that align with obligations of AI actors under the EU AI Act and other foreign instruments.

---

o **Rationale for Recommendations:**

▪ The inclusion of definitions for 'made available' and 'manage the operations' will serve to reduce regulatory compliance uncertainty by clarifying understanding about which specific requirements apply to an actor in the AI value chain.

▪ The concepts of 'made available' and 'manage the operations' are not defined under the EU AI Act or the *G7 Hiroshima Process International Code of Conduct for Organizations Developing Advanced AI Systems*. To ensure interoperability with such statutory and other instruments, it will be critically important to draft the proposed definitions to take into account the requirements imposed on AI actors within foreign instruments these and others currently being developed.

- **Include a risk-based qualification to the requirement for third-party conformity assessments**

> **Specific Recommendations:**
>
> - Amend the requirement for third-party conformity assessments under Section 7(1)(i) to clarify that they only apply to a subset of general-purpose systems that pose significant risk to health, safety or fundamental rights of natural persons in accordance with criteria prescribed by the regulations.

  o **Rationale for Recommendations:**

    ▪ The requirement under Section 7(1)(i) for third-party conformity assessments for all general-purpose systems introduces significant compliance burden that is not proportional to the risk.

    ▪ These conformity assessments for complex foundational models can be expensive and require the skills of a scarce supply of experts with the requisite skillset. This results in practical barriers for many organizations, and in particular, innovative AI companies with limited resources.

    ▪ An approach in AIDA where the application of third-party conformity assessments is limited to high-risk circumstances would be consistent with the approach adopted in the EU, which requires third-party conformity assessments only in the case of limited high-impact systems (e.g., if the AI system is part of biometric identification).

- **Limit the scope of retroactive application of AIDA to AI systems that have been significantly modified after AIDA has come into force.**

> **Specific Recommendations:**
>
> - Revise Sections 7(3), 8(2), 8.2(2), 9(3), 9.1(2), 10(3), 10.1(2), 11(2), 12(2), 12(4) to ensure that AIDA's requirements will not apply retroactively, except in circumstances where there has been a significant modification to the AI system after AIDA has come into force.

  o **Rationale for Recommendations:**

    ▪ Our proposed recommendation ensures interoperability with the provisions that apply retroactively under that regime with those set out in the EU AI Act, which applies to systems that experience significant changes in their design.

- ▪ Limiting the retroactive application of AIDA would reduce uncertainty and unnecessary burden for businesses that deploy or make available AI systems between now and when AIDA comes into effect.

- ● **Include a general proportionality clause.**

   - ○ **Rationale for Recommendations:**

      - ▪ Currently proportionality is not expressly referenced as a principle or a contextual consideration.

      - ▪ In contract, under the EU AI Act, proportionality is identified as an overarching objective and principle that informs its interpretation and enforcement.

We also recommend that in reviewing the obligations proposed under AIDA, INDU should employ a lens that balances the risk and safety considerations by also understanding whether these obligations: (a) impose compliance burdens that may disproportionately impact Canadian SMEs and scale ups (e.g., by imposing fixed costs irrespective company's size or growth stage); (b) impede AI innovation; and (c) hinder the responsible adoption of AI.

Lastly, we acknowledge comments made by stakeholders calling for the express inclusion of copyright related provisions within the statute. While copyright and related issues will need to be carefully considered in the context of AI systems, these are highly nuanced and complex issues that are far more appropriate and best served within a holistic consideration of the Canadian copyright framework, including as part of the *ISED Consultation on Copyright in the Age of Generative AI*. For example, the policy rationale to include training data disclosures for general purpose AI systems in the EU AI Act was made on the basis that under the *EU DSM Copyright Directive*[4] copyright holders are allowed to reserve the right to use *EU DSM Copyright Directive*'s text and data mining exemption or content that is publicly available online if they implement appropriate technological measures. Currently, there is no text and data mining exemption under Canada's copyright framework.

---

[4] *Directive (EU) 2019/790 on copyright and related rights in the Digital Single Market* (Article 4).