September 6, 2024

Innovation, Science and Economic Development Canada
235 Queen St,
Ottawa, Ontario
K1A 0H5

**Delivered via email: aicompute-calculia@ised-isde.gc.ca**

On behalf of Canada's business community, we thank you for the opportunity to participate in the consultation on Artificial Intelligence (AI) Compute.

The Canadian Chamber of Commerce is the country's largest and most activated business network — representing over 400 chambers of commerce and boards of trade and more than 200,000 businesses of all sizes, from all sectors of the economy and from every part of the country. As such, we consulted with members from a variety of sectors, including telecommunications, education, cyber security, finance, insurance, and healthcare among others, represented on the Canadian Chamber's Digital Economy Committee and Future of AI Council.

We strongly believe that this is an important and timely topic for discussion, and we are pleased to provide our input for your consideration.

Again, thank you for launching this discussion. We gladly make ourselves available at your convenience to meet and discuss our letter and recommendations further.

Sincerely,

**Ulrike Bahr-Gedalia**
Senior Director, Digital Economy, Technology & Innovation
Canadian Chamber of Commerce
**C: 613.410.6629 E: ubahr-gedalia@chamber.ca**

**Olha Sotska**
Policy Advisor
Canadian Chamber of Commerce
**C: 343.997.3027 E: osotska@chamber.ca**

**Re: Consultation on Artificial Intelligence (AI) Compute**


1. **How do you currently access compute? What are the most pressing issues you face in accessing AI compute in the ecosystem currently (i.e., level of access, cost, security etc.)?**

Industry utilizes different ways to access compute, ranging from on-prem solutions to relying on cloud computing. The on-prem approach keeps data private and on-premises while reducing or eliminating commercial cloud expenses. Meanwhile, cloud computing offers fast democratized access to technology, as everyone has equal access to the same tech and services as and when they need it. Moreover, cloud computing boasts unparalleled elastic ability to rapidly deploy multiple CPU and GPU cores (and scale down), enabling users to train models in hours rather than days or weeks. This is essential given the vast and growing scale of data use, which current on-prem solutions struggle to handle efficiently.

Looking ahead, advancements in access to energy and innovations that aim to reduce dependency on large datasets should be priorities with one approach being the leverage of open-source AI. These needs are driven by the growing complexity of models, improved architectures, the growing need to expand infrastructure investments to meet customer demand, and the desire for higher performance requiring large-scale compute platforms.

2. **What level and type of compute access do you require to grow and scale your business or conduct research? Are there existing programs or support mechanisms that could be leveraged in the short term?**

To appropriately serve different sectors of our economy, industry requires scalable access to high-performance CPU and GPU resources to handle a range of AI algorithm training and validation tasks, from small to large jobs. At the leading edge of AI commercialization, companies need state of the art compute (GPUs) to train models and to serve the models. None of the skills or technologies to build the compute to train models exist in Canada today. These supercomputers, with the latest and most advanced chips and networking are very challenging to build and operate.

Organizations have varying needs that are best met with models of different sizes. Training a new foundation model or other state-of-the-art AI from scratch can require significant resources that are not readily available for most developers. Open models and tools lower these barriers to entry. Developers can build on open models to create new AI tools or launch new AI ventures without spending significant amounts on research and training compute. This is critical given that smaller firms, especially start-ups, face difficult affordability and access challenges regarding the required compute needed.

3. **What new approaches should the Government consider to support researchers and industry access to AI compute in the short term?**

Government has made tremendous efforts in the research space with world-renowned trailblazers, such as Geoffrey Hinton and was the first country to launch a national AI strategy, the Pan Canadian AI Strategy, which successfully focused on growing a critical mass of AI

talent in Canada composed of both leading AI researchers and practitioners. Now, it is critical for government to consider new approaches that focus on advancing commercialization:

- Invest in advanced cloud infrastructure by aggregating demand and consolidating the purchase of compute at scale from cloud service providers (CSPs) that would allow to offer immediate compute capability to stakeholders. Providing access to government-funded compute resources could help offset costs and enhance computational capabilities, enabling businesses to drive innovation and support Canada's competitive edge.
- Develop state-of-the-art computing solutions with scalable, upgradable infrastructure. As a starting point, government could consider utilizing existing, underused infrastructure as a feasible, ultra-low-cost, immediate-term approach to increase compute capacity and accessibility.

4. **What is the benefit to you of having computing infrastructure that is Canadian-owned and controlled?**

- Collaboration and Innovation: Facilitate partnerships with independent not-for-profit Canadian institutions, fostering collaboration and innovation within the national AI ecosystem.
- Economic Development Support: Support the local economy by investing in Canadian technology and talent, promoting job creation and technological advancement.
- Open-Source AI: Open source-AI can help maintain these benefits as it facilitates more control over data access as open models can be downloaded onto local hardware, rather than requiring developers sharing data with external providers. Organizations with concerns over data access (i.e. data protection, proprietary, or other reasons) may be better positioned to experiment and innovate with open models.
- Data Sovereignty and Security: A sovereign compute infrastructure would mitigate risks associated with foreign controlled entities reallocating compute during high demand periods. Consistent access to critical compute is a matter of national security.

5. **What are the short-term opportunities or initiatives that would help to expand existing computing infrastructure in Canada?**

To increase our national capacity, the government could consider the following key short-term opportunities:

- Leverage compute from companies that may have underutilized or reserved compute.
- Leverage open-source AI.
- Offer cloud credits and technical support to boost innovation and adoption.
- Be an early adopter of new technologies and aim to integrate these new technologies into public sector operations where possible.
- Put in place strategic procurement policies that enable businesses to grow, scale and invest in innovation, ultimately further strengthening the economy and enhancing resilience and thereby build domestic capacity, de-risk investment/access, and enable broader uptake.

**6. In your view, what are the opportunities to incorporate Canadian-made computing hardware and software as part of Canadian AI Sovereign Compute Strategy?**

There are abundant opportunities to integrate Canadian-made technologies into Canada's AI Sovereign Compute Strategy. Part of this process should be deciding what aspects of the AI tech stack Canada wants to accelerate through these investments. As such, the following questions should be addressed:

- Do we want to be the hardware providers that support the training and running of LLMs?
- Do we want to provide the LLMs/FMs and tooling to build AI apps?
- Do we want to develop AI applications that leverage LLMs/FMs and said tools? Lack of clarity risks sub-optimizing this investment.

Canadian-made computing hardware also offers multiple advantages, from the development of solutions specifically tailored to the needs of Canadian businesses to stimulating economic growth, creating jobs, and fostering innovation within Canada. As introducing new tech can be daunting, leading many SMEs to potentially opt for off-the-shelf, more expensive cloud services, without government incentives, Canadian SMEs are unlikely to embrace adopting innovative Canadian-made tech.

**7. How can we leverage this investment to both retain and attract AI talent in Canada?**

Canada has the talent. What we are severely lacking is increased commercialization to keep the talent here. We have a long history of sluggish commercialization, and it has become a critical, time-sensitive issue Canada needs to address now.

The retention and attraction of talent is a function of creating a critical mass of opportunity and future potential. In many ways, Canada already possesses many of the requisite ingredients. This investment can further advance Canada's leadership, including attracting and retaining talent, by democratizing access to compute and emerging technologies. Providing the target talent cadre access to an attractive environment with cutting-edge technology, coupled with continuous learning and career development opportunities through partnerships with academic institutions, polytechnics and industry, will build a sustainable AI talent pipeline and reduce brain drain, ensuring Canada remains competitive in the global AI landscape.

**8. Are there specific collaboration models that would help to support industry/academic partnerships?**

- Partnering with academic institutions and polytechnics on joint research projects leverages both industry expertise and academic capabilities, encouraging knowledge exchange and co-development of innovative solutions.
- Offering internships, co-op programs, and fellowships provides recent graduates with practical industry experience while allowing businesses to tap into local fresh talent and new ideas.
- Funding specific research projects, with industry partners determining the scope of research rather than discovery researchers in the existing centers, to align with private industry goals can include shared intellectual property and commercialization rights.
- Forming advisory boards with industry and representatives from Universities and Polytechnics to guide collaborative projects ensures they meet both commercial and research objectives.

- Leveraging cloud computing capability to scale innovative ideas immediately, instantly reaching global markets, through a secure collaboration across institutions, sectors, industries, and continents.

**9. Which model, or combination thereof, do you think would achieve results most efficiently and effectively?**

There is no one model that solves every policy challenge. For example, the needs of start-ups are very different than those of scale-ups and researchers. Yet, a combination of the following models would most likely achieve the most efficient and effective results in supporting industry/academic partnerships:

- Expand the capacity of compute infrastructure providers in Canada to incentivize increases in the supply and availability of compute within Canada. Support could take the form of grants to help offset capital costs, or through tax or energy incentives.
- Democratize access to technology, leveraging the variety of technology, while taking advantage of the economies of scale.

**10. What would you view as the priority components or elements of National AI Compute Infrastructure for the next five years?**

- High-Performance Computing (HPC) Resources: Access to scalable and powerful CPUs and GPUs, at a competitive cost.
- Skilled Talent Development: Education and training programs for AI professionals
- Public-Private Partnerships: Collaboration between government, industry, academia, and polytechnics.
- Standardized Data Protocols: Interoperability standards for seamless data exchange.
- Sustainable Infrastructure: Energy-efficient computing solutions.

**11. How can public infrastructure be leveraged to support both the needs of researchers and industry? Are there unique requirements to these stakeholder groups which should be considered?**

Investments in public infrastructure can support both researchers and industry if tailored to their unique needs. Government might consider collaborative agreements with private companies that build, operate or manage public infrastructure in exchange for subsidies, investing in maintenance or upgrades.

Other key considerations include scalable HPC resources, storage, high throughput networking, analytics, AI, and ML services, and data lakes. Addressing these factors can enhance support for research and industry, driving innovation and AI development in Canada.

**12. What level of support to compute infrastructure providers would incentivize an increase in the supply and availability of compute in Canada and over what length of time could an increase in supply be expected?**

Streamlined regulations for compute infrastructure providers would incentivize growth. Support for public-private partnerships and innovation grants can further drive investment. However, the speed at which the private and public sectors operate is both quantitatively and qualitatively different. Quantitatively, the private sector often moves faster due to competitive pressures,

profit motives, and the need to respond quickly to market demands, leading to rapid decision-making and implementation.

### 13. How can Canadian compute be leveraged to collaborate and expand relationships and networks with like-minded countries?

A Federated Learning network can support secure, distributed AI training across borders. This approach enhances relationships with like-minded countries, enabling shared resources, expertise, and advancements in AI while maintaining data privacy. This network could also be used to monetize Canadian data through this worldwide access while improving the generalizability of algorithms. Additionally, through a partnership with CSPs, Canada could instantly get access to a robust global infrastructure that enables easy collaboration with like-minded countries.

### 14. Are there any considerations that we have missed or elements we should explore further when addressing this topic? Please provide any further feedback you feel should be considered.

- Commercialization: Canada has a long history of sluggish commercialization. It's time for Canada to reap the benefits of AI investments.
- Public Awareness and Engagement: Public Consultations and educational campaigns centered around transparency and impact of "AI for good" to foster a supportive environment for AI initiatives.
- Global Collaboration: Leveraging global expertise can position Canada as a leader in AI by participating in global AI initiatives.
- Integrating Ethical AI and Bias Mitigation Practices: Fundamental when developing infrastructure frameworks to ensure AI systems are fair, transparent, and accountable.
- Scalability and Flexibility: Infrastructure should be designed with flexibility to scale with the increasing demands of AI research and industry applications.
- Data Sharing and Privacy: Establishing secure data-sharing protocols between provinces and international partners can enhance collaborative research efforts and help build trust among stakeholders.
- Develop and Adopt Interoperability Standards: Facilitates seamless integration and collaboration across different platforms and institutions, maximizing the utility of the infrastructure and ensure smooth data exchange.
- Sustainability and Green Computing: Investing in energy-efficient and sustainable computing solutions minimizes environmental impact, while aligning with global efforts to combat climate change and enhancing the long-term viability of the infrastructure.
- Value of Compute Infrastructure in Isolation Compared to the Value for Ecosystem Partners:
  - How does investing in 'public' infrastructure materially help access?
  - What signal does government want to send to the private sector regarding foreign direct investment into digital infrastructure?
  - Government competing with the private sector will increase prices on finite tech supply chains, further exacerbating shortages and critical mass of compute.
  - Partnership with CPSs as a path to competitive advantage and leadership.
  - Potential exclusion of major CSPs/hyperscalers from federal hardware procurement, who could have made significant investments in hardware infrastructure in Canada, could be seen as a discriminatory policy that disadvantages global firms, potentially resulting in a trade challenge.