HealthCareCAN
Leading. Innovation. Together.

SoinsSantéCAN
Leadership. Innovation. Collaboration.

Submission to the Federal Government's Consultation on Artificial Intelligence (AI) Compute

September 6, 2024

Introduction

HealthCareCAN is the national voice of action for hospitals, health authorities, healthcare and health research organizations across Canada. HealthCare*CAN* advocates in support of health research and innovation and enhanced access to high-quality health services for people across Canada and empowers health professionals through our best-in-class learning programs.

HealthCare*CAN* welcomes the opportunity to provide our input to the federal government's consultation on Artificial Intelligence (AI) Compute.

HealthCare*CAN* is committed to advancing the Canadian health sector's leadership role in AI, seeking to contribute to the development of AI strategies that will drive innovation, enhance healthcare delivery, and strengthen Canada's position in the global AI landscape. Recognizing the importance of AI in transforming the healthcare sector, HealthCare*CAN* aims to ensure that Canada's AI initiatives are aligned with the needs of healthcare providers and patients alike.

HealthCare*CAN*, in collaboration with our members, specifically our Vice Presidents of Health Research (VPRs) Committee with members from health organizations across Canada, has identified several considerations for the federal government's as it moves forward in securing Canada's AI Advantage.

Demand for AI compute resources in the health system

AI has the potential to transform and enhance healthcare delivery for providers and patients, leading to better health outcomes. High-demand areas for enhanced AI compute resources include imaging data, pathology, and predictive analytics, where the volume and complexity of data make AI an invaluable tool for enhancing diagnostics and decision-making.

The demand for AI compute resources in Canada's health system is substantial, but the successful implementation of AI within healthcare depends on improving basic infrastructure first.

Interoperable electronic medical record (EMR) systems across institutions and the country, or at least standardized data models and legal frameworks for health data access, are critical for maximizing AI's potential within the health sector. Healthcare organizations continue to highlight the inefficiencies caused by the current fragmented EMR systems across institutions, provinces and territories that prevent AI technology from being utilized. To address this, establishing clear interoperability standards and mandatory specifications for EMR systems is essential to ensure health records are accessible and usable for AI applications across the country.

Feedback from our members organizations indicates current access to computing resources needs include:

- **HPC4Health (High-Performance Computing for Health):** A consortium of health providers offering high-performance computing certified to store personal health information, specifically designed for clinical research.
- **Research Institute or Faculty:** An institute or faculty with affiliations to AI research institutes can access the institute's high-performance computing infrastructure.
- **Commercial Cloud Services:** Computing resources available through major cloud providers such as Microsoft, Google, and Amazon.
- **Digital Research Alliance of Canada (DRAC):** National high-performance computing infrastructure used for heavy computational analyses. These sites do not handle any sensitive data.

HealthCare*CAN* member's key challenges in accessing compute resources include:

- **Increased access to graphics processing unit (GPU) servers:** There is a pressing need for more GPUs that are closely integrated with hospital datasets.
- **Compliance with hospital standards:** Compute resources must meet hospital standards, including coverage under the hospital's cyber insurance for personal health information. Cloud instances provided and managed by the hospital should adhere to the same security measures as on-premise hardware.
- **Data protection and custodianship:** Healthcare organizations are responsible and liable for data protection. Therefore, any systems implemented must allow for full isolation and control of resources by creating discrete "tenants" for each institution.
- **Cost of cloud computing:** The expense of using cloud computing resources is a significant concern.
- **Queueing for shared resources:** Researchers often face delays when waiting to access shared high-performance computing resources.

The health system must prioritize investments in robust information technology (IT) infrastructure, including adequately staffed IT departments, updated networks, and functional EMRs. By building this foundational capacity, Canada can support the adoption of AI across diverse healthcare settings, ultimately improving patient outcomes and system efficiency.

HealthCare*CAN* members have identified that to grow and scale their businesses or conduct research, they require access to:

- Large GPU servers, with a minimum of Nvidia A100 GPUs
- Substantial storage capacity with petabyte-scale storage
- Fast input/output (I/O) between storage and compute

To support infrastructure needs, existing programs and support mechanisms that could be leveraged include:

- Research credits on commercial clouds offer a viable short-term solution for accessing compute resources, although their availability is limited over time.
- Hybrid computing solutions would be advantageous due to being able to utilize a combination of on-premises and cloud resources to enable "cloud bursting" – where compute jobs are offloaded to the cloud when local infrastructure reaches capacity – can be effective.
- Current research programs cannot sufficiently fund high-performance computing at hospitals, making it challenging to secure resources and funding through Canadian Institutes of Health Research (CIHR) grants.

To support access to AI compute in the short term, the following new approaches should be considered:

- Establish agreements with cloud providers to secure research credits for academia. These credits should be used within private tenancies that meet hospital security standards.
- Collaborate with cloud providers to ensure the availability and maintainability of resources, as well as access to essential software tools and features (e.g., cloud bursting, multi-cloud strategies). This approach addresses challenges such as the need for maintenance, software tools, specialized technical staff, and the high costs associated with building and refreshing large compute and storage systems, which have a short lifecycle and are often located in expensive areas.

Level and consistency of demand for AI compute resources in the health system

The demand for AI compute resources within the health system is not only substantial but rapidly expanding, particularly in critical areas such as medical imaging, pathology, and clinical documentation. These domains generate vast amounts of data, making them ideal candidates for AI-driven innovations aimed at enhancing diagnostic accuracy, decision-making, and overall efficiency in healthcare delivery. This growing demand highlights the urgent need for sovereign and secure compute infrastructure that can support both current and future healthcare needs, as the digitization of clinical workflows and the integration of AI tools become increasingly widespread.

However, the readiness of healthcare organizations to adopt and implement AI technologies varies significantly. To address this inconsistency, it is essential to build capacity across the health system by ensuring that all organizations have access to robust, Canadian-owned compute resources that meet stringent privacy and security standards. HealthCare*CAN* members have emphasized the benefits of Canadian-owned infrastructure, including subsidized and controlled costs, adherence to local privacy laws, enhanced security and reliability, and the

potential to leverage renewable energy sources. Additionally, comprehensive training and support mechanisms are necessary to equip healthcare professionals and organizations with the skills and infrastructure required to fully leverage AI technologies, thereby enabling a consistent and sustainable adoption of AI across the entire health system.

HealthCare*CAN* members identified several opportunities to incorporate Canadian-made computing hardware and software as part of Canadian AI Sovereign Compute Strategy, including:

- Better control over deployment, monitoring, and updates, especially when accessing sensitive and confidential information.
- Job creation to build, manage, and maintain infrastructure.
- Although it would be challenging to compete with major manufacturers like Nvidia, Intel, and AMD, there could be opportunities for chip development in the Application-Specific Integrated Circuits (ASICs) and Field-Programmable Gate Arrays (FPGAs) space for the creation of specialized processors, such as tensor processing units (TPUs).

Regarding attracting and retaining AI talent in Canada, HealthCare*CAN* members indicated the need to:

- Support better access to compute resources so that more AI talent stays in Canada, as current AI professionals need reliable infrastructure to support their work.
- Address the challenge of recruiting and retaining talent in academia, where compensation standards often fall short compared to industry salaries.
- Consider non-monetary compensation options, such as providing housing or creating dedicated campuses, to support talent recruitment and retention.
- Develop unique, multimodal datasets connected to innovative computational resources (e.g., GPUs, high-speed RAM, storage, CPUs).
- Incorporate compute access training and foster industry partnerships within trainee programs to build expertise in data ecosystems and specialized environments, particularly in healthcare.

Collaboration models that would help support the industry/academic partnerships include:

- Develop AI-specific commercialization pathways, including intellectual property and data transfer agreements.
- Implement a co-commercialization strategy that prioritizes public and patient benefits.
- Clearly define personal health information (PHI) and conflict of interest (COI) guidelines, similar to Health Canada Guidance Documents.
- Partner with commercial cloud providers to bring both hardware and software into Canada.
- Utilize the HPC4Health model, where central teams manage shared hardware across hospitals, while individual institutions maintain control over software, configuration, and security to ensure compliance with hospital standards.

Which type of strategy/model will best serve the health system/healthcare organizations?

HealthCare*CAN* and our members propose that, although all the models would be beneficial to address the immediate needs when working towards long-term infrastructure development, the optimal strategy/model for serving the health system and healthcare organizations is to focus on investing in the development of robust and secure infrastructure.

Specifically, we recommend a model where government funding is directed towards building scalable, hospital-standard-compliant computing facilities. This approach ensures the creation of high-performance, Canadian-owned infrastructure that can securely handle sensitive personal health information (PHI). By prioritizing long-term infrastructure development, we can achieve both immediate and sustainable improvements in computational resources, aligning with the unique needs and security requirements of the healthcare sector.

Longer-term compute infrastructure

Investing in domestic compute infrastructure will require balancing priorities and addressing various stakeholder needs across the AI ecosystem. HealthCare*CAN* members noted that it is crucial to understand the unique requirements of different industry players and the research community to ensure that the solutions implemented have a lasting and positive impact.

In the next five years, the priority components or elements of National AI Compute Infrastructure include:

- Ensuring timely access to powerful and scalable computing resources.
- Implementing efficient data storage solutions to handle large volumes of data, provide fast access, and ensure data integrity.
- Ensuring that infrastructure can securely handle confidential and sensitive information, such as personal health information (PHI). Research hospitals should have direct access to compute resources to maintain security standards.
- Investing in education and training programs to develop a skilled workforce capable of advancing AI and supporting the necessary infrastructure.
- Providing access to software tools that support AI model development, training, deployment, and monitoring. These tools should facilitate collaboration among researchers and developers.

Contact:

Marissa Persaud, Research and Policy Analyst, HealthCare*CAN* mpersaud@healthcarecan.ca
Jonathan Mitchell, Vice-President, Research and Policy, HealthCare*CAN*
jmitchell@healthcarecan.ca