ADVANCING CANADA'S NATIONAL SOVEREIGN AI COMPUTE CAPACITY

AN ICTC POLICY BRIEF

ICTC CTIC

Information and Communications Technology Council

Conseil des technologies de l'information et des communications

Research by

ICTC CTIC

PREFACE:

The Information and Communications Technology Council (ICTC) is a neutral, not-for-profit national centre of expertise with the mission of strengthening Canada's digital advantage in the global economy. For over 30 years, ICTC has delivered forward-looking research, practical policy advice, and capacity-building solutions for individuals and businesses. The organization's goal is to ensure that technology is utilized to drive economic growth and innovation and that Canada's workforce remains competitive on a global scale.

ictc-ctic.ca info@ictc-ctic.ca

TO CITE THIS REPORT:

Erik Henningsmoen, Noah Lubendo, Mairead Matthews, Heather McGeer. Advancing Canada's National Sovereign AI Compute Capacity: An ICTC Policy Brief. Information and Communications Technology Council (ICTC), July 2025. Ottawa, Canada. Author order is alphabetized.

Researched and written by Erik Henningsmoen (Senior Research and Policy Analyst), Heather McGeer (Senior Research and Policy Analyst), Mairead Matthews, and Noah Lubendo (Research and Policy Analyst) with generous support from Namir Anani and Faun Rice and the ICTC Research & Policy team.

TABLE OF CONTENTS

Executive Summary 4

Background 6

Section I: Access to Compute 8

- How Canadian Industry Currently Accesses Compute 8
- The Most Pressing Issues Canadian Industry Faces When Accessing AI Compute Today 10
- Foreign Ownership and Environmental Sustainability of AI Compute Infrastructure 11
- Required Levels and Types of AI Compute Access Required by Industry to Scale and Existing Government Support Programs 12

Section II: Developing Sovereign & Sustainable Compute Capacity 14

- The Benefits Developing Canadian-Owned and Controlled Advanced Computing Infrastructure 15
- Short-Term Opportunities and Initiatives to Help Expand Access to Existing Computing Infrastructure in Canada 17
- Opportunities to Incorporate Canadian-Made Computing Hardware and Software into a Sovereign AI Compute Infrastructure 18
- Collaboration Models to Support Industry-Academic Partnerships 18
- Case Study: United States National AI Research Resource 19
- Developing Relationships and Collaborating with Likeminded Countries 19

Section III: Priorities for Longer-Term Compute Infrastructure 20

- Priority Elements to Develop National AI Compute Infrastructure for Canada in the Next Five Years 20
- Recommendations 23

Conclusion 25

EXECUTIVE SUMMARY

Artificial intelligence (AI) technology depends on extensive advanced computing infrastructure, including data centres equipped with high-performance cloud computing platforms and specialized, often expensive and scarce, AI chips. It also requires a range of enabling resources, including technical resources such as AI frameworks, libraries, models, and training data; human capital, such as skilled technologists and technicians, researchers and scientists, and engineers to develop, test, deploy, and operate AI technology; and supporting infrastructure, such as high-speed telecommunications networks.

A national economy's capacity to foster and host such AI technology and its enabling resources domestically, free of significant foreign ownership and control, comprises a country's sovereign AI compute capability. The rapid and astonishing development and adaptation of AI technology in the past few years has attracted increasing attention regarding the issue of developing Canadian sovereign AI compute. Economies that can effectively develop and adopt leading AI technology, and absorb AI as part of its economy, research and innovation ecosystem, education and training system, and public sector, are poised to enjoy a significant economic and technological advantage in an increasingly competitive and uncertain global economic order.

For Canada, the conversation on the development of a national sovereign AI compute strategy included a federal public consultation to shape its future policy approach to AI compute, with feedback culminating in the publication of the Government of Canada's Canadian Sovereign AI Compute Strategy and earmarking $2 billion in funding for sovereign AI compute in December 2024. In 2025, the federal government appointed Canada's inaugural minister of AI and digital innovation. It also provided international leadership on AI governance issues at the 51st G7 Summit held in Kananaskis, Alberta in June 2025.

Peer economies, such as France, the European Union, and United States, are also developing their own sovereign AI compute capacities with haste and at a significant scale. In the American case, the Stargate Project will direct USD $500 billion in investments into AI technology, including AI compute infrastructure, over the next four years.

Despite years of research and policy leadership in AI, Canada is at risk of falling behind in developing a sovereign AI compute capability necessary to define its own economic and technological future. Canada trails peer economies in its trajectory of investing in and building out sovereign AI compute capacity.

The consequences of failing to foster this capability are significant. Without sovereign AI computing infrastructure, Canadian researchers, innovators, entrepreneurs, businesses, and public sector institutions will remain dependent on access to foreign AI computing infrastructure. This will put the country's digital future in the hands of international providers and subject to the policy decisions of foreign governments. In a world of growing geopolitical tensions, ongoing diplomatic and trade disputes, and increasing uncertainty regarding global cooperation, continued access to foreign-controlled AI computing services to underpin Canada's AI ecosystem is no longer a given.

Currently, Canadian businesses access AI compute through a mix of personal workstations, internal data centres, cloud services, and vendor-provided tools. While cloud computing is growing due to flexibility and lower upfront costs, access remains limited by high costs, hardware shortages, lack of talent, and inadequate broadband in rural and remote areas.

Canada's limited domestic AI compute infrastructure also poses national challenges, including concerns around data sovereignty, intellectual property (IP) protection, and long-term access to critical computing resources.

Developing Canadian-owned, sustainable compute capacity can strengthen national security, support economic growth, reduce environmental impact, and help to retain and grow domestic AI talent. However, to ensure businesses prioritize Canadian ownership when making infrastructure decisions, public investment, incentives, and collaborative models that support industry-academic partnerships and align with global peers will be needed to strengthen Canada's AI innovation ecosystem. In addition, strategic policy intervention is needed to close critical gaps in the AI development pipeline, including AI hardware, models, training data, and talent. Canada's recent commitment to NATO defence spending targets is a potential opportunity to help secure Canada's sovereignty and global competitiveness in AI by scaling up its domestic AI compute infrastructure.

This policy brief recommends that this effort be guided by six key principles: a robust industrial strategy, intellectual property ownership and retention, supply chain resiliency, environmental sustainability, regional equity, and targeted tax and financial incentives. Strategic investment in Canadian-owned infrastructure—including support for domestic chip manufacturing, AI clusters, and high-performance computing—will help retain IP, reduce reliance on foreign platforms, and preserve Canada's position in critical AI sectors. Enhancing supply chain resiliency and prioritizing sustainability will mitigate geopolitical and environmental risks, while broader regional access and public-private co-investment mechanisms will ensure inclusive growth across Canada.

The policy brief is based on a September 2024 policy submission ICTC developed for Innovation, Science and Economic Development Canada's Fall 2024 public consultations on AI compute.[1] Based on ICTC's original policy submission document, the brief has been updated and expanded, accounting for new developments in public policy related to AI compute, advances in AI and digital technology, and changes to market conditions.

1 See: Innovation, Science and Economic Development Canada (Government of Canada), "Consultations on Artificial Intelligence (AI) Compute," last update December 5, 2024, https://ised-isde.canada.ca/site/ised/en/public-consultations/consultations-artificial-intelligence-ai-compute.

BACKGROUND

Canadian innovators, entrepreneurs, and researchers need significant data storage and high-end computing power to build, train, test, and deploy Artificial Intelligence (AI) models.[2] The AI Now Institute warns that scarcity of AI compute currently forms a bottleneck in training and deploying Large Language Models (LLMs), giving a substantial advantage in developing and operating AI technology to those with sufficient access to advanced computing resources.[3]

Compared to other advanced economies, Canada lags in physical infrastructure such as servers and high-performance processors essential for AI compute.[4] A 2023 study by the Tony Blair Institute notes that alongside Italy, "Canada trails all other G7 nations on compute investment."[5] Despite having some of the strongest overall AI capability in the world, driven by leadership in AI policy, strong research, and talent development, Canada may struggle to fully leverage these strengths due to a lack of national AI compute resources.[6]

These challenges come at a time of geopolitical uncertainty, rapid technological and economic change, and immense pressures on the rules-based international order. Amid uncertain long-term alliances and trade relationships, and threats to Canada's national sovereignty, developing a domestic and sovereign AI compute capability will be essential in the coming years for Canada to pursue the technological, economic, and scientific benefits of AI technology.

Acknowledging these challenges, in June 2024 the Government of Canada conducted a public consultation to shape its future policy approach to AI compute.[7] The consultations informed the publication of the Canadian Sovereign AI Compute Strategy in December 2024,[8] as well as the announcement of an AI Sovereign Compute Infrastructure Program and $300 million AI Compute Access Fund in early 2025.[9] Additionally, in November 2024, the Government of Canada launched the Canadian Artificial Intelligence Safety Institute.[10] These policy actions regarding AI compute build on forward looking government policy related to AI research and adoption over the past decade, most notably the 2017 Pan-Canadian Artificial Intelligence Strategy.[11]

Further, the federal government has acknowledged the critical nature of the present moment, particularly as it relates to investments into AI development. In May 2025, the newly elected Liberal government released a mandate letter, which outlined the federal government's key goals and priorities.[12] In the letter, Prime Minister Mark Carney stressed the importance of increased investments into AI infrastructure to boost productivity and

2 "What are large language models (LLMs)?," IBM, November 2, 2023, https://www.ibm.com/topics/large-language-models.
3 Jai Vipra and Sarah Myers West, "Computational Power and AI," AI Now Institute, September 2023, https://ainowinstitute.org/publication/policy/compute-and-ai.
4 Graham Dobbs and Jake Hirsch-Allen, "Can Canada Compute? Policy Options to Close Canada's AI Compute Gap," The Dais, Toronto Metropolitan University, March 2024, https://dais.ca/reports/can-canada-compute/, 5
5 Bridget Boakye, et al., "State of Compute Access: How to Bridge the New Digital Divide," Tony Blair Institute for Global Change, November 2023, https://www.institute.global/insights/tech-and-digitalisation/state-of-compute-access-how-to-bridge-the-new-digital-divide.
6 Ibid.
7 Innovation, Science and Economic Development Canada (Government of Canada), "Consultations on Artificial Intelligence (AI) Compute," last update December 5, 2024, https://ised-isde.canada.ca/site/ised/en/public-consultations/consultations-artificial-intelligence-ai-compute.
8 Innovation, Science and Economic Development Canada (Government of Canada), "Canadian Sovereign AI Compute Strategy," last update March 7, 2025, https://ised-isde.canada.ca/site/ised/en/canadian-sovereign-ai-compute-strategy.
9 Innovation, Science and Economic Development Canada (Government of Canada), "AI Sovereign Compute Infrastructure Program," last update February 27, 2025, https://ised-isde.canada.ca/site/ised/en/ai-sovereign-compute-infrastructure-program; Innovation, Science and Economic Development Canada (Government of Canada), "AI Compute Access Fund," last update March 14, 2025, https://ised-isde.canada.ca/site/ised/en/canadian-sovereign-ai-compute-strategy/ai-compute-access-fund.
10 Innovation, Science and Economic Development Canada (Government of Canada), "Canada launches Canadian Artificial Intelligence Safety Institute (press release)," November 12, 2024, https://www.canada.ca/en/innovation-science-economic-development/news/2024/11/canada-launches-canadian-artificial-intelligence-safety-institute.html.
11 Organization for Economic Cooperation and Development (OECD), "Pan-Canadian artificial intelligence strategy," OECD Digital Policy Platform, accessed May 29, 2025, https://depp.oecd.org/policies/CAN1230.
12 Office of the Prime Minister, "Mandate Letter," May 21, 2025, https://www.pm.gc.ca/en/mandate-letters/2025/05/21/mandate-letter.

technology adoption.[13] In addition to the mandate letter, Carney also appointed Canada's first minister of AI and digital innovation in May 2025.[14]

Canada has also continued to position itself on the world stage, specifically as it relates to fostering international collaboration in AI. In June 2025, Canada, which held the G7 Presidency for 2025, announced that it is launching the G7 Gov AI Grand Challenge to develop solutions to address AI adoption barriers, and announced plans to establish the G7 AI Network (GAIN).[15] As well, as part of the G7 AI Adoption Roadmap, commitments were made to support access to compute and digital infrastructure.[16]

These developments in Canadian AI compute capacity come at a time when other advanced economies are directing large public and private investments towards AI compute infrastructure. For example, in January 2025, the United States announced the Stargate Project. Through a partnership between SoftBank, OpenAI, Oracle, MGX, Arm, Microsoft, and NVIDIA, this USD $500 billion initiative aims to "secure American leadership in AI"[17] by building up to 20 new powerful AI data centres and supporting electrical infrastructure across the United States.[18] In July 2025, the United States unveiled a national AI action plan.[19]

Likewise, in February 2025, France announced €109 billion in pledged investments in AI infrastructure in the leadup to the AI Action Summit.[20] During the summit, the European Union announced the InvestAI initiative to "mobilise" €200 billion in AI investments, including a €20 billion to build four "AI gigafactories," consisting of largescale data centres for use in training AI models.[21] In January 2025, the United Kingdom published an AI Opportunities Action Plan, which seeks to expand its AI Research Resource, a pool of computing resources used in breakthrough AI research twentyfold by 2030.[22]

These significant national investments in peer economy AI compute capacities come at a time of heightened geopolitical tensions, a degrading international security situation, and concerns over national sovereignty in the digital realm. National sovereignty concerns include maintaining domestic control of advanced computing infrastructure, preserving digital sovereignty, and protecting and fostering national innovation pipelines. Furthermore, Canada's commitments to increased defense spending create an opportunity to invest in AI infrastructure with spillover benefits.

This policy brief outlines key considerations and policy recommendations for developing a national sovereign AI and advanced compute infrastructure for Canada. It includes an assessment of how Canadian industry currently accesses cloud and advanced computing resources; pressing issues faced by Canadian industry and new approaches being adopted to access AI compute; required supports needed by Canadian entrepreneurs, innovators, and academic researchers to access AI compute, including industry-academic collaboration models; cross-border risks and international collaboration opportunities; and opportunities and priority components for developing a sovereign AI compute capability for Canada in next five years.

13 Ibid.

14 Office of the Prime Minister, "Prime Minister Carney announces new Ministry," May 13, 2025, https://www.pm.gc.ca/en/news/news-releases/2025/05/13/prime-minister-carney-announces-new-ministry.

15 G7 2025 Kananaskis, "G7 Leaders' Statement on AI for Prosperity," June 17, 2025 https://g7.canada.ca/en/news-and-media/news/g7-leaders-statement-on-ai-for-prosperity/.

16 Ibid.

17 OpenAI, "Announcing the Stargate Project," January 21, 2025, https://openai.com/index/announcing-the-stargate-project/.

18 "Trump highlights partnership investing $500 billion in AI," Associated Press, January 22, 2025, https://apnews.com/article/trump-ai-openai-oracle-softbank-son-altman-ellison-be261f8a8ee07a0623d4170397348c41.

19 The White House (United States), White House Unveils America's AI Action Plan, July 23, 2025, https://www.whitehouse.gov/articles/2025/07/white-house-unveils-americas-ai-action-plan/.

20 "With announcement of investments worth €109 billion, Macron intends to take on US," Le Monde, February 10, 2025, https://www.lemonde.fr/en/economy/article/2025/02/10/ai-with-the-announcement-of-a-109-billion-investment-macron-intends-to-take-on-the-us_6737985_19.html.

21 European Commission (European Union), "EU launches InvestAI initiative to mobilise €200 billion of investment in artificial intelligence (press release)," February 11, 2025, https://luxembourg.representation.ec.europa.eu/actualites-et-evenements/actualites/eu-launches-investai-initiative-mobilise-eu200-billion-investment-artificial-intelligence-2025-02-11_en.

22 His Majesty's Government (United Kingdom), "AI Opportunities Action Plan," January 13, 2025, https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan.

SECTION I:

ACCESS TO COMPUTE

ICTC regularly consults with Canadian industry about their information and communication technology (ICT) infrastructure, such as for whether they acquire, build, and run their own data centres internally, or opt for external cloud services providers; how they choose between different cloud service providers; and how they decide what types of hardware, such graphics processing units (GPUs), to use. Relevant information from ICTC's engagement with industry is summarized below.

HOW CANADIAN INDUSTRY CURRENTLY ACCESSES COMPUTE

Canadian startups and small and medium-sized enterprises (SMEs) utilize a wide range of AI compute infrastructure to build, train, test, and deploy AI models, including:

- **Personal workstations, such as personal computers or machines**
- **Centralized computing infrastructure, including internal data centres or external cloud service providers**
- **AI frameworks and libraries**
- **AI models**

While some resources, such as AI frameworks, libraries, models, and training data, are open source and freely available, others are proprietary and require financial investment. In addition to ICT infrastructure, industry needs access to talent with the skills and knowledge necessary to integrate AI resources into products, services, and industry applications. Common AI development roles particularly in-demand by Canadian companies include AI engineers, AI researchers, computer vision engineers, data engineers, data scientists, machine learning engineers, and software developers.[23]

As of 2021, approximately half of Canadian businesses accessed centralized computing infrastructure through an internal, company-wide computer network, while a similar proportion used cloud computing.[24] Increasingly, Canadian businesses are choosing to outsource some or all their centralized computing infrastructure to cloud services providers, and the proportion using cloud infrastructure is projected to continue growing.[25] Cloud services require a smaller upfront investment, reduce the need for extensive internal talent and skills, and allow startups and SMEs to scale their ICT infrastructure dynamically as their computational needs fluctuate. It is also common for cloud services providers to offer startups and SMEs free "computing credits" to lower initial costs and encourage customer loyalty. The long-term cost-effectiveness of outsourced cloud services versus developing firm-owned, internal data centres depends on factors such as a company's overall demand for computational resources, the type of resources needed, and the stability and predictability of their computational workloads.

Industry members also access AI compute capacity through their AI product and service providers. For example, organizations that use ChatGPT via its web interface or API (application programming interface) outsource not only generative AI but also the required computational resources. Similarly, companies whose developers use tools like GitHub CoPilot or Cursor, which incorporate LLMs, access AI compute resources through their vendors.

23 As identified in the recent ICTC report: Todd Legere, Sheldon Lopez, and Noah Lubendo, "Canada's AI Ecosystem: A Brief Overview of In Demand Skills and Trends," Information and Communications Technology Council (ICTC), February 2025, https://ictc-ctic.ca/reports/canadas-ai-ecosystem-a-brief-overview-of-indemand-skills-and-trends/.

24 In 2021, 45.3% of businesses surveyed by Statistics Canada used cloud computing while 53% used a company-wide computer network. See: "Information and communication technologies used by industry and size of enterprises," September 2022, Statistics Canada, https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2210011701.

25 "Canadian organizations are increasingly moving away from their traditional data centre deployments to run their compute and storage capacity across dedicate private and public clouds." See: "Operating a Digital Business in a Hybrid Cloud World," IDC Canada, https://74388.fs1.hubspotusercontent-na1.net/hubfs/74388/2023%20Cloud%20Report%20-%20En.pdf; "93% of Canadian respondents said they're advanced in their adoption of cloud technology" while "59% of respondents say that 40% or more of their enterprise workloads are now in the cloud." See: "How to go from migration to modernization by building a strong cloud and automation strategy," May 2023, KPMG, https://kpmg.com/ca/en/home/insights/2023/05/building-a-better-cloud.html; From 2019 to 2021, the percentage of Canadian businesses that use cloud computing grew by 16.5%; the percentage that use a company-wide computer network meanwhile increased 4.5%. See: "Information and communication technologies used by industry and size of enterprises," September 2022, Statistics Canada, https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=2210011701; The IDC estimates that non-cloud infrastructure accounted for 34.9% of infrastructure in 2023 and projects that this will decline to 26.4% by 2028. Shared and dedicated cloud was meanwhile estimated to account for 65.1% of infrastructure in 2023, a figure that is projected to reach 73.6% by 2028. See: "Spending on Shared Cloud Infrastructure Continues to Lead the Way in Enterprise Infrastructure Investments, According to IDC Tracker," March 2024, IDC, https://www.idc.com/getdoc.jsp?containerId=prUS52001524.

THE MOST PRESSING ISSUES CANADIAN INDUSTRY FACES WHEN ACCESSING AI COMPUTE TODAY

The most pressing issues that Canadian industry faces in accessing AI compute within the ecosystem currently are technical requirements, cost, geographic location, talent availability, and hardware availability.

Technical Requirements:

Technical requirements are the primary consideration for Canadian industry when making ICT infrastructure decisions. For example, organizations subject to data residency requirements must ensure that data storage and processing occur within Canada. Organizations operating in critical infrastructure sectors must build redundancy into their ICT infrastructure by duplicating data processing and storage across multiple vendors and geographic locations. Similarly, companies integrating AI models into their products and services must use sufficiently advanced computing hardware, such as GPUs, to run their applications efficiently.

Cost:

Cost is a significant factor for Canadian industries when making ICT infrastructure decisions, such as whether to build internal data centres or outsource centralized computing infrastructure to cloud service providers; which cloud service provider to select; and what type of GPU to use for training, testing, and deploying AI models. Cost is often weighed alongside technical requirements. For example, an AI startup is likely to choose the least expensive GPU that can still run their application in a reasonable time, depending on user needs. Regardless of the method of access, the costs associated with AI compute can be substantial. Smaller organizations, such as startups and university researchers with limited financial resources, may be excluded from the AI innovation ecosystem due to lack of access to AI compute, despite having promising ideas and technical expertise.

Geographic Location:

Geographic location of computing infrastructure is a common consideration for companies operating in industries subject to data residency requirements, such as healthcare. For companies not bound by these requirements, geographic location is typically less critical. However, it may still be a factor due to practices like edge computing where data processing and storage are located near end users to reduce latency and improve bandwidth efficiency, or for reasons related to environmental sustainability, such as placing infrastructure in regions that utilize clean energy.

Talent Availability:

Talent availability is a common constraint on Canadian organizations' ability to access AI compute. Building, training, testing, deploying, and adopting AI models requires specialized skills. Industry feedback to ICTC often highlights difficulties in accessing the necessary talent for provisioning AI compute as well as for designing, constructing, and implementing AI solutions.

Hardware Availability:

Hardware availability can limit access to AI compute, especially for highly specialized or scarce computational resources. For example, ICTC recently interviewed a representative from a large Canadian technology company, who reported being placed on long waitlists when trying to rent H100 GPUs from their cloud service provider to train an AI model.

High-Speed Internet:

Utilizing remote AI infrastructure depends on high-bandwidth, low-latency networks, which are essential for rapid data transfer and processing in AI applications. Many rural, remote, and northern regions across Canada do not currently have high-speed broadband coverage.[26] Without access to high-speed internet, organizations are constrained in their ability to leverage remote AI infrastructure, such as that provided by cloud service providers, thereby limiting their capacity to develop, deploy, and adopt AI products and services.

FOREIGN OWNERSHIP AND ENVIRONMENTAL SUSTAINABILITY OF AI COMPUTE INFRASTRUCTURE

Whether computing infrastructure is Canadian-owned and controlled is generally not a primary consideration for Canadian industry when making decisions about their ICT infrastructure. Similarly, environmental sustainability is not commonly prioritized in ICT infrastructure procurement choices. When sustainability is considered, it typically focuses on energy consumption and carbon emissions rather than broader impacts like water use, natural resource use, or biodiversity.[27] Nevertheless, sustainability is an essential consideration for Canadian AI compute investments, as regional electricity source and renewability play a significant role in energy expenditure by data centre location.[28]

26 Download speeds of 50 megabits per second (Mbps) and upload speeds of 10 Mbps are necessary to support cloud-based software applications. In 2017, only 37% of rural households had access to 50/10 Mbps, compared with 97% of urban homes. See: "High-Speed Access for All: Canada's Connectivity Strategy," Innovation, Science and Economic Development Canada, March 2025, https://ised-isde.canada.ca/site/high-speed-internet-canada/en/canadas-connectivity-strategy/high-speed-access-all-canadas-connectivity-strategy.

27 Kaitlyn Carr, Allison Clarke, and Mairead Matthews, "Building a Sustainable ICT Ecosystem: Strategies and Best Practices for Reducing Environmental Harms in a Digital World," Information and Communications Technology Council (ICTC), January 2024, https://ictc-ctic.ca/reports/building-a-sustainable-ict-ecosystem.

28 Victor Schmidt et al., "Machine Learning CO2 Impact," accessed July 3, 2025, https://mlco2.github.io/impact/.

REQUIRED LEVELS AND TYPES OF AI COMPUTE ACCESS REQUIRED BY INDUSTRY TO SCALE AND EXISTING GOVERNMENT SUPPORT PROGRAMS

Canadian organizations that adopt off-the-shelf AI products and services can often access AI compute through their AI product and service providers. However, if they plan to train, fine-tune, and distill off-the-shelf AI models internally, they will need access to their own AI compute infrastructure.

Organizations that are building their own AI products and services internally, or in collaboration with a vendor, require a broad range of AI compute infrastructure to build, train, test, and deploy AI models. This includes personal workstations, such as computers or machines; centralized computing infrastructure with advanced hardware; AI frameworks and libraries; AI models; and training data. In addition to infrastructure, the industry needs access to skilled talent capable of integrating these resources into products, services, and industry applications.

Table 1 provides an overview of existing federal programs that support AI inputs and could be leveraged to improve access to AI inputs in the short term.

Federal Program	Personal Work-stations	Centralized Compute (On-Prem)	Centralized Compute (Cloud)	Advanced Compute R&D	AI Models	Data-sets	AI Talent
AI Compute Access Fund			✓				
AI for Logistics					✓	✓	✓
AI Sovereign Compute Infrastructure Program				✓			
Alliance Cloud Connect Pilot			✓				
Applied Quantum Computing Challenge				✓	✓		
Artificial Intelligence for Design Challenge					✓	✓	✓
Beaufort Sea Engineering Database Project						✓	
Business Expenses	✓	✓	✓		✓	✓	✓
Canada Research Chairs	✓	✓	✓		✓	✓	✓
Canada Research Data Centre Network						✓	
CanCode							✓
CIHR, NSERC, and SSHRCC Grants	✓	✓	✓		✓	✓	✓
Community Cloud Renewal			✓				
Computers for School Plus	✓						
Data Centre Upgrade		✓	✓				
Digital Research Infrastructure Strategy		✓	✓				

Federal Program	Personal Work-stations	Centralized Compute (On-Prem)	Centralized Compute (Cloud)	Advanced Compute R&D	AI Models	Data-sets	AI Talent
Digital Skills for Youth							✓
High-Performance Computing Renewal		✓					
National Research Council Research Programs				✓	✓	✓	✓
NRC Industrial Research Assistance Program							✓
NRC IRAP Youth Employment Program							✓
NSERC Student Awards							✓
Open Government Portal						✓	
Post-Secondary Institutions Strategic Investments Fund		✓					
Quantum and Nanotechnologies Research Centre				✓	✓		
Research Support Fund	✓	✓	✓	✓	✓	✓	✓
scale ai					✓	✓	✓
Student Work Placement Program such as ICTC's WIL Digital program							✓
Supporting Researchers Using Quantum Computers program		✓					
Upskilling for Industry Initiative such as ICTC's Beyond the Cloud program							✓

Table 1. Existing federal programs that provide support for AI inputs and could be leveraged to improve access to AI inputs in the short term. ICTC analysis based on public information.

SECTION II:

DEVELOPING SOVEREIGN & SUSTAINABLE COMPUTE CAPACITY

While questions regarding domestic ownership and control is not a common consideration for individual businesses when making ICT infrastructure decisions, developing sovereign and sustainable compute capacity can be viewed as a matter of economic security and national resiliency.

Training large AI models often requires access to vast computational resources that are typically hosted by foreign cloud providers (e.g., AWS, Google Cloud, Microsoft Azure). This means that sensitive Canadian data might flow through or reside in jurisdictions outside of Canada, exposing it to foreign laws and potential surveillance. When models are trained or fine-tuned using foreign compute resources, there is a risk that the models, their performance data, or even their architectures could be accessible to third-party service providers, undermining Canadian IP ownership and privacy.

Although Canada is a leader in fundamental AI research and talent production, it lags in domestic AI compute. According to the OECD.AI Policy Observatory, "Canada's research and enterprise-ready computational capabilities are lagging not just in comparison to its southern neighbour, the United States, but to all other G7 Nations."[29] The Tortoise's 2024 Global AI Index finds that, while Canada ranks eighth overall in the global AI ecosystem, it falls behind in infrastructure compared to AI leaders like the United States, China, Singapore, Japan, and South Korea. [30]

29 Graham Dobbs and Jake Hirsch-Allen, "Canada's plans to bridge the AI compute gap and how it can make industry policy inclusive and sustainable," OECD.AI Policy Observatory, April 16, 2024, https://oecd.ai/en/wonk/canadas-ai-compute-gap.

30 Tortoise, "Global AI Index," 2024, https://www.tortoisemedia.com/intelligence/global-ai/.

THE BENEFITS DEVELOPING CANADIAN-OWNED AND CONTROLLED ADVANCED COMPUTING INFRASTRUCTURE

Despite Canadian ownership and domestic control not often being common consideration for individual stakeholders in ICT infrastructure decisions, it can offer several national-level benefits:

Economic Benefits:

Global spending on ICT infrastructure is growing rapidly with cloud services alone projected to reach $1.35 trillion by 2027, driven by high-performance computing and AI.[31] Countries that own and control compute and storage infrastructure can benefit from increased revenues, GDP growth, and new job creation. In 2023, Canada's computing market generated approximately US $3.8 billion or US $96 per capita, compared to the US market's US $39.1 billion or US $114 per capita.[32] Expanding Canadian-owned infrastructure and using Canadian hardware and software in Canada's national infrastructure could help balance these figures and create economic opportunities for Canadian computing companies.

National Security and Public Safety:

The ICT sector is one of Canada's ten critical infrastructure sectors, vital to the health, safety, security, and economic well-being of Canadians and the effective functioning of government.[33] Computing infrastructure is central to the ICT sector and is critical, alongside other infrastructures like utilities, transportation, food, and finance. Canadian-owned and controlled computing infrastructure strengthens Canada's ability to operate independently, reduce dependency on infrastructure owned by foreign actors, secure data, and develop sensitive AI applications within a sovereign environment.

Environmental Sustainability:

ICT infrastructure has a broad range of environmental impacts throughout its lifecycle, including raw material extraction, energy use, emissions, water consumption, waste creation, land use change, and pollution.[34] Developing Canadian-owned infrastructure offers the opportunity to build more sustainable compute and storage facilities by leveraging clean energy sources and adhering to best practices, such as the International Telecommunications Union's L.1300: Best Practices for Green Data Centres.[35]

31 "Worldwide Spending on Public Cloud Services is Forecast to Reach $1.35 Trillion in 2027, According to New IDC Spending Guide," August 2023, IDC, https://www.idc.com/getdoc.jsp?containerId=prUS51179523.

32 "Revenue of the computing market worldwide by country in 2023," 2024 Statista, https://www.statista.com/forecasts/758685/revenue-of-the-computing-market-worldwide-by-country.

33 "National Strategy for Critical Infrastructure," Government of Canada, 2009, https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/srtg-crtcl-nfrstrctr/srtg-crtcl-nfrstrctr-eng.pdf.

34 Kaitlyn Carr, Allison Clarke, and Mairead Matthews, "Building a Sustainable ICT Ecosystem: Strategies and Best Practices for Reducing Environmental Harms in a Digital World," Information and Communications Technology Council (ICTC), January 2024, https://ictc-ctic.ca/reports/building-a-sustainable-ict-ecosystem.

35 "L.1300: Best practices for green data centres," ITU, June 2014, https://www.itu.int/rec/T-REC-L.1300-201406-I/en.

Digital Inclusion: Should Canadian-owned computing infrastructure be more cost-effective or subsidized, it could increase access for individuals and organizations currently excluded due to cost constraints.

Foster Domestic Innovation Pipeline: When AI compute resources are entirely concentrated abroad, Canadian AI talent may relocate to countries and companies with better access to computing resources. Losing highly qualified AI talent—including technologists, researchers, and entrepreneurs—amounts to brain drain and undermines Canada's AI innovation pipeline. Developing requisite AI compute resources within Canada can foster the country's domestic innovation pipeline.

Reliability of Compute Access and Geopolitical Risk Mitigation: Ongoing global trade disputes and geopolitical uncertainty suggest that reliable cross-border access to cloud, AI compute, and other digital infrastructure may no longer be a given going forward. Financial services provider BlackRock notes that "global technology decoupling" and "global trade protectionism" are both expected to be drivers of geopolitical risk through 2025.[36] Domestically owned and controlled advanced computing infrastructure can help mitigate risks of limited or restricted access to computing infrastructure—such as blocked access to cloud services or export controls on hardware like GPUs—across borders during future trade disputes and other geopolitical crises. This would reduce national economic vulnerability and mitigate the potential for economic cohesion should such situations arise.

Despite these benefits, Canadian startups and SMEs tend to prioritize technical requirements, cost, geographic location, and availability over Canadian ownership when choosing ICT infrastructure. If the national benefits of Canadian-owned infrastructure are deemed important for national sovereignty purposes, public investment or other economic incentives may be needed to promote its use. For instance, Canada already has in place some statutory data residency requirements in RFPs.[37] Although data residency requirements in Canada are not sweeping, and are most applicable to public sector activities, they suggest that targeted incentives or mandates may be necessary to encourage broader adoption of Canadian-based solutions.

36 BlackRock, "Geopolitical risk dashboard," accessed May 23, 2025, https://www.blackrock.com/corporate/insights/blackrock-investment-institute/interactive-charts/geopolitical-risk-dashboard.

37 Treasury Board of Canada Secretariat, "Government of Canada White Paper: Data Sovereignty and Public Cloud" July 28, 2020, https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/cloud-services/gc-white-paper-data-sovereignty-public-cloud.html.

SHORT-TERM OPPORTUNITIES AND INITIATIVES TO HELP EXPAND ACCESS TO EXISTING COMPUTING INFRASTRUCTURE IN CANADA

While investing in AI compute access for Canadian researchers and businesses is a long-term and ongoing endeavor, short-term initiatives can also shape Canada's AI compute ecosystem:

Support for Internal Data Centres:

Some organizations may prefer building internal data centres over outsourcing to cloud providers due to technical requirements, cost benefits, or other considerations. For example, organizations with stable and predictable AI workloads might find it more cost-effective to build internal data centres rather than renting specialized GPUs. Supporting these organizations could expand Canada's computing infrastructure in the long term.

Incentivizing Use of Canadian-Owned Infrastructure:

Since Canadian ownership is not typically a factor in ICT decisions, organizations may not choose Canadian compute options even if widely available. To ensure sustainable long-term AI compute infrastructure, there must be specific demand for Canadian-owned facilities. If this demand does not arise organically, incentives may be necessary, such as cost benefits or environmental advantages. Environmental advantages could include access to low-carbon energy sources like hydroelectricity, reduced emissions, or higher environmental standards, making Canadian facilities more attractive to organizations with sustainability goals. While sustainability is not yet a primary driver in most ICT procurement decisions, environmental advantages could become more influential if tied to regulatory requirements or ESG reporting obligations.[38]

38 Kaitlyn Carr, Allison Clark, and Mairead Matthews, "Building a Sustainable ICT Ecosystem: Strategies and Best Practices for Reducing Environmental Harms in a Digital World," ICTC, January 2024, https://ictc-ctic.ca/reports/building-a-sustainable-ict-ecosystem.

OPPORTUNITIES TO INCORPORATE CANADIAN-MADE COMPUTING HARDWARE AND SOFTWARE INTO A SOVEREIGN AI COMPUTE INFRASTRUCTURE

Developing national AI compute infrastructure presents an opportunity to leverage industrial strategy. Canada should adopt an industrial strategy lens while designing further AI compute policy by incentivizing the use of Canadian-owned and controlled infrastructure and creating procurement opportunities for Canadian suppliers. For example, ongoing research by ICTC highlights that Canadian semiconductor firms have developed strong research and development capabilities in next-generation semiconductor technology relevant to producing specialized chips for advanced computing applications, such as AI and quantum computing.[39]

COLLABORATION MODELS TO SUPPORT INDUSTRY-ACADEMIC PARTNERSHIPS

Industry is outpacing academic institutions in basic AI research partly due to disparities in compute capacity.[40] If academic institutions continue to face barriers to accessing AI compute, there is a risk that private actors will increasingly dominate the AI research agenda. Public–private research alliances can bridge this gap, allowing academic researchers to access necessary AI compute while providing industry with specialized research expertise. Such collaborative models can create "shared value"[41] opportunities for academic institutions, industry, and not-for-profits, with each partner contributing their individual strengths.[42] To foster AI compute capacity domestically, public-private research alliances should prioritize cooperation between domestic private sector partners and Canadian academic institutions.

Collaborative models could involve partnerships with major cloud service providers. Alternatively, private sector organizations could provision their own AI compute resources for collaborative public–private projects or provide financial resources to purchase new AI compute capacity. Academic institutions with specialized computing power, like the Digital Research Alliance of Canada's supercomputing National Host Sites, can also contribute to joint academic–industry projects, especially for early-stage AI startups needing access to powerful computing resources to develop proofs of concept for new AI technologies to attract investment.[43]

39 See: "ICTC Partners with Georgetown's CSET to Map Semiconductor Industries in Canada and the United States (press release)," Information and Communications Technology Council (ICTC), November 14, 2024, https://ictc-ctic.ca/news-and-events/news-articles/ictc-partners-georgetowns-cset-map-semiconductor-industries-canada.

40 Brian Eastwood, "Study: Industry now dominates AI research," MIT Sloan School of Management, May 18, 2023, https://mitsloan.mit.edu/ideas-made-to-matter/study-industry-now-dominates-ai-research.

41 Michael E. Porter and Mark R. Kramer, "Creating Shared Value," Harvard Business Review, January-February, 2011, https://hbr.org/2011/01/the-big-idea-creating-shared-value.

42 Mark R. Kramer and Marc W. Pfitzer, "The Ecosystem of Shared Value," Harvard Business Review, October 2016, https://hbr.org/2016/10/the-ecosystem-of-shared-value.

43 "National Host Sites," Digital Research Alliance of Canada, accessed August 16, 2024, https://alliancecan.ca/en/services/advanced-research-computing/federation/national-host-sites.

Case Study: United States National AI Research Resource

The United States National AI Research Resource (NAIRR) is a pilot program by the US National Science Foundation that provides academic researchers and educators access to advanced computing, datasets, models, software, and support for AI applications.[44] The program partners with 10 U.S. federal agencies and 25 private companies, including Nvidia, Microsoft, Meta, and OpenAI, to develop the US national AI capacity, establish research and development partnerships between industry, academic institutions, and civil society, and support open, secure, and classroom-focused AI research.[45] It includes program streams for open AI research (NAIRR Open), secured AI research with privacy safeguards (NAIRR Secure), AI software and interoperability tools (NAIRR Software), and AI in the classroom (NAIRR Classroom).[46]

The NAIRR has been popular among academic institutions, highlighting the significant demand for AI compute access. However, it also provides some useful lessons learned. A 2024 report by the Brookings Institution recommends creating distinct funding streams for smaller academic institutions and encouraging partnerships between smaller and larger entities to address imbalances in the resources these two groups can dedicate to proposal writing.[47] Program funders should be aware of this resource imbalance and design calls for proposals accordingly.

Canadian universities have also created innovation, commercialization, and business incubation ecosystems, such as the MaRS Discovery District in Toronto and the Creative Destruction Lab,[48] to help student and faculty entrepreneurs bring new products and technology to market. Facilitating access to AI compute through private providers, or a national public AI compute program, may be a valuable service these ecosystems could offer in the future.

DEVELOPING RELATIONSHIPS AND COLLABORATING WITH LIKE-MINDED COUNTRIES

Canada's international peers face similar challenges in AI compute, including ICT and AI supply chain concentration, barriers to accessing AI compute, and growing environmental impacts from ICT. One of Canada's largest trading partners recently adopted legislation that, for the first time, ties AI systems to their environmental sustainability impacts by requiring the providers of general-purpose AI models to calculate and disclose the energy consumption of their models and encouraging AI systems to be developed and used in an environmentally sustainable manner.[49]

Canadian sovereign AI compute could be leveraged in Canada's relationships with like-minded countries to help solve these challenges, providing a way for Canada's peers to diversify and increase the resiliency of their ICT supply chains and sectors, address barriers to accessing AI compute, and adopt ICT infrastructure that is aligned with standards for sustainable ICT. As emphasized in AI Sovereignty and Economic Growth, ICTC's joint report with the AI, Data and Robotics Association (Adra), deeper collaboration between Canada and the European Union could leverage complementary strengths in governance, research, regulation, and industrial applications. Such partnerships are vital for reinforcing both AI sovereignty and global competitiveness, while fostering responsible development pathways for enabling joint investment in compute hardware, sovereign cloud infrastructure, and clean energy innovation to meet the demands of large-scale AI systems sustainably.[50]

44 "Democratizing the future of AI R&D: NSF to launch National AI Research Resource pilot (press release)," U.S. National Science Foundation, January 24, 2024, https://new.nsf.gov/news/democratizing-future-ai-rd-nsf-launch-national-ai.

45 "The U.S. Just Took a Crucial Step Toward Democratizing AI Access," Time, January 26, 2024, https://time.com/6589134/nairr-ai-resource-access/.

46 U.S. National Science Foundation, 2024, op cit.

47 Jennifer Wang and Mark Muro, "How the National Artificial Intelligence Research Resource can pilot inclusive AI," Brookings Institution, July 9, 2024, https://www.brookings.edu/articles/how-the-national-artificial-intelligence-research-resource-can-pilot-inclusive-ai/.

48 "MaRS Discovery District," https://www.marsdd.com/; "Creative Destruction Lab (CDL)," https://creativedestructionlab.com/.

49 "REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL," Official Journal of the European Union, 2024, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689.

50 Amaya Garmendia, Todd Legere, and Noah Lubendo. "AI Sovereignty and Economic Growth: Strengthening Transatlantic Leadership Between the EU and Canada". AI, Data and Robotics Association (Adra) / Information and Communications Technology Council (ICTC), May 2025, https://ictc-ctic.ca/reports/ai-sovereignty-and-economic-growth.

SECTION III:

PRIORITIES FOR LONGER-TERM COMPUTE INFRASTRUCTURE

Investing in domestic compute infrastructure will require balancing priorities and addressing various stakeholder needs across the AI ecosystem. Understanding the unique requirements of different industry players and the research community will be essential to ensure that solutions have a lasting and positive impact.

PRIORITY ELEMENTS TO DEVELOP NATIONAL AI COMPUTE INFRASTRUCTURE FOR CANADA IN THE NEXT FIVE YEARS

Canada's strategy for national AI compute infrastructure should focus on inputs for AI that may be difficult for businesses, organizations, and individuals to access through other means, whether due to cost, availability, or other constraints. Table 2 identifies which AI inputs are open source and freely available versus those that are proprietary and come with financial costs. It suggests that while Canadian startups, companies, researchers, and students generally have good access to AI frameworks and libraries, the strategy should focus on strengthening the availability and accessibility of AI hardware, models, training data, and talent. These components are discussed in more detail below.

Input		Description	Licensing Model	Cost
AI Hardware	**Personal Workstations**	Personal workstations, such as personal computers, with powerful enough hardware to develop, test, and train AI models on a small scale	Proprietary	Paid access
	Centralized Computing Infrastructure	Private data centres or publicly accessible cloud computing platforms with high performance GPUs, ASICs, TPUs, or FPGAs for AI training and deployment for inference, etc.	Proprietary	Paid access
AI Frameworks and Libraries		Frameworks (e.g., TensorFlow, PyTorch, Keras) and libraries (e.g., Nvidia CUDA, Intel ONEAPI, Caffe) which are used to develop and deploy AI models	Open source	Freely available
AI Models		Pre-trained and custom-built models that can be used to accomplish different types of tasks	Open source and proprietary	Freely available and paid access
AI Training Data		Datasets used to train AI models, which may be either proprietary or open source depending on the source and use case	Open source and proprietary	Freely available and paid access
AI Talent		People with the skills needed to develop, test, train, and deploy AI models	NA	Paid access

Table 2. The characteristics of AI inputs.

Component I: AI Hardware

Organizations building and adopting AI solutions require access to sophisticated hardware that is powerful enough to develop, test, train, and deploy AI models. Personal machines must be capable of handling AI models locally and on a small scale before moving to production. Powerful, centralized computing infrastructure is needed to develop, test, train, and deploy AI models quickly enough at scale. While foreign-owned and controlled AI compute infrastructure meets the needs of many Canadian startups, companies, researchers, and students, it can be costly and inaccessible to some due to budget constraints. To ensure access to necessary hardware, Canada should explore ways to reduce barriers such as cost, availability, geography, and internet access. Efforts to diversify sourcing and build local capacity would help mitigate risks associated with global semiconductor bottlenecks and support investments in specialized Canadian AI hardware.

Canada should also prioritize and encourage the development of a local semiconductor industry. While it may not be viable to develop a broad semiconductor industry in the same way as has been done in Taiwan, South Korea, and the United States, Canada can strengthen capabilities in manufacturing for select hardware, such as AI-specific chips,[51] by focusing on the design and manufacturing of application-specific integrated circuits (ASICs) tailored to machine learning and deep learning tasks, as well as funding research and commercial efforts to develop Canadian AI accelerators or niche chip semiconductor technologies. Canada should also encourage partnerships with global leaders in AI chip design, such as Nvidia and AMD, which foster the production of high-performance chips for AI applications.

Component II: AI Models

AI models are the foundation of AI product and service development. ICT professionals have a wide array of AI models to choose from, capable of tasks such as event detection; data recognition, identification, and categorization; forecasting; personalization; system optimization; reasoning; and content generation, including text, audio, and visual outputs.[52]

AI models can be either open source, making them publicly available and free for use, or proprietary, requiring payment for integration into products and services. While some analysts view open-source models as risky due to reduced safeguards, others see benefits in their customizability, adaptability, and the ability to run in closed environments without sharing data.[53] Open-source models also allow researchers not affiliated with large tech firms to conduct independent AI research, including research on AI safety.[54]

The distribution of open source versus proprietary models varies across AI types. Until recently, most advanced large language models (LLMs) were proprietary, like OpenAI's GPT-4, Anthropic's Claude, and Google DeepMind's Gemini.[55] However, in July 2024, Meta AI released Llama 3.1 405B, the first open-source LLM rivaling closed models.[56] In January 2025, AI startup DeepSeek released DeepSeek-R1, a novel, open-source LLM that purports to use less training data and thus less computing power than other contemporary leading AI applications.[57]

To shape the availability of AI models for Canadian researchers and industry, the government should continue funding AI-focused R&D through organizations like the National Research Council of Canada, the Natural Sciences and Engineering Research Council of Canada, Canada Research Chairs, and the Canadian Global Innovation Cluster, Scale AI.[58]

---

51 See: Saif M. Khan and Alexander Mann, "AI Chips: What They Are and Why They Matter," Center for Security and Emerging Technology (CSET), April 2020, https://cset.georgetown.edu/publication/ai-chips-what-they-are-and-why-they-matter/.

52 "OECD Framework for the Classification of AI systems," OECD, February 2022, https://www.oecd.org/en/publications/oecd-framework-for-the-classification-of-ai-systems_cb6d9eca-en.html.

53 Will Knight, "Meta's New Llama 3.1 AI Model Is Free, Powerful, and Risky," WIRED, July 2024, https://www.wired.com/story/meta-ai-llama-3/; "Introducing Llama 3.1: Our most capable models to date," Meta, July 2024, https://ai.meta.com/blog/meta-llama-3-1/.

54 "Introducing Llama 3.1: Our most capable models to date," Meta, July 2024, https://ai.meta.com/blog/meta-llama-3-1/.

55 Ibid.

56 Ibid.

57 "What is DeepSeek? The Chinese OpenAI rival sparking chaos in tech markets," CBC News, January 27, 2025, https://www.cbc.ca/news/business/deepseek-ai-startup-1.7442382.

58 See: "Who we are," Scale AI, https://www.scaleai.ca/about-us/.

Component III: AI Training Data

Training data is essential for ICT professionals to finetune AI models for industry applications. Like AI models, training data can be open source or proprietary. Recent legal disputes between AI firms and proprietary data providers underscore the importance of accessible training data.[59] To support AI innovation, the government should continue developing accessible datasets through initiatives like the Open Government Portal, Canada Research Data Centre Network, and Beaufort Sea Engineering Database projects.

Component IV: AI Talent

Organizations building and adopting AI solutions need access to skilled talent with the knowledge and skills to turn AI inputs into new products, services, and industry applications. Canadian organizations frequently face difficulties in accessing the necessary skills to work with AI, whether designing, building, or adopting AI solutions. To address this, the government should continue to support programs focused on developing AI talent and skills.[60] Key skills required include:

- Building centralized computing infrastructure
- Provisioning hardware resources like GPUs from internal data centres and cloud providers
- Managing and engineering data (e.g., data collection, cleaning, processing, and preparation)
- Selecting AI models based on characteristics like quality, speed, and price
- Building, training, finetuning, and distilling AI models
- Address AI ethics, fairness, and bias mitigation to ensure responsible development
- Securing AI systems and ensuring data privacy
- Deploying AI models and monitoring performance
- Using AI frameworks and libraries to integrate AI into applications
- Collaborating across disciplines to incorporate domain expertise into AI solutions

Developing AI talent with compute expertise can be supported through the creation of specialized training programs at Canadian colleges and universities that are focused on AI infrastructure, systems engineering, and distributed computing. To mitigate the issue of brain drain and ensure Canadian AI researchers and engineers are not being pulled to compute-rich ecosystems located abroad, the government should provide competitive funding and tax incentives for businesses.

59 "Getty Images v. Stability AI," 2024, BakerHostetler, https://www.bakerlaw.com/getty-images-v-stability-ai/; Dylan Walsh, "The Legal Issues Presented by Generative AI," MIT Sloan School of Management, August 2023, https://mitsloan.mit.edu/ideas-made-to-matter/legal-issues-presented-generative-ai.

60 Maryna Ivus, Akshay Kotak, and Ryan McLaughlin, "The Digital-Led New Normal," The Information and Communications Technology Council (ICTC), August 2020, https://ictc-ctic.ca/reports/the-digitalled-new-normal; Rosina Hamoni, Olivia Lin, Mairead Matthews, and Peter Taillon, "Building Canada's Future AI Workforce," March 2021, The Information and Communications Technology Council (ICTC), https://ictc-ctic.ca/reports/building-canadas-future-ai-workforce.

RECOMMENDATIONS

In addition to the above components, the development of Canada's national AI compute infrastructure should be guided by six principles: industrial strategy, intellectual property (IP) ownership and retention, supply chain resiliency, environmental sustainability, regionality, and tax and financial incentives.

A) Industrial Strategy:

Canada should leverage industrial strategy when developing national AI compute infrastructure. This includes using demand-side policies to incentivize the development and use of Canadian-owned infrastructure; prioritizing Canadian suppliers when developing national AI compute; building out local manufacturing capacity for AI-specific chips; creating national or regional AI clusters that integrate compute resources with research labs, universities, and industry partners; continuing to support AI compute R&D; and encouraging substantial investment in high-performance computing (HPC), cloud infrastructure, and quantum computing capabilities to support Canada's competitive advantage in the AI compute race. Further, to safeguard Canada's competitiveness on the global stage, Canada should invest proportionately, relative to peer economies in domestic AI compute. Canada's recent commitments to increase its national defence spending may be an opportunity to leverage defense funding for AI infrastructure with spillover benefits for the rest of the economy.

B) Intellectual Property Ownership and Retention:

Despite significant public investment in AI, Canadian inventors own relatively little AI IP relative to the amount that they create, and valuable technology startups are often acquired by foreign buyers or exit the Canadian ecosystem.[61] Canada risks becoming a consumer rather than a creator of foundational AI technologies, which can stifle the development of homegrown IP and limit competitive advantage in AI-intensive sectors like health, fintech, manufacturing, and climate tech.

Without domestic compute infrastructure, Canadian AI researchers and companies may be forced to rely on foreign platforms for developing cutting-edge models. This can shift the value chain—including model ownership, optimization expertise, and derivative technologies—out of Canada. With new public investment in AI on the table, integrating IP considerations into the national strategy is crucial to avoid losing a strategic IP advantage. In addition, introducing more robust data residency laws may help ensure that sensitive Canadian data is processed within Canadian borders.

61 Mairead Matthews and Faun Rice, "Context matters: Strengthening the Impact of Foreign Investment on Domestic Innovation," 2022, The Information and Communication Technology Council (ICTC), https://ictc-ctic.ca/reports/context-matters.

C) Supply Chain Resiliency:

ICT manufacturing and production rely on a complex and highly globalized supply chain.[62] ICT hardware and devices are built using a long list of raw materials, such as indium, lithium, tantalum, gallium, copper, silver, gold, and rare earth elements, and components, including printed circuit boards, backplanes, enclosures, cables, precision machine components, and optical modules.[63] Mining and production of these inputs occur on a global scale, with a large proportion taking place in regions that are geographically very distant from Canada, such as China and the Democratic Republic of Congo.[64]

During the use phase of the ICT lifecycle, ICT hardware located within data centres also depend on a large and reliable supply of energy to power devices and building operations as well as large volumes of water to remove heat waste from ICT equipment.[65] A study from 2019 estimates that data centres with 15MW of IT capacity consume between 0.8 and 1.3 million litres of water per day, while another estimated that in 2014, in the United States alone, data centres consumed a collective 165 billion gallons of water.[66] Because of these dependencies, building national capacity to produce ICT inputs is essential for reducing supply chain risks.

D) Environmental Sustainability:

The government should prioritize sustainability when developing AI compute infrastructure, following standards such as the International Telecommunications Union's L. 1300: Best Practices for Green Data Centres.[67]

E) Regionality:

Most of Canada's data centres are concentrated in a few provinces, such as Ontario, Quebec, British Columbia, and Alberta, and large cities, such as Toronto, Montreal, Vancouver, Mississauga, Winnipeg, Ottawa, and Halifax. Expanding infrastructure should consider the needs of urban and rural areas across all regions.

62 "Assessment of the Critical Supply Chains Supporting the US ICT Industry," February 2022, U.S. Department of Commerce and US Department of Homeland Security," https://www.commerce.gov/sites/default/files/2022-02/Assessment-Critical-Supply-Chains-Supporting-US-ICT-Industry.pdf.

63 "Assessment of the Critical Supply Chains Supporting the US ICT Industry," United States Department of Commerce and United States Department of Homeland Security, February 2022, https://www.commerce.gov/sites/default/files/2022-02/Assessment-Critical-Supply-Chains-Supporting-US-ICT-Industry.pdf; "Digital Economy Growth and Mineral Resources: Implications for Developing Countries," December 2020, UNCTAD, https://unctad.org/system/files/official-document/tn_unctad_ict4d16_en.pdf.

64 United Nations Conference on Trade and Development (UNCTAD), "Digital Economy Growth and Mineral Resources: Implications for Developing Countries," UNCTAD Technical Notes on ICT for Development No. 16, December 2020, https://unctad.org/system/files/official-document/tn_unctad_ict4d16_en.pdf.

65 Kaitlyn Carr, Allison Clarke, and Mairead Matthews, "Building a Sustainable ICT Ecosystem: Strategies and Best Practices for Reducing Environmental Harms in a Digital World," January 2024, Information and Communications Technology Council (ICTC), https://ictc-ctic.ca/reports/building-a-sustainable-ict-ecosystem.

66 Ibid.

67 International Telecommunication Union (ITU), "L.1300: Best practices for green data centres," Telecommunication Standardization Sector of ITU, June 2014, https://www.itu.int/rec/T-REC-L.1300-201406-I/en.

68 See: Flow-through shares (FTSs), Government of Canada, last updated April 28, 2008, https://www.canada.ca/en/revenue-agency/services/tax/businesses/topics/flow-through-shares-ftss.html.

F) Tax and Financial Incentives:

Tax instruments, and other financial incentives such as grants, could be used to align private domestic investment in AI compute infrastructure with wider Canadian industrial policy. Large domestic corporate actors, such as financial institutions, energy companies, and telecommunications companies, could be provided with tax incentives to encourage public-private co-investment in AI compute infrastructure.

Innovative tax mechanisms, such as flow-through shares, could be adopted for Canadian AI infrastructure providers and start-ups to encourage domestic investment in AI infrastructure. Flow-through shares are a commonly used tax instrument in Canada's energy and mining sectors.[68] Other instruments, such as tax rebates, could be provided to Canadian startups and other small- and medium-sized enterprises to help subsidize access to domestically provided AI compute power. Such policies could help build a critical mass of domestically based AI compute infrastructure in Canada.

CONCLUSION

Canadian startups and SMEs use a broad range of AI compute infrastructure, including personal workstations, centralized data centres, cloud services, frameworks, models, data, and talent in developing and deploying novel AI technologies. Despite Canada's early strengths in AI research, talent, and adoption, it lags in domestic AI compute, which could hinder its technology leadership potential and put its future prosperity at risk. Peer economies such as the United States, France, and the European Union are outpacing Canadian investment in sovereign compute by magnitudes.

Expanding Canada's compute capacity to train and deploy AI technology will help underpin the country's digital sovereignty in an uncertain and constantly evolving geopolitical environment.

As Canada designs public AI compute access programs and future sovereign AI compute capabilities, it is vital to engage in consultations with private companies, academic institutions, and non-profits to fully understand the nation's AI compute needs and its global connections.


ICTC CTIC
www.ictc-ctic.ca