

Canada as a Champion for Public AI: Data, Compute and Open Source Infrastructure for Economic Growth and Inclusive Innovation

This is an authors' working draft of an article for an upcoming collected volume on Canadian internet policy. It was presented in draft form at The Future of Canadian Internet Policy at McMaster University and is currently undergoing further review.

Nicholas Vincent, Mark Surman, and Jake Hirsch-Allen

ABSTRACT

Current AI models and products are predominantly built by a very small number of actors within only two incentive systems and political contexts, namely (1) big tech, centred in Silicon Valley and (2) China. This poses serious concerns about how AI will impact the rest of the world. The concentration of influence could prevent building AI as a public good, inhibit nations like Canada from unlocking transformative economic change, and increase risks – including democratic deficits, greater inequality, and existential concerns – worldwide. In this article, we outline a tractable roadmap for Canada to work with a coalition of public and private actors around the world to champion a third option: “Public AI”. These actors should include government institutions, private funds and companies, research institutes, open source projects, and civic organizations. This proposal builds on others’ extensive research on AI as a public good or utility [53], such as calls for a “CERN for AI” or “Airbus for AI”. If successful, this proposal would allow Canada and allies around the world to build and deploy competitive public AI infrastructure that drives economic growth and public benefit. Publicly backed AI research and compute programs, as well as commercial and nonprofit open source AI projects, already provide a solid foundation. What we still need is: 1) greater and faster international collaboration, 2) explicit, publicly backed open source and standards-based strategies that make it possible for innovations and improvements in one country, company, or lab to accrue into a common pool of interoperable public goods used by all actors, and 3) explicit data strategies that enable AI development and public interest uses of AI.

1. INTRODUCTION

Building robust and healthy digital economies is a dance between public and private actors. Researchers in universities, often supported by public grants, make fundamental breakthroughs and invent new technologies. Companies looking for new or expanded markets turn these inventions into products used by millions. And public interest organizations – such as governments, as well as standards and open source communities – create public goods that fuel economic growth and social benefit for all of us. This was the dance we performed with the

internet, the web and the open source software stack that currently underpins most digital businesses and public services [3]. Importantly, the public goods produced during these eras created widely spread social and economic benefits in many parts of the world. For instance, a recent study from Harvard estimated that every \$1 of investment in open source software creates \$2000 in value for businesses, governments and others who freely use and build on this software [23]. Economists have highlighted the potential for open source AI to “be a critical driver of innovation, economic growth, and shared prosperity” [26].

The challenge: the current AI landscape is dominated by the market – and by a very small number of companies – in a way that may short circuit the economic and social benefits we have seen with earlier eras of digital technology. The present landscape involves too little AI investment in areas like safety, democracy, trust, healthcare and education relative to their societal value. Companies highly concentrated on the west coast of the US (“big tech”) and China are rapidly locking down the current wave of innovation in ways that will make it hard for researchers and companies in other places to compete (see, for example, coverage of OpenAI’s letter to the White House [36], other lobbying efforts from big tech [45], and China’s AI boom [52]). While the definition of big tech and China’s equivalent companies changes as fast as their technologies, the list of companies includes Google, Meta, Microsoft, Apple, Amazon, Baidu, Alibaba, Tencent and ByteDance.

We need a counterweight to the private players who currently dominate the AI landscape. Our vision for public AI – inspired heavily by the Public AI Network [24] and Mozilla and Columbia University’s recent work on open source [9], responsible data [8] and AI safety [2] – is a robust and geographically distributed ecosystem of initiatives that create public goods, have a public orientation, and facilitate public use throughout every step of AI development and deployment, including compute, data, tooling and talent. The idea is not to replace the private AI companies, but rather to ensure they have public interest dance partners focused on broad economic growth and benefits to society.

Canada is ideally positioned to help build a coalition of governments, companies, non-profits, and open source projects investing in and stewarding public AI, based on its history of AI innovation, international coalition building, and leadership in areas like human rights and multiculturalism. Canada’s existing public investment in a national network of research labs and startups can serve as a foundation for these efforts: the Pan-Canadian AI Strategy modelled public AI before it had a name. Furthermore, Canada has immediate incentives to use these strengths to help build an international coalition for public AI. We need to find a way past its singular dependence on the resource economy [1, 19] and to build a much larger and more sustainable digital sector – and making the right choices on AI could help us to achieve this diversification and corresponding economic growth.

Despite tremendous horsepower, investment, knowledge and potential for developing global tech companies, we've struggled to sustain industry leadership even where we have developed important beachheads. From Corel to Nortel to BlackBerry, Canada has seen early leads evaporate. Even our current successes like Shopify and Cohere face uncertain futures in a global marketplace dominated by big tech and China. The question is: what can Canada do to break the cycle of false starts and build durable technological advantage?

We believe the path forward is to band together with others who want to build a “third way” on AI that is based on neither American big tech nor its Chinese equivalent. This way is **public AI**. Building this third way will require active, elegantly designed collaboration amongst governments, startups, civil society and open source communities in many different countries.

This paper outlines how Canada can leverage its unique strengths to champion public AI as both an economic catalyst and social good. Moving from the current fragmented public AI ecosystem to an effective alternative infrastructure requires:

1. Greater and faster international collaboration amongst a coalition of startups, open source projects, research labs, and national and regional governments.
2. An open source strategy that makes it possible for innovations and improvements in one country, company, or lab to accrue into a common pool of interoperable public goods used by all actors.
3. A data strategy specifically focused on enabling public AI development and public interest uses of AI including legal and technical support for ethical capture, storage, governance and sharing of high-quality data including vetted datasets.

By pursuing “digital sovereignty through collaboration” – pooling resources with others to create an “Airbus or CERN for AI”, ideas for which other scholars have developed substantive and tractable proposals – Canada can help establish a third pole in AI development distinct from the big tech and China, enable pluralistic uses of AI, and secure both economic independence and social benefit.

To be very upfront about the challenges ahead of us, we must note: the vision laid out in this paper will require coordination – in public and private investment, in political action and policy making, and in research efforts – across institutions and domains. This will be a monumental undertaking. Throughout this paper, we take an optimistic tone, aiming to highlight a path that can bypass numerous obstacles. We return in our conclusion to the topic of how follow up work (which is already in progress) can further document likely challenges and increase the tractability, political and otherwise, of this plan.

2. WHAT WE MEAN BY PUBLIC AI

Public AI infrastructure includes the foundational elements upon which any AI application can be built, such as compute, data, standards, models and the software and talent to build and orchestrate all these elements into a working AI system. Put simply, public AI means building AI systems that are made available as public goods (like utilities, roads, or open source software [53]), and these systems must maintain public accessibility, democratic accountability, and financial and environmental sustainability. This definition draws on definitions from the Public AI Network [24] and Mozilla [31].

Public AI is not about governments monopolizing AI development, nor is it focused on regulations to control or restrict private innovation. Rather, it's about a coalition of public, private and nonprofit actors building foundational elements that enable a diverse ecosystem of applications, both commercial and non-commercial. Ensuring that the basic building blocks are accessible and open to everyone everywhere – and not controlled by a tiny handful of private players – will require stewards for sustainability, standards for interoperability and robust collaboration to create a compelling alternative to the offerings of big tech and China.

2.1 Plentiful Analogies for Public AI

History provides a number of successes to look towards beyond the previously mentioned roads and electrical grids, such as Canadian and British broadcast media, the Internet, and multinational projects like CERN and Airbus. Other examples include the consolidation of weather data, the implementation of GPS, scientific endeavors like the Human Genome Project, scholarly research data sharing, as well as ongoing projects that focus on “digital public infrastructure” [16].

British and Canadian approaches to broadcasting, culture, and the arts can serve as strong exemplars facilitating “digital abundance”. Both media and AI systems fundamentally deliver information to people, and thus are affected by the “public goods” properties of information [43], in particular that information is non-rival (everyone can “consume” a unit of information at once) and hard to exclude (it’s very easy to copy and distribute information, whether it’s a news headline spread via word-of-mouth or AI outputs from the latest model spread via online platforms). Looking to successful institutions such as the CBC [48], BBC [25], and the Canadian Council for the Arts [21] that deliver informational public goods can provide an excellent roadmap for the governance of AI.

Looking to science, CERN stands as one of humanity's most successful examples of international collaboration. The former president of Switzerland commented on CERN that “the provision of global public goods is a major issue for global survival” [41]. CERN’s success demonstrates that shared costs make advanced infrastructure possible when no single nation can afford it. This international public infrastructure has generated both scientific and economic benefits. As one

such example: Tim Berners-Lee invented the world wide web while at CERN [42], paving the way for numerous future innovations (including web data-hungry generative AI systems).

2.2 Digital Sovereignty through Collaboration

The reality is that neither a single government in a country like Canada nor mid-sized tech companies can compete alone against the massive AI investments of big tech and China. The solution is what we might call digital sovereignty through collaboration: pooling resources with like-minded governments, companies, academia, civil society and open source projects to create collective capacity.

This is the Airbus model - named after the European aircraft manufacturer created as a multinational consortium to compete with American aviation dominance. An "Airbus for AI" would bring together democratic nations with shared values to build public AI infrastructure at sufficient scale to serve as a meaningful third pole.

There have been numerous explicit calls for Airbus models for AI, including from Daniel Crespo and Mateo Valero [17]. Brandon Jackson and Josh Tan of the Public AI Network among others championed an Airbus for AI approach at the 2025 Paris AI Action Summit. Similarly, there have been many calls [12, 27] for a CERN for AI, including from European Commission president Ursula von der Leyen [40]. Cutting across both sets of calls, in 2023, Yoshua Bengio, director at the Mila, advocated for a network of publicly-funded AI labs [5].

Our argument for digital sovereignty through collaboration builds on numerous national digital sovereignty and subsequent AI sovereignty initiatives such as the 'national champions' strategies of France, Canada and Germany [13]. Public AI makes success on sovereign AI more likely by allowing governments and private actors to benefit from shared public goods. Mistral and Cohere have hybrid closed and open source strategies that are examples of the public and private upsides of public goods. The public, and these private companies, are contributing to and benefiting from the open source elements of Public AI as part of these AI sovereignty initiatives.

This proposal is designed to build on the calls for an Airbus, CERN for AI and network of AI labs in addition to the digital and sovereign AI movements including numerous public compute investments, to ensure these public investments generate sufficient ROI across compute, data, algorithms and the applications they support. We aim to answer the questions: what is needed to make Public AI a viable counterweight to big tech and Chinese dominance and what role can Canada play?

We need a coordinated strategy across three layers: compute, an open software stack, and high quality data. We need timelines and commitments across each to be coordinated across the public and private sectors and internationally. In the sections that follow, we describe the steps of

the dance that will take us the last mile towards public AI, expand on Canada’s unique strengths, and then summarize concrete actions Canadian institutions can take now.

3. THREE LAYERS FOR PUBLIC AI

3.1 Data for Public AI

AI is built on data. This crucial infrastructure layer constitutes the raw material that AI systems need to learn from, and which is the bedrock of science and computing. With no data, there is no model. Over the course of AI development, the data available on the internet — such as text on public websites crawled by nonprofits like Common Crawl — has been foundational to training AI models. Now, as that is no longer enough data to train increasingly large and powerful models, big tech is striking exclusive partnerships with social media platforms and news companies to acquire more content [37, 49] and exploring “synthetic data” generated for the purposes of training AI models (see e.g. “Phi” models from Microsoft [4]).

Data represents a potential “Achilles Heel” [32] for the current generation of generative AI products. These models are trained on data acquired through potentially illegal, anticompetitive, and certainly unsustainable methods, from scraping creative works without compensation to harvesting personal information without meaningful consent [30]. In some jurisdictions, current commercial AI products could be deemed illegal because of their data dependencies or could be rejected by consumers on moral grounds. This backlash could elevate open-access and openly licensed LLM data-based public AI models to become the most competitive AI offering.

As AI systems increasingly rely on expert human feedback throughout their lifecycle—from training to monitoring and evaluation to push the frontier of AI and add layers of expertise to the systems —access to data becomes critical. Unlike traditional software, AI systems uniquely benefit from keeping humans in the loop, and this means a growing need for the AI development and deployment to be anchored in high-quality, curated human input. Unfortunately, human generated data is resource-intensive and often inaccessible to both open source and public-interest efforts. To counter this, organizations like the Human Feedback Foundation and AI Objectives Institute envision a data commons that enables public-interest AI by providing access to vetted datasets, tools for responsible data stewardship, and legal scaffolding that supports safe, open collaboration, e.g. citizen assembly deliberations at scale.

Furthermore, the types of data sourced for commercial AI development are often not the kind of data needed for training smaller, more specialized AI models for public interest AI applications; there is an overarching focus on acquiring as much clean data as possible to train foundation models (e.g., scraped data from the internet) rather than specialized datasets (e.g., disaggregated data about health outcomes).

While OpenAI has made an explicit case that scraping is necessary (and that the US will “lose” to China if OpenAI is not allowed to scrape indiscriminately) [36], others have argued this is a false dichotomy [8]. With sufficient investment, open and ethically sourced datasets can achieve competitive performance. The industry's reliance on questionable data practices is a choice driven by speed and cost concerns, not technical necessity. Public AI efforts can prove that high-performing models are possible through the careful curation of properly licensed and public domain data, combined with innovative approaches to data governance and efficiency and targeted collection of specialized datasets in areas of public importance such as healthcare, climate science, and education. Further, there is an argument that because modern AI systems depend on broad, collective artifacts of human creativity, a public AI approach is a better moral fit for the nature of the technology.

Concretely, the specific programs by which a public AI coalition might build openly licensed datasets might involve a combination of (1) data archiving and digitization processes (curating openly licensed web content) and (2) establishing opportunities for peer production (e.g., leveraging ongoing interactions with public services like government chatbots, and creating a platform for users of LLMs to contribute feedback to open projects, as an alternative to providing “thumbs up” feedback data only to private companies).

There are numerous examples in the world that we can look towards: Wikipedia, the Internet Archive, and Serpentine Labs all actively participate in AI-aware data creation and curation. Open government datasets, including data.europa.eu, data.gov, and AI Sweden’s “data factory” all provide examples. In the realm of language, efforts like the First Languages AI Reality initiative and Mozilla’s Common Voice [6, 18] have already made great strides towards usable public interest projects.

Even if the current data paradigm is not overturned entirely (i.e., if ChatGPT and Llama-style products dependent on scraping, torrenting, etc. remain on the market), open data can substantially close the gap between public and private AI. By also prioritizing datasets in areas like healthcare, education, climate (see e.g. focus areas of organizations like Current AI), as well as known data gaps (e.g. along cultural, linguistic, urban-rural, and other lines), public AI can compete in those domains and provide more real utility to end-users.

Another opportunity for public AI allies to advance public interest data would be building new data intermediaries [28]. This would involve radically reshaping the flow of data between users and companies, by injecting new intermediary organizations. For instance, this might mean that when a user sends a query to an LLM or a post to social media, the actual content of that message is stored first on a server operated by the intermediary, which engages in governance processes to make governance decisions: what types of data will be sent to private companies and what will be contributed to a commons? How will compensation work? An ecosystem of intermediaries would make it possible to honour creator preferences and compensate data

creation while accumulating high-quality data in a public pool. Practically, intermediaries could be instantiated as public benefit companies, nonprofits or even governments in places where the rule of law is strong.

All of these efforts should involve commitment to proposed best practices for open data [8]: a focus on enabling a competitive LLM ecosystem, reproducibility, taking seriously creator preferences and harms from data collection, collecting diverse data, and stewardship for the maintenance and durability of clean and ethical data.

3.2 The Compute Layer

The second foundation of AI development is compute - the raw processing power needed to train and run AI models. While many democratic economies are investing public money in AI compute resources, these efforts remain fragmented and subscale compared to American and Chinese investments, which are centralizing compute resources in the hands of a few large companies. The United States has 90 times more raw computing power than Canada [22]. Adjusting for population size, this becomes 7.6x. This compute gap threatens to lock Canadian, and most other countries' researchers, startups, and public institutions into dependency on foreign, and primarily big tech's, platforms, with associated risks to data sovereignty and economic autonomy.

As was discussed amongst several leading public labs at the 2025 Paris AI Action Summit, in an Airbus for AI model, compute resources would be pooled and distributed across participating nations - much as the Stargate consortium allows member companies to share compute capacity. Canada, the Nordic nations, Taiwan, Brazil, India, Switzerland, Germany, Spain, France, Japan, and other leaders from the global majority could contribute data centres that together form a powerful distributed network.

Critically, by establishing norms around direct model checkpoint sharing and other processes that require close collaboration, it will be possible for a coalition to save on compute. For instance, members can pool training checkpoints to support distributed experimentation. Technical work has suggested a "pool of models" approach can be promising and even has benefits for understanding how particular subsets of data affect AI capabilities [33]. Similarly, cleaner, better structured and more specific data sets could decrease compute requirements, as could supporting exploration of less compute intensive methods.

Canada's climate, abundant clean energy (and long-term planning around energy), and world-class institutes like Mila, the Vector Institute, and Amii [10] all provide unique strengths for contributing towards compute pooling. However, this doesn't mean building all the capacity within Canada. Rather, we would contribute strategically to a distributed network that provides resilience, diversity, and scale far beyond what any single participant could achieve alone,

building on existing initiatives in Europe and conversations like those at the Paris AI Action Summit.

Our partners should not just be European nor just nations: they should include countries across Asia and the global majority, large cities and regions, as well as non-state actors including corporations like Cohere and Hugging Face, experts in decentralization like the Ethereum Foundation and Project Liberty, foundations like the Future of Life Institute and key civil society actors from standards organizations to open source projects. To use one example, Canada should form a closer relationship with the leading chip manufacturer Taiwan, as we have discussed with its Tech Ambassador and former Minister of Digital Affairs Audrey Tang (who has written extensively on Plurality and decentralized collaboration [44]).

3.3 An Open Software Stack

The third critical layer of AI infrastructure is the software stack - the complete set of tools needed to build, train, and deploy AI. This includes training and deployment code, the actual model weights that must be deployed to a server, safety guardrails, and the documentation needed to fine tune and otherwise customize AI for use in a finished product [9]. Or, put more simply, this stack includes all the software someone needs to make something using AI.

For public AI to succeed, we need to build a fully open stack for AI that makes these components available to anyone. This assertion is based on experience from the first era of the internet, where open source came to serve as the core underpinning of most infrastructure and a huge driver of economic growth. Over the last 20 years, the Linux stack has become the standard infrastructure upon which all things digital are built, underpinning 90% [46] of publicly available cloud computing services and running 100% of the top 500 super computers [38]. The core open source stack including Linux represents \$4 billion in supply side value (how much it cost to create) and \$8 trillion in demand side value (the value it created for businesses and governments) [23]. The technical and economic value created by open source in the last era of the internet emerged by many businesses and governments — including competitors — pooling resources to create robust core infrastructure that everyone can build on. Public AI needs to take a similar approach.

Unfortunately, the current dominant model of delivering generative AI hides the stack behind a chatbot or an API and is built mostly from proprietary components. For example, all the components of the AI stack used to deliver ChatGPT or Claude are opaque to users — they appear only as product features. For big tech, this approach can be beneficial as it creates a moat against competitors. However, it does not make sense for public AI efforts, where the goal is to rapidly create an AI stack that entrepreneurs, researchers, communities and governments can use freely, customize and build on.

The good news is that many parts of the AI ecosystem are already open source — and we have seen a move towards open source in base models over the last two years (see e.g. Hugging Face’s AI Action plan [34] and the Stanford’s AI Index reports [50] for coverage of these

trends). This new wave of base models spans a spectrum of openness [2, 9] from those with openly licensed software, datasets and model weights (e.g., OLMo from the nonprofit AI2, the Allen Institute for AI [35] and Salamandra from BSC, the Barcelona Supercomputing Center [20]) to those with open model weights, training and fine tuning software but no access to the pretraining data (e.g. various models from Mistral) to those that simply offer open weights (e.g., Llama from Meta). While the most rapid progress has been in open weight models like Llama, fully open models are catching up.

It's important for champions of public AI to understand the upsides of investing in a fully open AI stack and the risks of relying on more proprietary players. Having a fully open stack – with all training code and documentation alongside data for pretraining, post training, and evaluation – provides the freedom to fork any element of the stack (e.g. to build a model deeply tuned to a specific cultural or language context by intervening in the pretraining process at a particular checkpoint) and the transparency to fully scrutinise a trained model (e.g. for cultural or language bias, or for hidden security threats). It also makes it easier for many different parties to contribute back to a common stack as they fork, innovate and extend things in their own context. Open weight models like LLaMa offer none of these advantages. Meta has also signalled [54] that they may not release the weights for future Llama models.

These factors make the case even stronger that public AI efforts must build — and be built on top of — open source technology. Companies like Meta have the resources to continually train cutting edge models — and nonprofits like AI2 and Kyutai do not. While no single startup, nonprofit or country can easily match the resources of major tech companies, a consortium approach that pools compute and research talent creates a competitive open alternative — providing developers with the guarantee that the AI model will continue to be maintained and improved over time, even if it continues to gain in performance and quality.

Beyond open source software, we will also need a concerted effort to help developers embrace open source AI in practice. There are likely to be practical barriers to using outputs from the open software stack that mirror experience with Linux. Twenty-five years ago, Linux was free and high quality, but much harder to use than proprietary offerings like SunOs (Unix) and Windows NT. Projects like RedHat and Ubuntu made it easier for developers to quickly pull together many elements from an open source stack into a single server setup or development environment. A number of startups including Hugging Face are focused on making it easier to use open source AI — but an explicit open source strategy from players like Canada (described below) will be needed.

As we expand on below, Canada brings both strengths in international and multi-party coordination, AI and open source. Of particular note is the Aya project led by Cohere for AI, a non-profit research lab run by the Toronto based AI company Cohere. Aya is a global initiative to advance the state-of-the-art in multilingual AI and bridge gaps between people and cultures across the world, involving 3,000 independent researchers across 119 countries [47].

4. CANADA'S UNIQUE ADVANTAGES

World-class AI leadership: Canada has been at the forefront of artificial intelligence research for decades – even when machine learning was considered a fringe approach. The Pan-Canadian AI Strategy launched in 2017 was the world's first national AI strategy, with a focus on research, talent and ecosystems, it established three national AI research institutes, Mila in Montreal, the Vector Institute in Toronto, and Amii in Edmonton [39]. Geoffrey Hinton, Yoshua Bengio and Richard Sutton, three of the foremost AI pioneers, built their careers in Canada, and Canada is at the top of global AI talent rankings [11, 29].

These institutes have created clusters of excellence, attracting global talent and incubating innovative startups. Companies like Cohere and Borealis AI demonstrate the commercial potential of Canadian AI research (and the potential for non-profits like Cohere for AI).

Many of the breakthroughs developed in Canadian labs have been commercialized elsewhere, with the economic benefits flowing primarily to American companies. Expanding on these early Public AI investments offers a path to retain more of that value within the Canadian economy.

Trusted Diplomatic Soft Power: Building a public AI consortium aligns perfectly with our diplomatic tradition of multilateralism, coalition-building, pluralism. Under leaders like Lester B Pearson and Lloyd Axworthy we championed the International Declaration of Human Rights, treaties on landmines, acid rain and CFCs, trade and more, becoming renowned for our Soft Power and ability to “punch above our weight” internationally [7, 14]. Canada remains highly trusted and intentionally connected making us an ideal champion to form this coalition. Public AI also offers an opportunity for Canada to re-establish its world-leading diplomatic status.

Canada's strong institutions, commitment to rights-respecting governance, fiscal capacity for strategic investments and core democratic values all naturally align with the public AI ethos. As tensions rise between the U.S., Europe and China over technological dominance, Canada can offer neutral ground for essential conversations about safety, access, and governance.

Strategic Resources: Canada's abundance of clean hydroelectric power, particularly in Quebec and British Columbia, provides a sustainable foundation for AI infrastructure. While fossil fuel energy in places like northern Texas and Alberta may be cheaper in the short term, we're not building Public AI for the next two years – we're building it for the next 10-20 years. As carbon pricing and environmental regulations tighten globally, Canada's clean energy advantage will only grow stronger.

Cultural and linguistic diversity and good governance: Our immigrant populations and Indigenous communities provide connections to diverse perspectives that can inform more inclusive AI development. Further, Canada remains attractive to global talent, with a reputation and immigration policies attracting top skilled professionals, particularly from the US and China.

Canada is also at the forefront of governance innovation (e.g., from civic assemblies to federalism) that will be crucial to ensure AI is democratically legitimate and aligns with human values.

These advantages create a unique opportunity for Canada to pioneer a different approach to AI development - one that balances economic growth with public benefit, and short-term commercial interests with long-term social welfare.

5. CONCRETE ACTIONS FOR CANADIAN PUBLIC AI LEADERSHIP

To push forward the vision outlined above, Canada should: (1) assemble a public-private, international, and interprovincial public AI collaboration strategy, (2) champion an internationally focused open source AI strategy, and (3) build a public AI focused national data strategy.

5.1 International and Public Private Collaboration

Canada is well positioned to act as a catalytic and central force within the public AI coalition. Specific actions – each of which will be an undertaking – that Canada should take include:

- Convene an "Airbus for AI" consortium consisting of a core set of governments, companies and international and civil society organizations.
- Develop and champion a proposal to coordinate AI compute investments with the EU, other nations, key international players like the OECD, UN, IGF and non-governmental players like foundations
- Prioritize clean energy locations and the \$15 billion the federal government proposed to match \$30 billion in pension fund support to build Canadian data centres [51]. Make these available to consortium members at cost.
- Use public AI strategy to coordinate investments between Canadian provinces.
- Fund civic organizations, cultural institutions, and public services to work with their international and interprovincial peers on public AI initiatives.

Overall, Canada should move away from trying to lead on regulation and instead prioritize the creation of public AI infrastructure underlying a vibrant ecosystem of domestic and international, private and public actors. This alone could vault it back to international diplomatic leadership and offer significant social and economic benefits.

5.2 Open Source Strategy

Canada should find ways to encourage researchers and companies to use and contribute to the broader open source AI stack:

- Open source artifacts funded by public dollars such as those produced by national labs, and public-private partnerships
- Work with allies around the world to develop practical ways to pool and rapidly improve the public AI stack through a joint commitment to open source, open content and open data
- Ensure that base models are kept truly open and unrestricted while also allowing elements of the stack to be used in more restrictive settings
- Encourage international collaboration between existing open source efforts. Open sourcing is already happening with contributions from Canadian firms like Cohere and Transformer Labs, European firms like Mistral, nonprofit labs like AI2 and Kyutai, and national projects like AI Singapore, AI Sweden, BSC, and the UK AI Safety Institute.

As noted above, no one country or mid-sized company can create an alternative to big tech or China on its own. A coherent, well-resourced international open source strategy would underpin Canada's development of a strong, independent, globally connected economy for the AI era.

5.3 Data Strategy

Canada should establish a data strategy that at once enables rapid progress on AI innovation and at the same time prioritizes fair compensation, privacy and representation. Specifically, Canada should:

- Create and support both data trusts (looking to examples like the Department of Canadian Heritage's framework [15]) and sectoral data cooperatives in areas of Canadian strength, such as healthcare, clean energy, and financial services.
- Create new publicly managed "data flywheels" that feed high-quality data into said trusts, cooperatives, and open source projects
- Fund institutions that are producing, archiving, and curating public domain and openly licensed data, looking to the examples described above like the Internet Archive.
- Implement frameworks for indigenous data sovereignty to ensure that indigenous communities maintain control over their own data.
- Lead in establishing cross-border data sharing agreements specifically for public AI development

Building a thriving economy and society for the AI era will require an approach to data that both fuels AI progress and considers the rights and livelihoods of those who contribute data. Canada has a chance to play a leading role in showing that these two goals can be met in tandem creating an economically competitive alternative to big tech and China.

6. CONCLUSION: THE TIME FOR CANADIAN LEADERSHIP IS NOW

The window for establishing alternatives to the current trajectory of AI development — and to create economic opportunity for players beyond big tech and China — is rapidly closing. As compute resources concentrate, data sources are exhausted or locked up, and technical talent flows to a handful of companies that are led by a small number of billionaires with decreasing democratic checks on their power, AI's potential future benefits for the average human relative to its risks diminish.

Canada stands at a crucial juncture. We have the expertise, resources, and diplomatic credibility to champion an alternative approach - not just for our benefit, but for democratic societies worldwide. By focusing on public infrastructure, international collaboration, and balancing public and private interests, we can create a distinctive and economically viable position in the global AI landscape. For the Canadian economy, this will have immediate benefits: we can support Canadian-based AI companies, attract international talent, and reduce dependencies on foreign AI platforms and associated economic leakage. Further, we can embed Canadian values of democracy and respect for rights into AI systems, preserve Canadian cultural sovereignty and dignity in an era of AI-driven content and services, and contribute broadly to making AI systems that lead to fair outcomes for diverse populations.

This is not merely idealism - it represents a pragmatic economic strategy for a mid-sized power navigating between technologically advanced, economic superpowers. By leveraging our unique advantages and addressing the fundamental challenges of the current AI paradigm, Canada can secure both economic benefits and social goods. By pooling resources across compute, software, and data layers, this consortium could achieve the scale necessary to present a meaningful alternative to existing models.

For Canada, this represents both a return to our strengths in multilateralism and an innovative approach to technological development. Rather than attempting to win a game designed for larger players, we can help change the game itself - establishing rules and infrastructure that better serve democratic values and diverse interests. We have the tools to navigate the tensions holding AI development back. Canada can demonstrate the viability of alternatives to data acquired through potentially illegal, anticompetitive, and unsustainable methods. It can better balance regulation with funding and flexibility for innovation and talent to compete.

Collaboration can overcome big tech and China's concentration of power and money by connecting the many existing publicly backed and open source AI initiatives so that they add up to more than the sum of their parts. Values can be a greater draw for talent than dollars. Academia, government and civil society can work more effectively with companies to create value through carefully choreographed public-private partnerships and ecosystems than black and white zero sum competition with the private sector. While this pooling approach can help to mitigate many of the reasons an ambitious undertaking like building public AI might fail, there is

still work to be done in terms of documenting specific obstacles that will arise and producing reactive plans (e.g., how can public AI proceed if current compute investments targets are not met?)

The time to act is now. With clear vision, strategic investment, and international leadership, Canada can help ensure that the future of AI serves not just market interests, but public good. In doing so, we can secure our place in the next technological era while staying true to our values of pluralism, inclusion, and shared prosperity. And we can convert one of the most influential AI communities into one of the most lucrative and socially beneficial AI economies in the world.

REFERENCES

- [1] 10 Key Facts on Canada's Natural Resources – 2023: 2024. <https://natural-resources.canada.ca/science-data/data-analysis/10-key-facts-canada-s-natural-resources-2023>. Accessed: 2025-03-18.
- [2] A different take on AI safety: A research agenda from the Columbia Convening on AI openness and safety | The Mozilla Blog: <https://blog.mozilla.org/en/mozilla/ai/columbia-convening-research-agenda/>. Accessed: 2025-03-21.
- [3] Abbate, J. 1999. *Inventing the Internet*. MIT Press.
- [4] Abdin, M. et al. 2024. Phi-4 Technical Report. (2024).
- [5] AI and Catastrophic Risk: 2023. <https://www.journalofdemocracy.org/ai-and-catastrophic-risk/>. Accessed: 2025-03-20.
- [6] Ardila, R. et al. Common Voice: A Massively--Multilingual Speech Corpus. *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*.
- [7] Axworthy, L. 1997. Canada and human security: the need for leadership. *International Journal*. 52, 2 (1997), 183–196.
- [8] Baack, S. et al. Towards Best Practices for Open Datasets for LLM Training.
- [9] Basdevant, A. et al. 2024. Towards a Framework for Openness in Foundation Models: Proceedings from the Columbia Convening on Openness in Artificial Intelligence. arXiv.
- [10] Can Canada Compute? - The Dais: 2024. <https://dais.ca/reports/can-canada-compute/>. Accessed: 2025-03-22.
- [11] Canada's leadership in AI – talent, ecosystems, and responsible AI - OECD.AI: <https://oecd.ai/en/wonk/canada-national-ai-strategy-2021>. Accessed: 2025-04-07.
- [12] CERN for AI | CAIRNE - Confederation of Laboratories for Artificial Intelligence Research in Europe CAIRNE - Confederation of Laboratories for Artificial Intelligence Research in Europe AI: <https://cairne.eu/cern-for-ai/>. Accessed: 2025-03-23.
- [13] Chavez, P. 2024. Sovereign AI in a Hybrid World: National Strategies and Policy Responses. *Lawfare*. (Jul. 2024).
- [14] Copeland, D. 2001. The Axworthy years: Canadian foreign policy in the era of diminished capacity. *Canada among nations*. (2001), 152–172.
- [15] Digital Content Governance and Data Trusts — Diversity of content in the digital age: 2021. <https://www.canada.ca/en/canadian-heritage/services/diversity-content-digital-age/digital-content-governance-data-trust.html>. Accessed: 2025-03-18.
- [16] Eaves, D. and Rao, K. Digital Public Infrastructure.

- [17] Es la hora de ‘AIbus’: por qué Europa debe crear una gran empresa de AI: 2024. <https://elpais.com/tecnologia/2024-04-11/es-la-hora-de-aibus-por-que-europa-debe-crear-una-gran-empresa-de-ai.html>. Accessed: 2025-03-20.
- [18] First Languages AI Reality | Mila: <https://mila.quebec/en/ai4humanity/applied-projects/first-languages-ai-reality>. Accessed: 2025-04-07.
- [19] Gelb, A. 2014. Should Canada worry about a resource curse? *The School of Public Policy Publications*. 7, 2 (2014).
- [20] Gonzalez-Agirre, A. et al. 2025. Salamandra Technical Report. arXiv.
- [21] Granatstein, J.L. 1984. Culture and scholarship: The first ten years of the Canada Council. *Canadian Historical Review*. 65, 4 (1984), 441–474.
- [22] Hirsch-Allen, J. 2024. Toronto Star Op-Ed and OECD AI Observatory post on Canadian AI Compute Capacity and Strategy. *Jake Hirsch-Allen*.
- [23] Hoffmann, M. et al. 2024. The Value of Open Source Software. Social Science Research Network.
- [24] Jackson, B. et al. 2024. *Public AI: Infrastructure for the Common Good*. Public AI Network.
- [25] Jackson, B. 2024. *The UK needs a ‘British AI Corporation’, modelled on the BBC*. Chatham House.
- [26] Joint Statement from Economists on the Importance of Open Source AI: 2024. <https://open.mozilla.org/economists/>. Accessed: 2025-03-23.
- [27] Juijn, D. et al. The EU’s seat at the table.
- [28] Lanier, J. and Weyl, E.G. 2018. A Blueprint for a Better Digital Society. *Harvard Business Review*. (2018).
- [29] Lockhart, A. 2023. *Automation Nation? AI Adoption in Canadian Businesses*. The Dais.
- [30] Longpre, S. et al. 2023. The Data Provenance Initiative: A Large Scale Audit of Dataset Licensing & Attribution in AI. arXiv.
- [31] Marda, N. et al. A vision for a robust ecosystem of initiatives that promote public goods, public orientation, and public use throughout every step of AI development and deployment.
- [32] Mylles, R. 2024. Is data supply AI’s Achilles’ heel? *UBS Articles*.
- [33] Na, C. et al. 2024. Scalable Data Ablation Approximations for Language Models through Modular Training and Merging. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (Miami, Florida, USA, Nov. 2024), 21125–21141.
- [34] Nuñez, M. 2025. Hugging Face submits open-source blueprint, challenging Big Tech in White House AI policy fight. *VentureBeat*.
- [35] OLMo, T. et al. 2025. 2 OLMo 2 Furious. arXiv.
- [36] OpenAI lobbies Trump admin to remove guardrails for the industry: 2025. <https://www.cnn.com/2025/03/13/openai-lobbies-trump-admin-to-focus-ai-on-speed-light-regulation.html>. Accessed: 2025-03-18.
- [37] OpenAI to start using news content from News Corp. as part of a multiyear deal: 2024. <https://apnews.com/article/openai-news-corp-a49144d381796df5729c746f52fbef19>. Accessed: 2025-03-22.
- [38] Operating system Family / Linux | TOP500: <https://www.top500.org/statistics/details/osfam/1/>. Accessed: 2025-03-23.
- [39] Pan-Canadian Artificial Intelligence Strategy: 2024. <https://ised-isde.canada.ca/site/ai-strategy/en/pan-canadian-artificial-intelligence-strategy>. Accessed: 2025-03-24.

- [40] President von der Leyen launches initiative for AI: https://ec.europa.eu/commission/presscorner/detail/en/ac_25_481. Accessed: 2025-03-23.
- [41] Schaeffer, A. et al. 2017. The CERN model, United Nations and global public goods: addressing global challenges: Report on the symposium that took place on 2 November 2015 at the Palais des Nations, Geneva. (2017).
- [42] Segal, B. 1995. A short history of Internet protocols at CERN. (1995).
- [43] Stiglitz, J.E. 1999. Knowledge as a global public good. *Global public goods: International cooperation in the 21st century*. 308, (1999), 308–325.
- [44] Tang, A. and Weyl, G. 2024. Plurality: The future of collaborative technology and democracy. *Plurality. net*. (2024).
- [45] The AI Copyright Battle: Why OpenAI And Google Are Pushing For Fair Use: <https://www.forbes.com/sites/virginieberger/2025/03/15/the-ai-copyright-battle-why-openai-and-google-are-pushing-for-fair-use/>. Accessed: 2025-03-21.
- [46] The Linux Threat Landscape Report | Trend Micro (US): 2022. <https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/the-linux-threat-landscape-report>. Accessed: 2025-03-23.
- [47] Üstün, A. et al. 2024. Aya Model: An Instruction Finetuned Open-Access Multilingual Language Model. arXiv.
- [48] Viewing the CBC as a public good: <https://policyoptions.irpp.org/magazines/november-2016/viewing-the-cbc-as-a-public-good/>. Accessed: 2025-03-20.
- [49] Wiggers, K. 2024. OpenAI inks deal to train AI on Reddit data. *TechCrunch*.
- [50] Zhang, D. et al. 2022. The AI Index 2022 annual report. *Stanford Institute for Human-Centered AI, Stanford University*. 123, (2022).
- [51] 2024. Canada proposed \$15 bln incentive to boost AI green data centre investment, Globe and Mail reports. *Reuters*.
- [52] China's AI boom is reaching astonishing proportions. *The Economist*.
- [53] 2025. *Expanding Access to Public AI Models - The Aspen Institute Recorded Videos*.
- [54] 2024. *Mark Zuckerberg - Llama 3, \$10B Models, Caesar Augustus, & 1 GW Datacenters*.