Memo #3 - Building safe AI systems and public trust in AI

The analysis and positions expressed are independently developed by the author and do not represent any institutional views.

Executive Summary

Canada's cyber resilience is central to ensuring that AI supports the interests and values of its citizens. This memo proposes how Canada can treat trust as a strategic asset — not a constraint — and outlines why building trusted AI systems must be pursued proactively rather than reactively in a world with AI and quantum computing. Based on observations of the global innovation ecosystem, specifically in AI, quantum computing, and cybersecurity, recommendations to Minister Solomon's AI Task Force are:

1. Proactively innovate on trust and adopt AI- and quantum-secure technologies

- In parallel to the recommendations provided in *Memo #1*, Canada must be an early customer for AI- and quantum-secure cybersecurity solutions.
- Embed trust into the development of emerging technologies from the start.
- Treat trust (reliability, accountability, fairness) as a design requirement as technologies are being developed, not a compliance step after they are deployed.

2. Sovereignty through security and cyber resilience

- Leverage Canada's strengths in AI, quantum, and cybersecurity. The intersection and depth of research and skill in these areas is unique, and presents an opportunity for Canada to secure its digital infrastructure and establish global leadership in trusted AI. Scale champions and attract investment in these industries (recommendations provided in *Memo #2*).

3. Establish clear leadership of AI and quantum cyber resilience within the public sector

- Cyber resilience is mission critical to building a strong future for Canada. This is a dynamic, fast-moving space that requires constant landscape analysis and situational awareness. Identify a single entity in the federal government that is responsible for understanding emerging technologies, distributing trusted solutions and best practices across federal and provincial public services. This entity can also serve as a single point of contact for private Canadian enterprises.

Introduction

The rapidly improving performance of AI tools make cyberattacks faster and harder to detect. AI-based systems can identify vulnerabilities and exploit them without a human in the loop. The threat landscape is rapidly changing with AI available, and quantum computing maturing faster than most realize. For decades, cybersecurity has depended on encryption standards such as RSA-2048, which a conventional computer could break in ~300 trillion years. Google scientists estimate that a 1M qubit quantum computer could do this in a week [12].

Trust is becoming a scarce and costly resource in the era of AI and quantum computing. Each breach raises the perceived and actual cost of adopting emerging technologies. The accelerating pace of AI adoption underscores a key reality: trust can no longer be an afterthought built once technologies are mature — it must be a core design principle from the outset.

Trusted AI systems are those that demonstrate reliability, transparency, and accountability — where outcomes are explainable, performance is monitored, and failures can be traced and corrected. Building accountability into the technology itself is what enables lasting public trust. Canada has an opportunity to innovate on trust — embedding it into AI systems, quantum computing, and advanced processes, rather than waiting to rebuild after failures occur.

Be Proactive: Change Is Happening Quickly

AI adoption is far more rapid than prior technologies. While personal computers took years to reach even modest adoption, generative AI tools achieved widespread public use within about two years. ChatGPT reached 100 million users only 2 months after its initial public release. Like the early computer era, the full economy-wide impact will take time to materialize, but the speed and breadth at which AI is spreading suggest its ultimate transformation could be even larger and arrive much sooner than the computer revolution. Similarly, AI is spreading faster than the internet did, reaching users and widespread business experimentation within just a few years, whereas the internet took a decade or more to reach similar adoption levels globally.

In Canada, the adoption of artificial intelligence, the internet, and personal computers has unfolded at varying paces, each leaving a distinct mark on society and the economy (**Table 1**). AI adoption in Canada has accelerated rapidly in recent years. As of 2024, approximately 6.1% of Canadian businesses reported using AI in their operations [1]. By 2025, this figure nearly doubled, with 46% of business leaders indicating they were "definitely using" AI, up from 24% the previous year [2]. The most common AI applications include data analytics (26.4%), virtual agents or chatbots (24.8%), and marketing automation (23.1%) [3]. Notably, larger firms are adopting AI at higher rates, with 14% of businesses employing generative AI tools by 2024.

Table 1: Adoption Diffusion of computers, the internet, and AI in Canada

Technology	Speed of Adoption	Scale of Adoption	Societal Impact
AI	Rapid growth in recent years; ChatGPT launched in Nov 2022	14% of businesses adopted generative AI by 2024	Significant integration into workplaces; government incentives proposed
Internet	Accelerated in the late 1990s and early 2000s	94.3% of Canadians online by 2024	Fundamental to daily life; enables remote work and access to information
Computers	Increased adoption in the 1990s	Widespread in households and businesses	Transformed business operations and communication

Today: AI Creates Software

"Software is eating the world" by Marc Andreessen https://a16z.com/why-software-is-eating-the-world/

In 2011, Marc Andreessen argued that software was transforming every major industry, enabling new companies to disrupt long-standing incumbents and driving a major economic shift. With internet access now global, cloud infrastructure inexpensive, and billions of smartphone users, software-driven businesses like Amazon, Netflix, and Google were rapidly overtaking traditional firms across retail, entertainment, finance, transportation, and more. Andreessen saw this as a long-term trend, where Silicon Valley-style innovation will continue to "eat the world," creating huge opportunities for new companies while forcing others to adapt or fail.

In 2025, modern AI systems make probabilistic predictions and continuously adapt, which makes after-the-fact regulation impractical — by the time a rule is written, the underlying model may already have changed. These features are unlike earlier digital tools that performed static, rules-based functions, with predictable outcomes.

AI is accelerating the disruptive power of software. Open-source models, API-based platforms, and no-code tools mean that individuals and small teams can now deploy capabilities once reserved for major organizations in the public and private sector.

So today, AI is not just eating the world — its accessibility makes it such that anyone can do the eating. This same accessibility makes innovation faster, but it also means risks can scale at unprecedented speed if trust is not built into the technology from the start.

Pillars of Trust

Abraham Maslow posited in 1943 that humans have a hierarchy of psychological needs. This has been a foundational element of behavioural psychology for the past 80 years. Modern reinterpretations of humans' needs explicitly frame **"trust and attachment"** as core evolved mechanisms that enable social cooperation — making trust a *functional prerequisite* to higher needs. Trust among humans has been studied extensively and found to critically depend on several features:

1. **Predictability & Reliability:** Humans only trust those whose actions are consistent and predictable. Erratic or unpredictable behaviour quickly destroys trust.
2. **Competence**: Is someone capable of doing what they claim?
3. **Transparency and Accountability:** Openness about motives, limits, and mistakes reinforces credibility. Accountability mechanisms (social, legal, reputational) maintain it.
4. **Benevolence:** Showing good intentions
5. **Integrity**: Trust depends on alignment between values and actions.
6. **Shared Values**: Trust grows faster within groups that share goals, norms, or identities

Humans apply these same features to systems and institutions. A useful counter example of building institutional trust is media. StatsCan reports that in 2023 trust in media was 21% (Quebec), 15% (Atlantic), 13% (Ontario & British Columbia) and 12% (Prairies). Key drivers include perceived bias, concerns about media independence, and the challenges of the modern information environment (misinformation, online news fragmentation). Transparency, accountability, and integrity are perceived to be very low, contributing to low trust.

Algorithms play an integral role in public perception of trust in online media, particularly social media. AI greatly expands the ability to influence incoming content to individuals about news, social media, products and services. AI is therefore starting from a position of suspicion by a considerable proportion of Canadians. A 2025 KPMG survey found that only 34% of Canadians trust AI, 83% are concerned about it being used for misinformation and disinformation, and 79% of Canadians are concerned about AI producing negative outcomes.

Trust and cybersecurity for AI systems, as well as quantum readiness are competitive advantages that Canada must actively cultivate. Organizations that can guarantee data integrity, system reliability, and AI accountability will lead global markets increasingly defined by trust.

Key Takeaways and Recommendations

Building trusted AI systems is not a regulatory afterthought — it is Canada's opportunity to lead. By advancing trust proactively as an innovation, not a reaction, Canada can define how intelligent systems serve society and strengthen public confidence in the technologies shaping our future.

1. Proactively innovate on trust and adopt AI- and quantum-secure technologies

- In parallel to the recommendations provided in *Memo #1*, Canada must be an early customer for AI- and quantum-secure cybersecurity solutions.
- Embed trust into the development of emerging technologies from the start.
- Treat trust (reliability, accountability, shared values) as a design requirement as technologies are being developed, not a compliance step after they are deployed.

2. Sovereignty through security and cyber resilience

- Leverage Canada's strengths in AI, quantum, and cybersecurity. The intersection and depth of research and skill in these areas is unique, and presents an opportunity for Canada to secure its digital infrastructure and establish global leadership in trusted AI. Scale champions and attract investment in these industries (recommendations provided in *Memo #2*).

3. Establish clear leadership of AI and quantum cyber resilience within the public sector

- Cyber resilience is mission critical to building a strong future for Canada. This is a dynamic, fast-moving space that requires constant landscape analysis and situational awareness. Establish a single entity in the federal government that is responsible for understanding emerging technologies, distributing trusted solutions and best practices across public services. This entity can also serve as a single point of contact for private Canadian enterprises.

References

1. Statistics Canada. *Analysis on artificial intelligence use by businesses in Canada*. June 16, 2025. Available from: https://www150.statcan.gc.ca/n1/pub/11-621-m/11-621-m2025008-eng.htm
2. Public First. *Exploring Canada's AI Opportunity: Google Canada Economic Impact Report 2024*. October 7, 2025. Available from: https://www.publicfirst.co.uk/exploring-canadas-ai-opportunity-google-canada-economic-impact-report-2024.html
3. Business Data Lab. *Generative AI Adoption by Canadian Businesses*. 2024. Available from: https://businessdatalab.ca/publications/prompting-productivity-generative-ai-adoption-by-canadian-businesses/
4. Reuters. *Canada proposed $15 bln incentive to boost AI green data centre investment*. December 12, 2024. Available from: https://www.reuters.com/technology/artificial-intelligence/canada-proposed-15-bln-incentive-boost-ai-green-data-centre-investment-globe-2024-12-12/
5. Statistics Canada. *Household Internet Use Survey*. September 18, 2003. Available from: https://www150.statcan.gc.ca/n1/daily-quotidien/030918/dq030918b-eng.htm
6. DataReportal. *Digital 2024: Canada*. February 22, 2024. Available from: https://datareportal.com/reports/digital-2024-canada
7. Kurt's Copy. *Canadian Internet Statistics 2024: Insights with a Global Perspective*. December 11, 2024. Available from: https://kurtscopy.com/blog/canadian-internet-statistics-2024-insights-with-a-global-perspective/
8. Baldwin JR, Sabourin D. *Growth of Advanced Technology Use in Canadian Enterprises*. 1999. Available from: https://www150.statcan.gc.ca/n1/pub/11f0019m/11f0019m1999105-eng.pdf
9. Bitler M. *Small business and computers: Adoption and performance*. 2001. Available from: https://www.frbsf.org/wp-content/uploads/wp01-15bk.pdf
10. The Canadian Encyclopedia. *Digital Economy in Canada*. October 24, 2019. Available from: https://thecanadianencyclopedia.ca/en/article/digital-economy-in-canada
11. World Health Organization. Ethics and governance of artificial intelligence for health. 2021. Available from: https://www.who.int/publications/i/item/9789240029200
12. Lambert L, Mosca M. *Security delayed is security denied.* 2025. Available from: https://www.hilltimes.com/story/2025/10/31/security-delayed-is-security-denied/480693/