Intelligence MEMOS

From: Charles Eagan

To: AI Observers

Date: November 11, 2024

Re: A Path to Trusted AI

Artificial Intelligence (AI) has infiltrated our lives for decades, but since the public launch of ChatGPT showcasing generative AI in 2022, society has faced unprecedented technological evolution.

With digital technology already a constant part of our lives, AI has the potential to alter the way we live, work, and play – but exponentially faster than conventional computers have. With AI comes staggering possibilities for both advancement and threat.

The AI industry creates unique and dangerous opportunities and challenges. AI can do amazing things humans can't, but in many situations, referred to as the black box problem, experts cannot explain why particular decisions or sources of information are created. These outcomes can, sometimes, be inaccurate because of flawed data, bad decisions or infamous AI hallucinations. There is little regulation or guidance in software and effectively no regulations or guidelines in AI.

How do researchers find a way to build and deploy valuable, trusted AI when there are so many concerns about the technology's reliability, accuracy and security?

That was the subject of a recent C.D. Howe Institute conference. In my keynote address, I commented that it all comes down to software. Software is already deeply intertwined in our lives, from health, banking, and communications to transportation and entertainment. Along with its benefits, there is huge potential for the disruption and tampering of societal structures: Power grids, airports, hospital systems, private data, trusted sources of information, and more.

Consumers might not incur great consequences if a shopping application goes awry, but our transportation, financial or medical transactions demand rock-solid technology.

The good news is that experts have the knowledge and expertise to build reliable, secure, high-quality software, as demonstrated across Class A medical devices, airplanes, surgical robots, and more. The bad news is this is rarely standard practice.

As a society, we have often tolerated compromised software for the sake of convenience. We trade privacy, security, and reliability for ease of use and corporate profitability. We have come to view software crashes, identity theft, cybersecurity breaches and the spread of misinformation as everyday occurrences. We are so used to these trade-offs with software that most users don't even realize that reliable, secure solutions are possible.

With the expected potential of AI, creating trusted technology becomes ever more crucial. Allowing unverifiable AI in our frameworks is akin to building skyscrapers on silt. Security and functionality by design trump whack-a-mole retrofitting. Data must be accurate, protected, and used in the way it's intended.

Striking a balance between security, quality, functionality, and profit is a complex dance. The BlackBerry phone, for example, set a standard for secure, trusted devices. Data was kept private, activities and information were secure, and operations were never hacked. Devices were used and trusted by prime ministers, CEOs and presidents worldwide. The security features it pioneered live on and are widely used in the devices that outcompeted Blackberry.

Innovators have the know-how and expertise to create quality products. But often the drive for profits takes precedence over painstaking design. In the AI universe, however, where issues of data privacy, inaccuracies, generation of harmful content and exposure of vulnerabilities have far-reaching effects, trust is easily lost.

So, how do we build and maintain trust? Educating end-users and leaders is an excellent place to start. They need to be informed enough to demand better, and corporations need to strike a balance between caution and innovation.

Companies can build trust through a strong adherence to safe software practices, education in AI evolution and adherence to evolving regulations. Governments and corporate leaders can keep abreast of how other organizations and countries are enacting policies that support technological evolution, institute accreditation, and financial incentives that support best practices. Across the globe, countries and regions are already developing strategies and laws to encourage responsible use of AI.

Recent years have seen the creation of codes of conduct and regulatory initiatives such as:

- Canada's Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems, September 2023, signed by AI powerhouses such as the Vector Institute, Mila-Quebec Artificial Intelligence Institute and the Alberta Machine Intelligence Institute;
- The Bletchley Declaration, Nov. 2023, an international agreement to cooperate on the development of safe AI, has been signed by 28 countries;
- US President Biden's 2023 executive order on the safe, secure and trustworthy development and use of AI; and
- Governing AI for Humanity, UN Advisory Body Report, September 2024.

We have the expertise to build solid foundations for AI. It's now up to leaders and corporations to ensure that much-needed practices, guidelines, policies and regulations are in place and followed. It is also up to end-users to demand quality and accountability.

Now is the time to take steps to mitigate AI's potential perils so we can build the trust that is needed to harness AI's extraordinary potential.

Charles Eagan is the former CTO of Blackberry and a technical advisor to AIE Inc.

To send a comment or leave feedback, email us at blog@cdhowe.org.

The views expressed here are those of the authors. The C.D. Howe Institute does not take corporate positions on policy matters.