TRUST IN AI IS POSSIBLE BUT CERTAIN PRINCIPLES NEED TO BE FOLLOWED

GRI | GLOBAL RISK INSTITUTE

The increasing use of artificial intelligence (AI) and its integration into our daily lives has created a sense of urgency for organizations to ensure they are harnessing its benefits while also mitigating its risks.

This is particularly important in Canada's financial services industry, where AI has the potential to revolutionize the sector with more personalized products and better customer service, while enhancing risk management, cybersecurity and asset allocation.

To that end, the Office of the Superintendent of Financial Institutions (OSFI) and the Global Risk Institute brought together experts from Canada's financial services sector, government and academia. Together, we listened to AI experts and influential regulators from several countries and debated the learnings.

We developed a framework for the financial services sector that would help institutions manage risks more effectively. Institutions should use the framework to analyze and mitigate the risks associated with AI, which will allow for better risk reward assessments.

To gain a competitive edge, organizations must safeguard public and consumer confidence and trust while using the new technology. This requires examining the ethical, legal, financial, and social implications of AI's development, deployment and oversight. Organizations must prioritize the safe and responsible use of AI to maintain public trust, and in turn, secure their long-term success in the industry.

The framework we developed concentrated on the need for guardrails that address explainability, data, governance and ethics, or the "EDGE principles". This lens is universal in its application, agnostic of the underlying technology, and any organization can follow this approach with AI to fortify their business model deployment.

Regulators, as well as customers and investors, are already questioning the risks of AI. One of the most prominent and persistent challenges of AI is the potential difficulty of explaining how those models reach a conclusion. To address concerns about the difficulty of explaining AI models, organizations should be certain their models are explainable, which will promote fairness, transparency and accountability, and generally help to protect against unintended bias and discrimination. In other words, when customers understand the reason for the AI response or decision, they become empowered to act upon it.

Organizations also need to address the issue of bias in AI decision making. If the data being used to train AI models is incomplete or inaccurate for the task at hand, then bias will be intrinsically present. Advanced data management with human oversight will help ensure data is accurate, complete, consistent, and private, which is crucial for AI systems to work effectively in financial services.

A robust governance framework within an organization promotes a culture of responsibility and accountability around the use of AI as society's expectations of financial institutions maintain high ethical standards increase. Organizations should consider disclosing how they maintain standards for their AI models to promote internal and external

transparency. This will allow financial institutions to fully realize the benefits made possible by AI while building trust with their customers.

It is this need for trust that will see the application of AI in Canada's financial institutions target use cases that lead to faster decision-making and better customer service outcomes, rather than a release of control to autonomous machines.

Effective governance structures are essential for ensuring that AI models make the right decisions for the right reasons, and that intervention points are in place to prevent unintended consequences.

This will help to maintain trust and confidence in AI systems, while enabling organization to achieve their business objectives through the use of AI.

Regulators must strike a balance between regulation and innovation. For example, OSFI will release policy guidelines that will include AI applications in their enhanced E-23 Enterprise Model Risk Management Guideline, due for public consultation later in 2023. In the interim, the financial services industry should implement the spirit and intent of the EDGE principles with a sense of urgency to ensure they meet (or hopefully) exceed their stakeholder expectations.