CHARTERED
PROFESSIONAL
ACCOUNTANTS
CANADA

Building a Risk Management Framework for Trustworthy AI

January 2022

CPA Canada
Foresight
REIMAGINING THE PROFESSION.

22-2315

ABOUT CPA CANADA

Chartered Professional Accountants of Canada (CPA Canada) works collaboratively with the provincial, territorial and Bermudian CPA bodies, as it represents the Canadian accounting profession, both nationally and internationally. This collaboration allows the Canadian profession to champion best practices that benefit business and society, as well as prepare its members for an ever-evolving operating environment featuring unprecedented change. Representing more than 220,000 members, CPA Canada is one of the largest national accounting bodies worldwide. cpacanada.ca

Electronic access to this report can be obtained at **cpacanada.ca**.

Table of Contents

Weather forecasts, e-mail spam filtering, Google's search predictions and voice recognition machines, such as Apple's Siri, are all examples of AI systems that are adding remarkable value to businesses, consumers and society in general.

Technologies that use machine-learning algorithms to react and respond in real time without human intervention are already improving business productivity; future growth prospects are nothing short of mind boggling. A survey by management consultancy McKinsey estimated that AI analytics could add around US13 trillion or 16 per cent of annual global GDP by 2030.[1]

[1] McKinsey Global Institute. Notes From the Ai Frontier: Modeling the Impact of AI on the World Economy. McKinsey. September 2018, 61 pages. https://www.mckinsey.com/~/media/mckinsey/featured%20insights/artificial%20intelligence/notes%20from%20the%20frontier%20modeling%20the%20impact%20of%20ai%20on%20the%20world%20economy/mgi-notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy-september-2018.pdf?shouldIndex=false

As impressive as that sounds, however, deploying AI systems comes with its share of risks – risks that organizations in the midst of their own digital transformation process must be aware of. Misused, AI systems can provide novel and powerful tools for manipulative, exploitative and social control practices. In response, recent developments particularly in the U.S., and E.U. are expected to frame the ethical use of AI systems which will drive demand for risk management frameworks to guide development, testing and use of AI.

Rachel Kirkham, VP Analytics and Data Science at MindBridge points out, "There's lots of evidence about the potential harm of improper use of this technology, so now is the time for people to put appropriate frameworks in place to manage this from a corporate risk perspective as well as a regulatory perspective."[2]

Accountants play a role in the collection, analysis, interpretation and provision of information for decision-making processes that help both internal and external stakeholders to understand and influence performance drivers. By providing relevant information from risk assessments,[3] they also link risk to business performance indicators. As such, CPAs are well positioned to help design and implement systems and controls to achieve trustworthy AI.

This primer proposes concrete actions to make progress on that front.

[2] CPA Canada, Foresight Podcast https://www.cpacanada.ca/en/foresight-initiative/podcast/cpas-double-edged-sword-technology

[3] CPA Canada and IFAC, From Bolt-On to Built-In: Managing Risk As An Integral Part Of Managing An Organization, May 2015. https://www.cpacanada.ca/en/business-and-accounting-resources/strategy-risk-and-governance/enterprise-risk-management/publications/from-bolt-on-to-built-in-managing-organizational-risk

Addressing the growing trust deficit

Trust is the most powerful force underlying the success of every business – and it can be shattered in an instant. When looking at trust in digital technologies, all indications are that we are nearing a breaking point. From systemic data misuse by big tech platforms, to fake news designed to create division and conflict, to privacy breaches and ransomware attacks, the tech sector seems to be caught in a growing trust maelstrom. Increasingly, organizations undergoing digital transformation processes are finding out they can be impacted by this trust deficit, often with grave implications. Prejudices flowing from flawed datasets can be baked into algorithms, resulting in decisions that can harm organizations, their stakeholders and society as a whole.

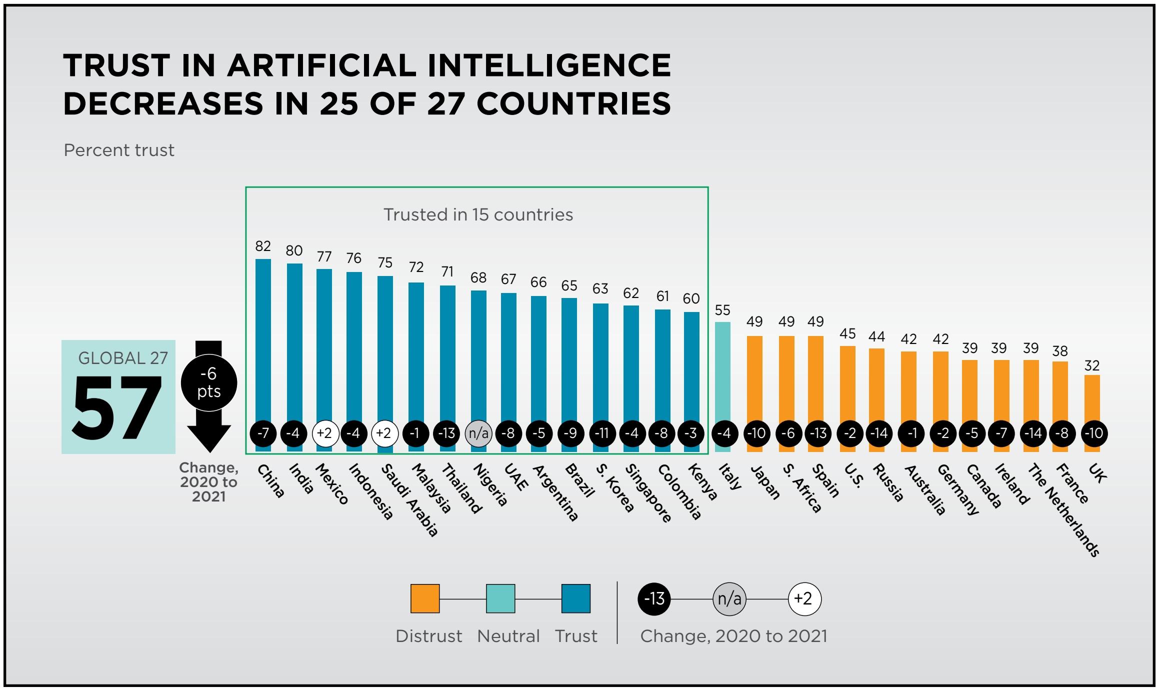
Once a system is developed, it's too late to ask the truly vital questions such as "Should this be system built at all?" and "How do we safeguard against bias and ensure fairness?" The proverbial genie is, at that point, out of the bottle, especially if an ethics shortcoming is discovered after a significant amount of time, money and creative energy has been invested in development.[4]

The 2021 Edelman Trust Barometer, an annual survey of 33,000 people in 27 countries, found that trust in AI decreased in 25 out of 27 countries over a period of a year.[5] In Canada, trust in AI fell by five percentage points to reach 39 per cent. Consumers have growing concerns about harm that AI can inflict on vulnerable populations because of the opacity, complexity, bias, unpredictability and the partially autonomous behaviors of certain AI systems.

[4] CPA Canada, in collaboration with IFAC, ICAS and IESBA, Technology is a double-edge sword with both opportunities and challenges for the accountancy profession" https://www.cpacanada.ca/en/foresight-initiative/trust-and-ethics/technology-double-edged-sword

[5] The Edelman Annual Trust Report's survey methodology can be accessed on page 2 and in relevant annexes of the 2021 Trust Barometer Report: https://www.edelman.com/sites/g/files/aatuss191/files/2021-03/2021%20Edelman%20Trust%20Barometer.pdf

Figure 1



Source: 2021 Edelman Trust Barometer. https://www.edelman.com/sites/g/files/aatuss191/files/2021-03/2021%20Edelman%20Trust%20Barometer%20Tech%20Sector%20Report_0.pdf p. 44

> "In light of the increasing importance of trust in the digital age, there are greater expectations for leaders – including professional accountants – to be accountable and act responsibly. Failure to do so could be seen as a lack of integrity and could discredit the profession under the principle of professional behaviour."
>
> Excerpt from *Technology is a double-edge sword with both opportunities and challenges for the accountancy profession*; CPA Canada, in collaboration with IFAC, ICAS and IESBA

What is trustworthy AI?

In response to trust concerns, governments, industry and civil society have outlined new approaches to manage risks associated with AI systems. An outdated concept of ethical AI, developed to frame the deployment of AI, has been replaced by a broader framework encompassing "trustworthy AI".[6] According to William Diab, a world expert on AI systems who helped develop a new ISO standard on trustworthy AI, "Every customer – whether it's a financial services company, whether it's a retailer, whether it's a manufacturer – is going to ask: 'Who do I trust?' Many aspects including societal concerns, such as data quality, privacy, potentially unfair bias and safety must be addressed."[7]

There have been many declarations and statements on ethical and trustworthy AI through the years.[8] Nowadays, trustworthy AI typically embodies the following concepts:

- **accuracy** – AI should make the right decisions.
- **explainability** – The process used by the system to make decisions should be documented, understood and replicated by humans.

[6] For example, new regulations framing AI are under development in the E.U. through the Artificial Intelligence Act: A Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence. https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/AUTRES_INSTITUTIONS/COMM/COM/2021/06-02/COM_COM20210206_EN.pdf. A new voluntary risk management framework for AI is being developed by the National Institute of Standards and Technology in response to an Executive Order from the White House. https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf. Internationally, new voluntary standards regarding the trustworthiness of AI have been published by organizations such as the International Organization for Standardization (ISO) https://www.iso.org/standard/77608.html?browse=tc. The Institute of Electrical and Electronics Engineers (IEEE) is also in the process of developing a comprehensive series of standards under its 7000 series under its Ethically Aligned Design initiative. Issues being tackled range from transparency, data privacy processes and algorithmic bias considerations to child, student and employer data governance. https://ethicsstandards.org/p7000/. Regarding Canadian voluntary standards framing digital governance, the CIO Strategy Council has published a standard on ethical AI entitled "Artificial Intelligence: Ethical design and use of automated decision systems." https://ciostrategycouncil.com/standards/1012019/

[7] https://www.iso.org/news/ref2530.html

[8] Among the many declarations and statements on this issue, one notes the Asilomar Principles proposed by the Future of Life Institute https://futureoflife.org/ai-principles/; The Open Data Charter https://opendatacharter.net/principles/; The 2017 Montréal Declaration for a Responsible Development of Artificial Intelligence. https://www.montrealdeclaration-responsibleai.com/the-declaration; and the Top Ten Principles for Ethical AI www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf; UNESCO member states adopt global agreement on Ethics of Artificial Intelligence. https://www.unesco.org/en/articles/unesco-member-states-adopt-first-ever-global-agreement-ethics-artificial-intelligence

- **resiliency** – When new data causes an AI system to operate outside of its nominal boundaries, it should be able to adapt to new conditions or to alert humans in order to avoid catastrophic failure.
- **safety** – Systems should not create health or safety hazards to humans or the environment.
- **reliability** – AI systems should be designed to operate continuously and consistently.
- **objectivity** – AI systems should be devoid of prejudice or bias against individuals or groups.
- **inclusivity in growth, sustainable development and well-being** – AI systems should contribute to beneficial outcomes for people and the planet including fundamental values such as democratic rights, fairness and privacy.

Three categories of AI systems

AI systems are not created equal. Some, like algorithms that propose music playlists and movie suggestions through streaming services or those that promote products through online advertising, can generally be managed without much concern about creating harm. On the other hand, AI systems making life and death decisions such as self-driving vehicles or critical infrastructure operations require a high degree of oversight. Recent advances in Europe point to segmenting AI according to various categories to focus scarce government/regulatory resources on high-risk AI systems. In its recently tabled regulation on trustworthy AI, the European Commission outlines a risk-based approach modulated on three categories of AI systems. These categories can serve as useful guidance to organizations aiming for trustworthy AI.

The first category encompasses **"unacceptable" high-risk AI systems** that may contravene laws or violate fundamental rights. Examples of unacceptable high-risk AI include:

- practices that have a potential to manipulate individuals through subliminal techniques that are beyond their consciousness, that distort human behaviour or exploit vulnerable groups such as children or persons with disabilities
- systems that create social scoring of persons or evaluate or classify the trustworthiness of persons based on their social behaviour

It is expected that unacceptable high-risk systems will be prohibited from use in the E.U.

The second category encompasses **high-risk AI systems that can be managed**. AI systems that create high health and safety risk or could threaten fundamental rights and freedoms fall into this category. Examples include:

- AI systems intended to be used as a safety component of products already covered by public safety regulations, including electrical, plumbing, pressure vessels, heating and cooling equipment, elevators, toys, worker safety equipment, radio equipment and equipment used in dangerous environments. In addition, it is expected that a wide range of new product categories, including autonomous robots in manufacturing and personnel assistance and care; health-care diagnostics and systems supporting health-care decisions based on sophisticated autonomous AI systems will be included.
- AI systems that may impact the right to human dignity; private and family life; personal data discrimination as well as other rights and freedoms. AI systems that may impact rights and freedoms can be found across a wide variety of sectors including finance, credit ratings, insurance, education, human resources management (in functions such as such as recruitment and hiring), law enforcement and administrative proceedings and the administration of justice and democratic processes.

The third category includes all other AI systems which are deemed **low risk and can be deployed without severe constraints**. Regulators recommend voluntary codes of practices be put in place to ensure low-risk AI systems remain safe.

Managing high-risk AI

> "Various risks that can impact information integrity exist throughout the information lifecycle and increase the possibility of material errors and omissions in information leading to erroneous or sub-optimal decisions arising from the use of the information."
>
> Excerpt from CPA Canada's *A Framework for Information Integrity Controls*

Organizations deploying AI systems must take steps to manage the associated risks and therefore carry out trustworthy AI. The steps proposed below are drawn from recent regulatory and standards initiatives in the U.S., Europe and Canada.[9]

- **implement an accountability framework.** Organizations should ensure that AI systems meet the characteristics of trustworthy AI throughout their lifecycle. This may include developing, testing, deploying, operating, upgrading and decommissioning phases of current AI systems. It should be noted that the new E.U. regulation will require that organizations deploying high-risk AI systems designate a provider to manage this accountability framework.
- **use existing quality management systems** to track and report on trustworthy AI. The E.U. regulation will require organizations to set up appropriate management systems such as ISO 9001 whenever high-risk AI systems are developed, used or sold. Making a commitment to achieve

[9] For example, new regulations framing AI are under development in the E.U. through the Artificial Intelligence Act: A Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence. https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/AUTRES_INSTITUTIONS/COMM/COM/2021/06-02/COM_COM20210206_EN.pdf
A new voluntary risk management framework for AI is being developed by the National Institute of Standards and Technology in response to an Executive Order from the White House. https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf
Internationally, new voluntary standards regarding the trustworthiness of AI have been published by organizations such as the International Organization for Standardization (ISO) https://www.iso.org/standard/77608.html?browse=tc. The Institute of Electrical and Electronics Engineers (IEEE) is in the process of developing a comprehensive series of standards under its 7000 series under its Ethically Aligned Design initiative. Issues being tackled range from transparency, data privacy processes and algorithmic bias considerations to child, student and employer data governance. https://www.iso.org/standard/77608.html?browse=tc
Regarding Canadian voluntary standards framing AI and machine learning, the CIO Strategy Council is working on a series of digital governance standards, including AI. Recent publications of interest include "Artificial Intelligence: Ethical design and use of automated decision systems". https://ciostrategycouncil.com/standards/101_2019/

quality and continuous improvement benchmarks for AI trustworthiness through a quality management framework will allow your organization to set goals, document efforts, benchmark performance across supply chains and manage risks.[10] Alternatively, AI performance can be tracked through the enterprise **risk management systems**, strategies and processes. Risks associated with AI systems can be documented and tracked through management systems using standards such as ISO 31000.[11]

- **set policies and procedures to manage AI systems.** Procedures are necessary to foster high quality data sets, upkeep technical documentation, record keeping and archiving of datasets. This may require adjustments to your corporate data policy to reflect a new accountability framework.[12]
- **create and maintain an inventory of AI systems** currently operating and those under development. This will allow your organization to react should a problem occur, for example if faulty data has been used to train multiple algorithms.
- **classify AI systems by category.** As outlined above, AI systems can be flagged as *unacceptably high risk; high risk; and low risk.*
- **consider using alternatives to unacceptably high-risk AI systems.** As noted previously, the use of AI systems that can nudge or manipulate people will be made illegal in the E.U. As such, it is expected that organizations doing business with the European single market will be required to abide by these new restrictions.
- **plan for human oversight** of high-risk AI systems in the organization. As outlined above, your organization should avoid using AI systems that make autonomous decisions impacting health and safety without some form of human oversight. To reduce organizational risks avoid so called "black box" algorithms where decisions cannot be explained or verified by humans. And ensure that those who are accountable for human oversight of AI systems have the necessary competencies, training and authority to carry out that role.
- **consider creating an advisory committee or board on AI systems**. The committee can be empowered to identify potential risks from unintentional, unanticipated, or harmful outcomes that may arise from intended uses and misuses of all AI systems, including low risk AI. It can also provide guidance on the use of trustworthy AI principles when AI systems are being designed

[10] https://www.iso.org/iso-9001-quality-management.html

[11] https://www.iso.org/iso-31000-risk-management.html

[12] https://www.cpacanada.ca/en/foresight-initiative/data-governance/mastering-data/corporate-data-policy-and-its-elements

and deployed.[13] Canadian standards such as the CIO Strategy Council's National Standard of Canada (NSC) on the ethical design and use of automated decision systems contain valuable guidance on the creation and operation of ethical AI advisory committees or boards.[14]

- **consider third-party attestation of high-risk AI systems** before deployment. The CIO Strategy Council's standard on automated decision systems contains clauses allowing for assessors to conduct ethical impact assessments of AI systems. Through attestation or direct engagements, organizations can obtain assurance on the entity's claim of conformity to digital governance standards, including the CIO Strategy Council's NSC on automated decision systems. Under the Canadian Standards on Assurance Engagements (e. g. CSAE 3000 and CSAE 3001) published in the *CPA Canada Handbook – Assurance*, CPAs can undertake such assurance engagements and obtain limited or reasonable assurance for AI system developers or users.[15]
- **focus on data quality.** Although it is recognized that prejudices can be baked into algorithms, they can also appear in incomplete datasets. Organizations deploying AI systems need to ensure that datasets are of high quality and fit for purpose.
- **aim for transparency.** Your organization should systematically inform users and customers when they interact with AI or bots, or when decisions affecting them are chiefly based on AI systems. Consider establishing a recourse mechanism for consumers to use in the event of disagreement regarding a decision generated by an AI system.[16]

[13] For example, the OECD, through its Council on Artificial Intelligence, issued recommendations on trustworthy AI in 2019 https://www.oecd.org/digital/artificial-intelligence/.

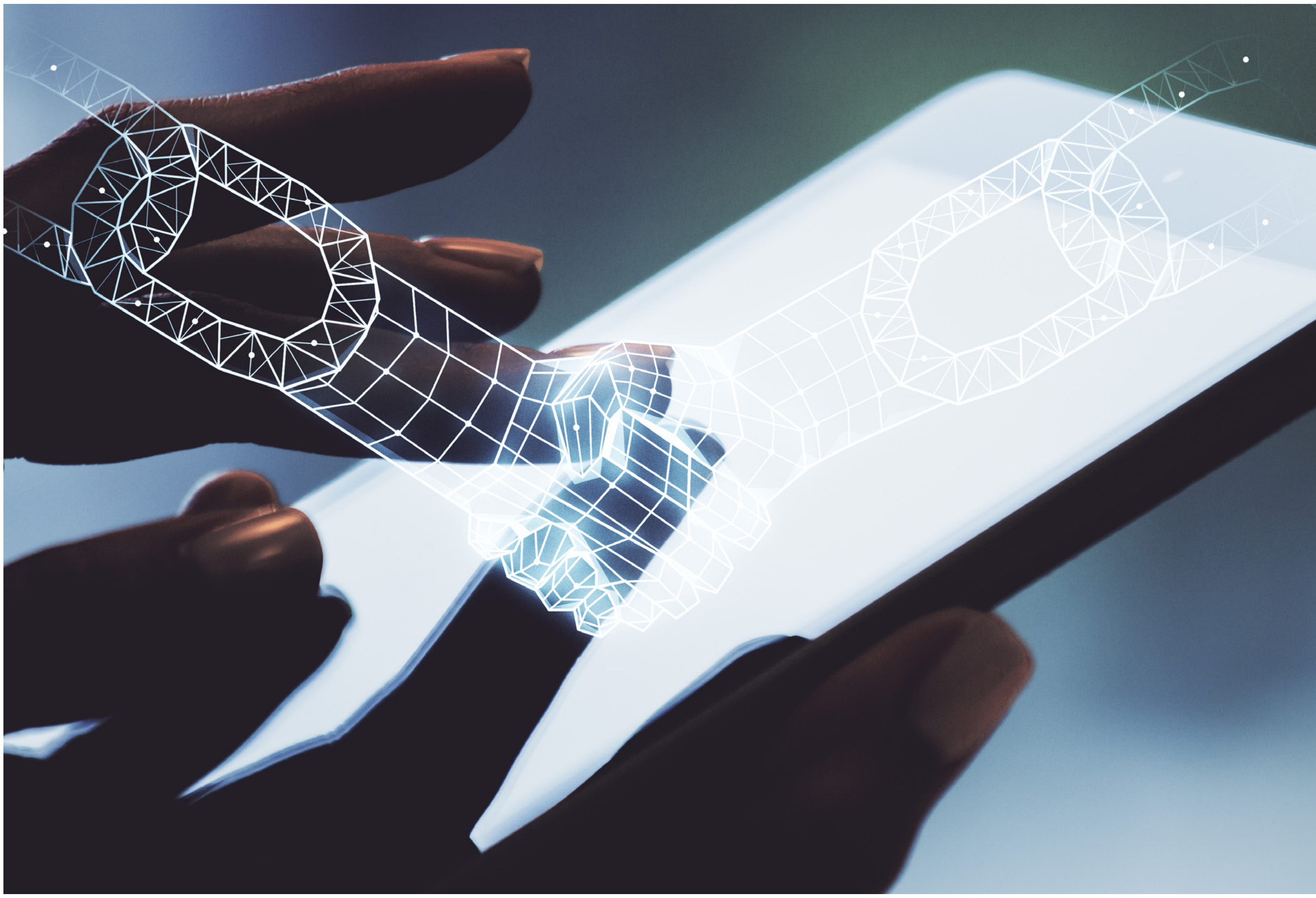
[14] CIO Strategy Council. CAN/CIOSC 101:2019 National Standard of Canada, corrected version, 2020-09, Ethical design and use of automated decision systems. 2020-09. 23 pages. https://ciostrategycouncil.com/standards/

[15] What auditors need to know about attestation engagements and direct engagements (cpacanada.ca). https://www.cpacanada.ca/en/business-and-accounting-resources/audit-and-assurance/standards-other-than-cas/publications/attestations-direct-engagements-auditors-should-know.

[16] Creating high-quality data to reach your digital transformation goals (cpacanada.ca) https://www.cpacanada.ca/en/foresight-initiative/data-governance/quality-data-vital

Looking forward

Organizations planning to deploy high risk AI systems need an appropriate risk management framework. Accountants are accustomed to designing, planning, implementing, and monitoring risk management programs. They can play a vital role in helping bridge the gap between data and trust. By managing risks that encourage trustworthy AI, accountants can help position their organizations to manage risks and make progress towards successful digital transformation.



CHARTERED
PROFESSIONAL
ACCOUNTANTS
CANADA

277 WELLINGTON STREET WEST
TORONTO, ON   CANADA   M5V 3H2
T. 416 977.3222   F. 416 977.8585
CPACANADA.CA