McCarthy Tétrault *Advance*™
Building Capabilities for Growth

Responsible AI – From Principles to Governance

Tuesday, November 26, 2019

Jennifer Reynolds, President & CEO, Toronto Finance International

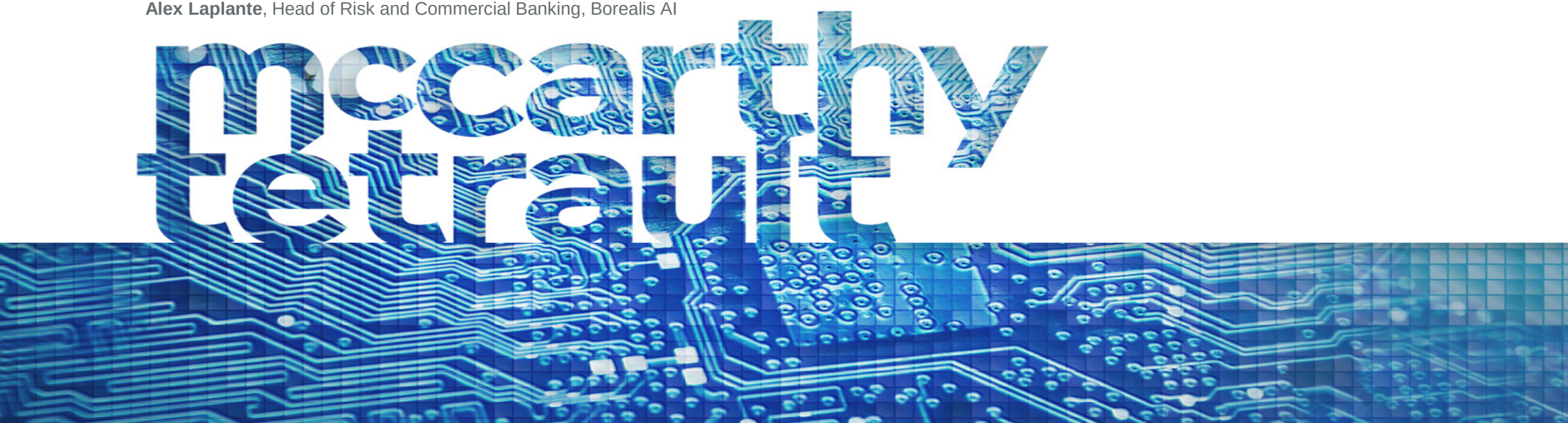
Charles Morgan, Partner, McCarthy Tétrault LLP

Christine Ing, Partner, McCarthy Tétrault LLP

Naresh Mudunuru, AVP Enterprise Machine Learning, TD Bank

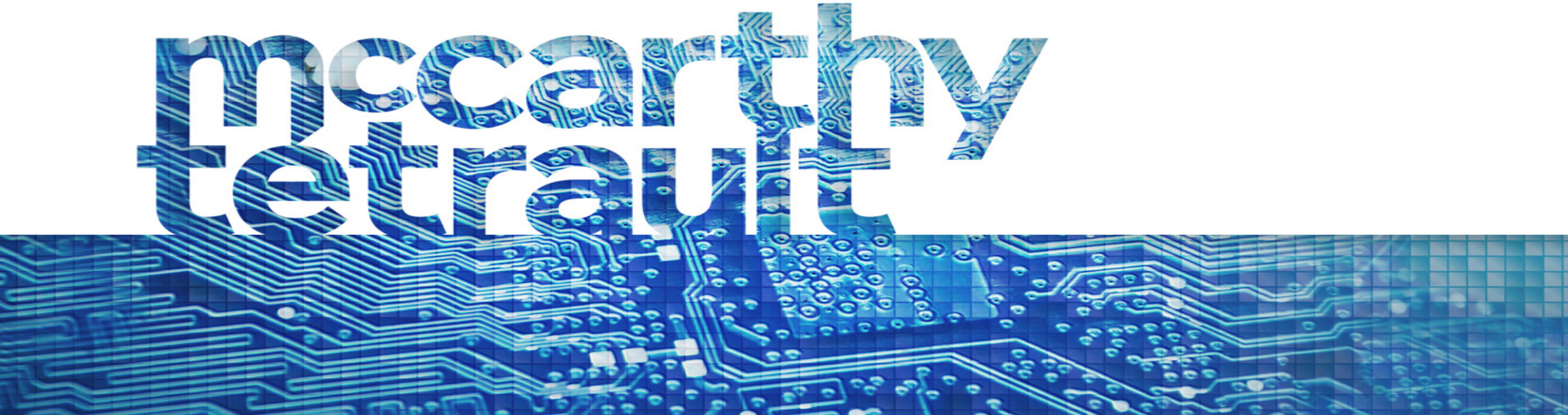
John Colthart, General Manager, MindbridgeAi

Alex Laplante, Head of Risk and Commercial Banking, Borealis AI



Taking Responsibility for Responsible AI

Charles Morgan
Partner, McCarthy Tétrault LLP

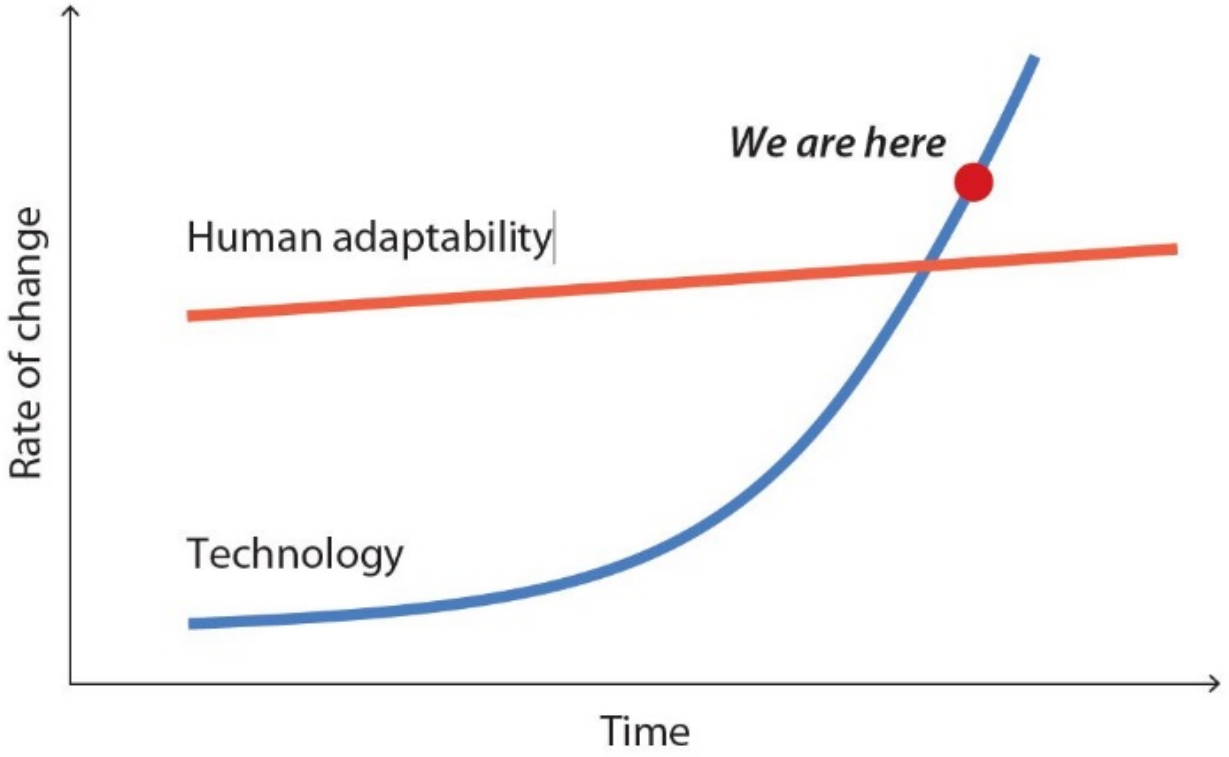


Presentation

- Context
- Principal Sources of Inspiration
- The Principles

Context

- Cambridge Analytica
- Microsoft's *Future Computed*
- Thomas Friedman's *Thank You for Being Late*



Principal Sources of Inspiration

- Principles are based in part on a growing consensus:
  - Accountability
  - Human-centered approach
  - Limiting biases
  - Explainability

- Inspired by
  - The Future Computed
  - Montreal Declaration Responsible AI
  - European Commission High-Level Expert Group on AI's *Ethics Guidelines for Trustworthy AI*
  - Singapore's *Proposed Model AI Governance Framework*

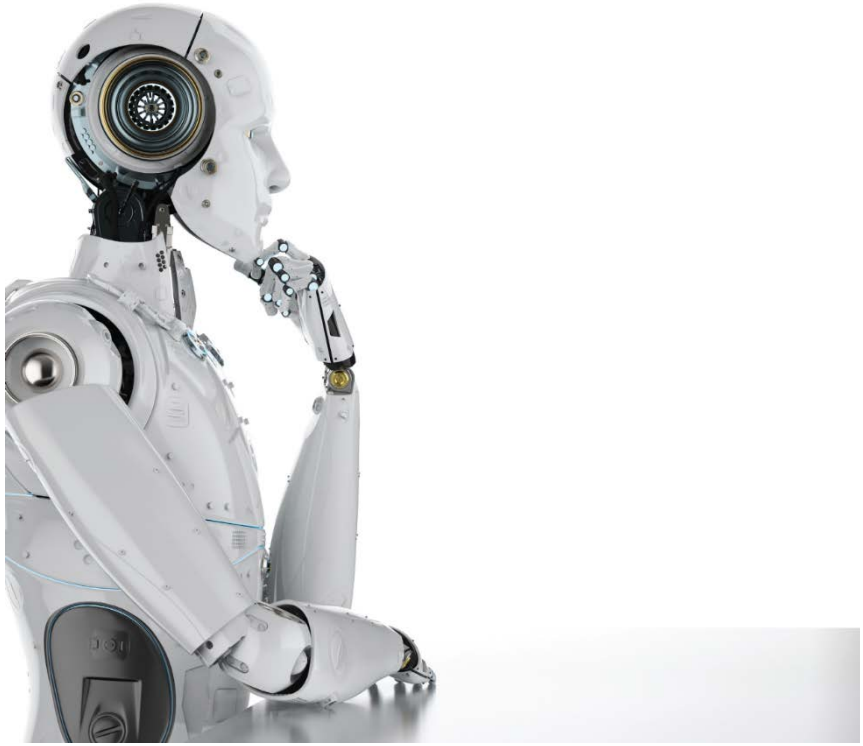
The Contributors

- 54 contributors
- 16 countries
- 5 continents
- 27 law firms
- Academics (Oxford, MIT, Berkeley, McGill)
- Industry representatives (Google Deepmind, Thomson Reuters, National Bank, ABB)

The Principles

1. Ethical Purpose and Societal Benefit
2. Accountability
3. Transparency and Explainability
4. Fairness and Non-discrimination
5. Safety and Reliability
6. Open Data and Fair Competition
7. Privacy
8. AI Rights and Intellectual Property

Principle 1: Ethical Purpose and Societal Benefit



Organisations that develop, deploy or use AI systems should require the purposes of such implementation to be identified and ensure that such purposes are consistent with the overall ethical purposes of beneficence and non-maleficence, as well as the other principles of the Policy Framework for Responsible AI.

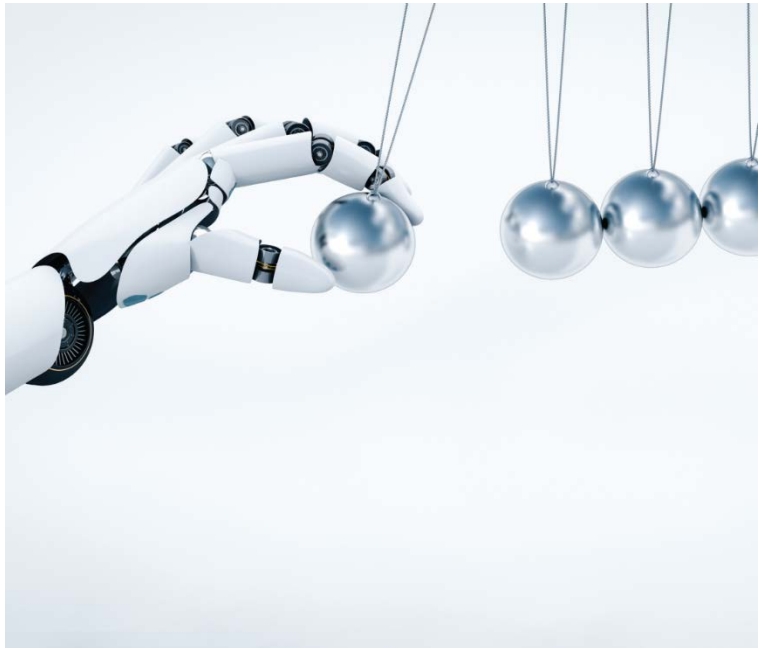
Principle 1: Ethical Purpose and Societal Benefit

- The critical threshold question: “to what end?”

- We assessed four specific areas of great societal impact, where AI will likely have direct and potential unintended consequences:
  - Work & Employment
  - Ecology and the environment
  - Warfare & conflicts
  - Democratic life

- Implications for financial institutions:
  - AI Impact Assessments should include consideration of broader social impact.

Principle 2: Accountability



Organisations that develop, deploy or use AI systems shall respect and adopt the eight principles of this Policy Framework for Responsible AI (or other analogous accountability principles).

In all instances, humans should remain accountable for the acts and omissions of AI systems.

Principle 2: Accountability

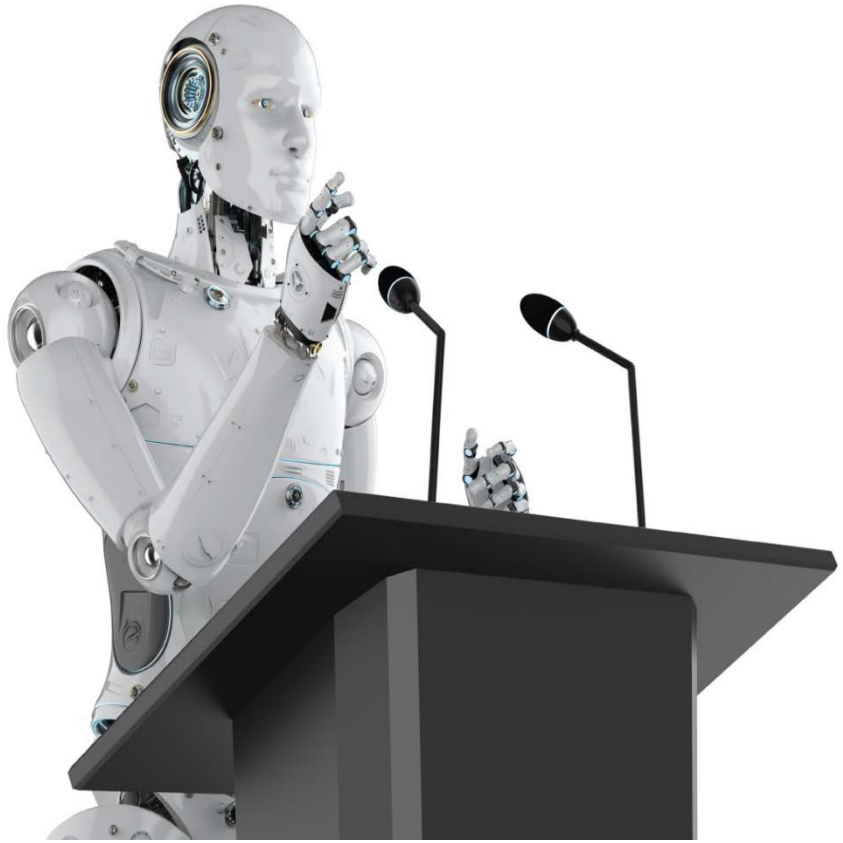
- Human-centered approach has a counterpoint: human accountability

- Redress the potential for “accountability gaps” that might arise as a result of AI implementation

  - Granting legal personality to AI systems:
  - Revisiting the logic of liability shield for internet intermediaries

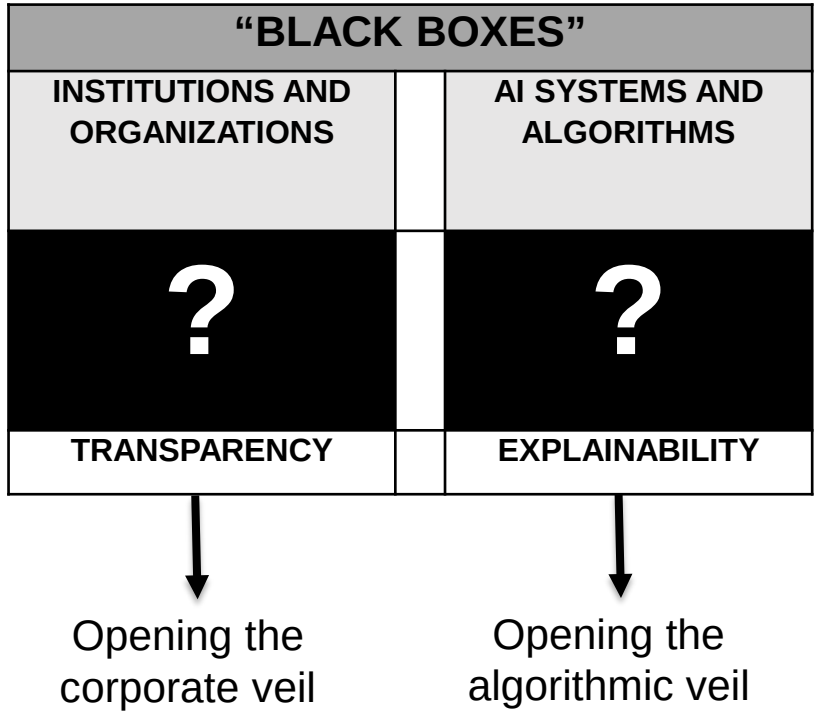
- Implications for governance for financial institutions
  - Chief Data Officer
  - AI Impact Assessment
  - Processes for confirming data readiness and data security
  - Human in the loop design

Principle 3: Transparency and Explainability



Organisations that develop, deploy or use AI systems shall ensure that, to the extent reasonable given the circumstances and state of the art of the technology, such use is transparent and that the decision outcomes of the AI system are explainable.

Principle 3: Transparency and Explainability



- Essential to reinforce accountability
- Essential to develop trust in the fairness of outcomes by general public (ultimate consumers)

- Parallel to administrative law principles of natural justice and fairness obligations
- Intensity of the obligations will vary depending on the sensitivity of the information and the criticality of the decision outcomes

- Implications for financial institutions:
  - "explainability by design"
  - capacity to audit results

Principle 4: Fairness and Non-Discrimination

Organisations that develop, deploy or use AI systems shall ensure the non-discrimination of AI outcomes, and shall promote appropriate and effective measures to safeguard fairness in AI use.

Principle 4: Fairness and Non-Discrimination

- Data bias (conscious and unconscious)
- Algorithmic bias (conscious and unconscious)

English-Turkish-English

He is lazy	*O tembel*
O tembel	**She's lazy**

She is smart	*O akıllı*
O akıllı	**He's smart**

English-Finish-English

He is a nurse	*Hän on sairaanhoitaja*
Hän on sairaanhoitaja	**She's a nurse**

She is an engineer	*Hän on insinööri*
Hän on insinööri	**He is an engineer**

- Implications for Financial institutions:
  - •ensure diversity on research teams
  - •focus on data readiness
  - •« neutral » data may in fact hide implicit bias
  - •fairness/non-discrimination “by design” is necessary but insufficient
  - •must also monitor outcomes to test for unintended discriminatory outcomes (feedback loop)

Conclusion

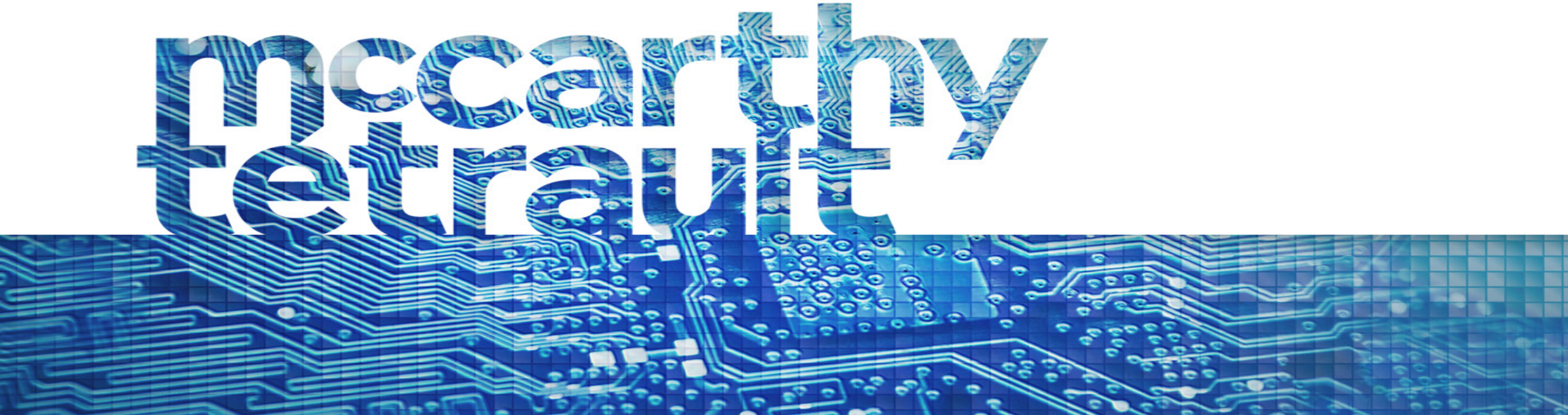
- AI will almost certainly have an enormous societal impact
- The rate of change and development is both exhilaration and alarming

- The time is now, in the relative infancy of AI development, to establish solid guideposts for the responsible development, deployment and use of AI systems

- The issues call out for active participation by a broad range of stakeholders

- The Responsible AI public consultation process is underway … we encourage comments and feedback on the draft framework

Questions? Comments?



This program is eligible for up to 2 Substantive Hours under the Law Society of Ontario's mandatory CPD Regime.