Adopting Responsible AI in Practice

“the practice of designing, building and deploying AI in a manner that empowers people and business, and fairly impacts customers and society — allowing companies to engender trust and scale AI with confidence.”

WORLD ECONOMIC FORUM, 2019

Agenda

- Background
- Government of Canada Approach to Responsible AI
- Learning from Experience
- Building a Responsible Certification Mark for AI
- Q&A

About Me

2008 ⟷ 2019

Data Driven Decision Making

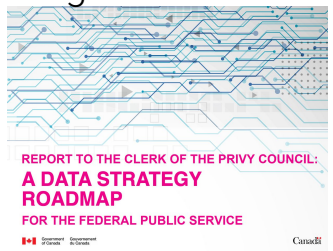
Community organizing

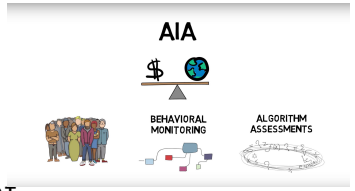
Corporate Governance

Open Data

Open Government

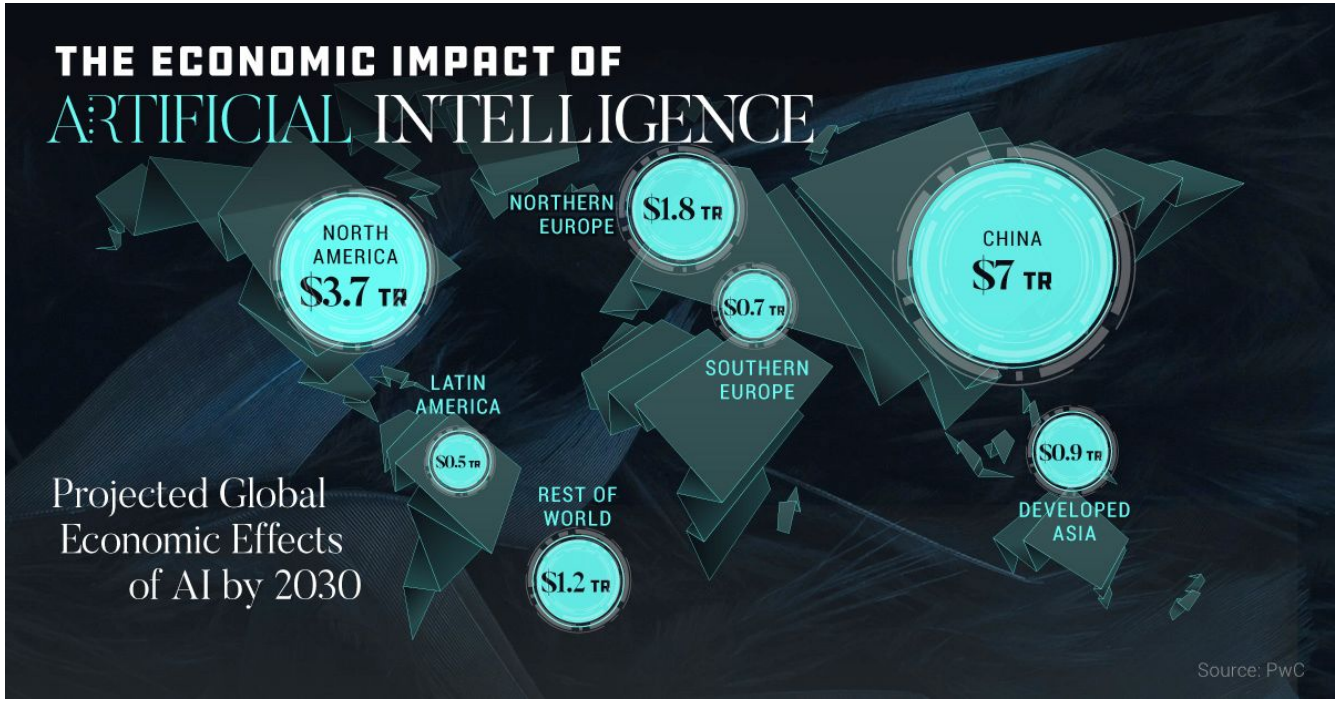
Data as a foundation for Digital Government





Implementing AI Responsibly

AI expected to add $15 Trillion to World Economy By 2030

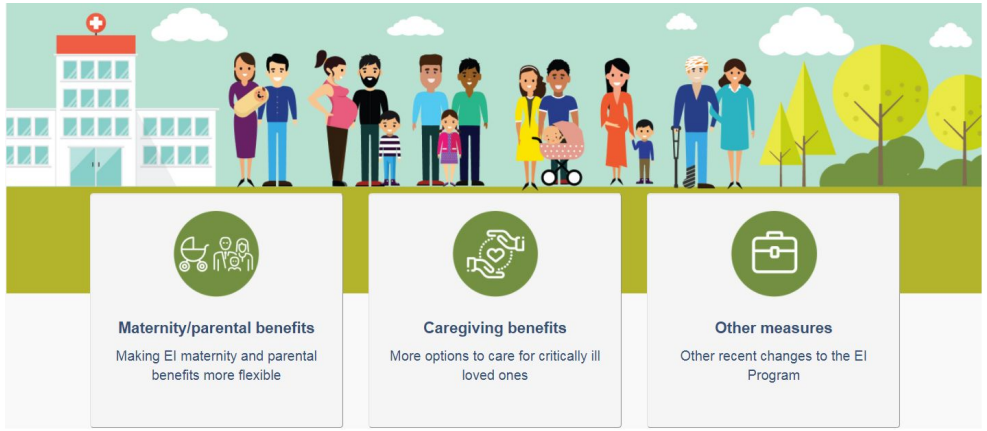


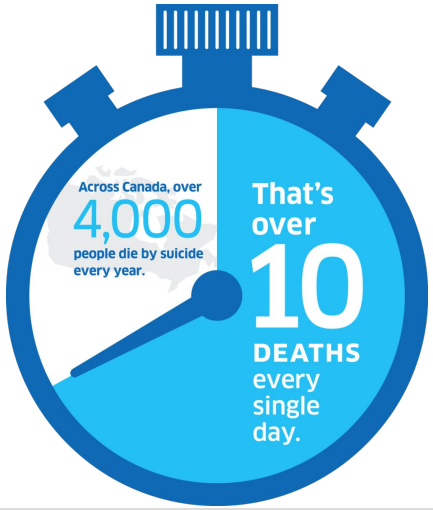
"In five years every decision will be impacted by Cognitive Computing."
IBM CEO

"AI will be as transformative to human kind as fire and electricity."
Google CEO

"Human-AI partnership can help solve society's challenges and release human creative potential."
Microsoft CEO

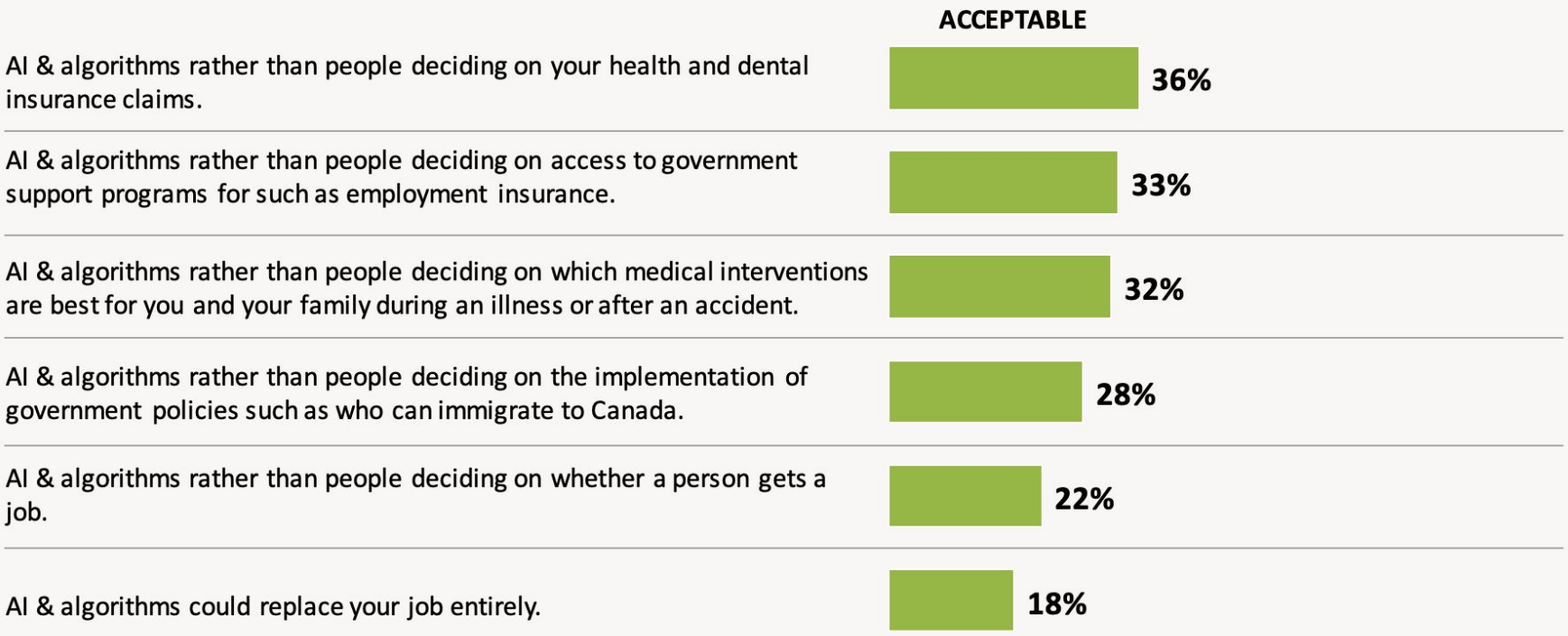
Limitless Opportunities





Diminishing Public Trust

The More it Matters the Lower the Support for AI

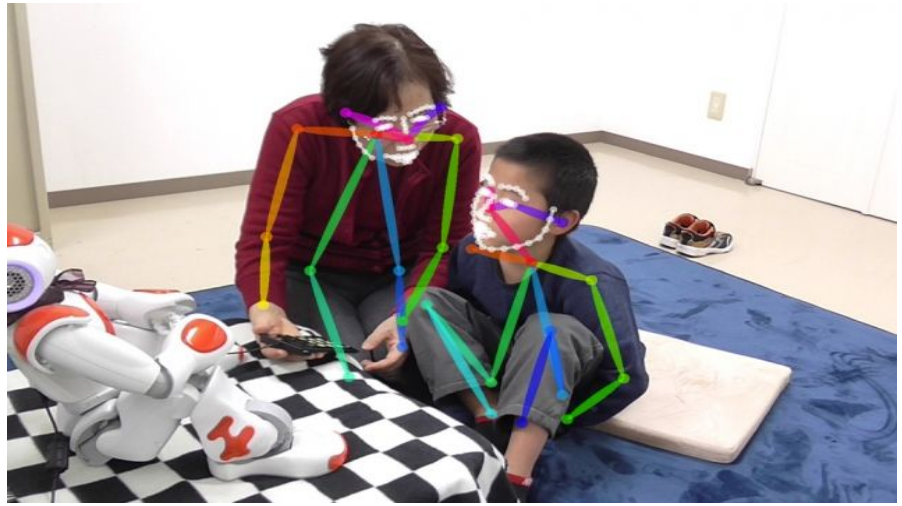


Ipsos

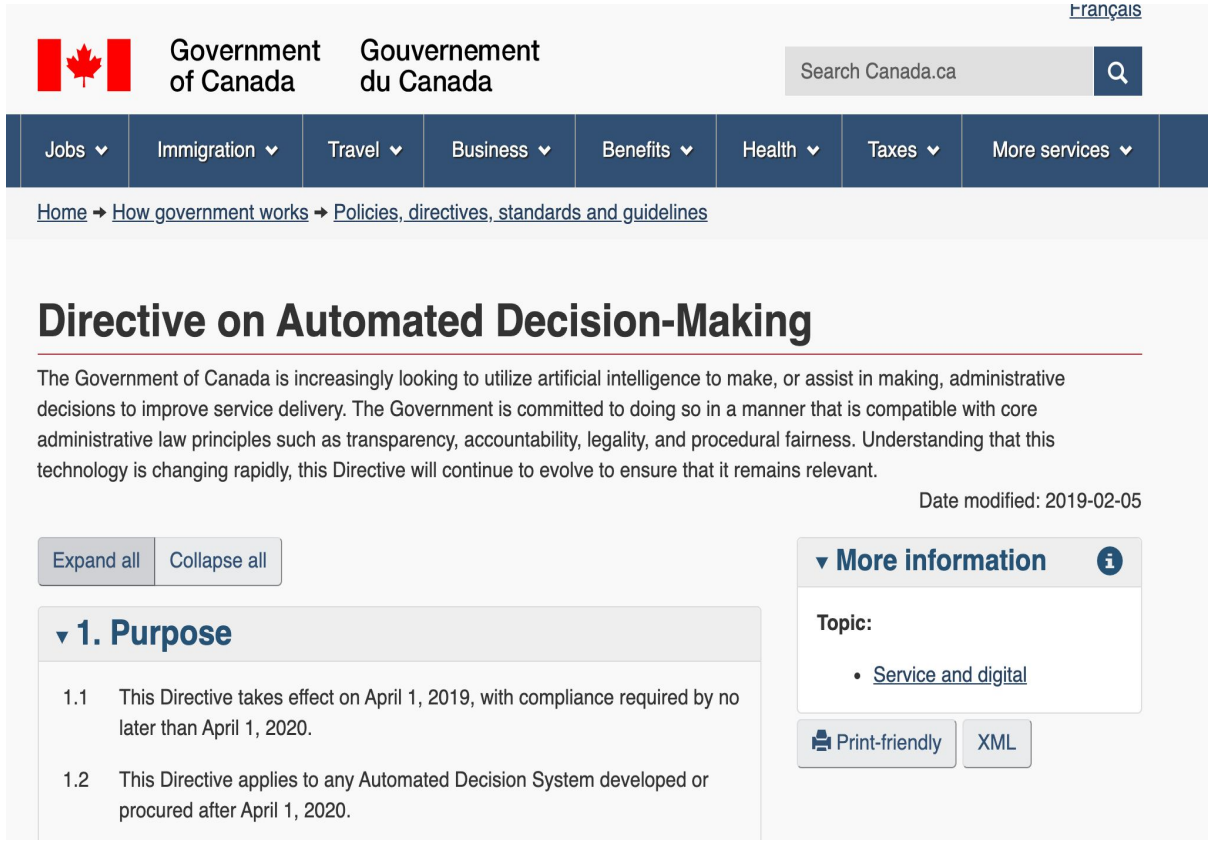
For each of the following, please indicate...: - assuming that it does happen do you think it is very acceptable, somewhat acceptable, not very acceptable, not at all acceptable. - Top 2 Box Summary.
Base: All Respondents. Total (n=2,001)

24

Contributing to Fear



First Generation AI Policy



Français

Government of Canada | Gouvernement du Canada

Search Canada.ca

Jobs | Immigration | Travel | Business | Benefits | Health | Taxes | More services

Home → How government works → Policies, directives, standards and guidelines

Directive on Automated Decision-Making

The Government of Canada is increasingly looking to utilize artificial intelligence to make, or assist in making, administrative decisions to improve service delivery. The Government is committed to doing so in a manner that is compatible with core administrative law principles such as transparency, accountability, legality, and procedural fairness. Understanding that this technology is changing rapidly, this Directive will continue to evolve to ensure that it remains relevant.

Date modified: 2019-02-05

Expand all | Collapse all

1. Purpose

1.1 This Directive takes effect on April 1, 2019, with compliance required by no later than April 1, 2020.

1.2 This Directive applies to any Automated Decision System developed or procured after April 1, 2020.

More information

Topic:

- Service and digital

Print-friendly | XML

- Peer review
- Notice
- Human in the loop
- Explanation requirements
- Testing
- Monitoring
- Training
- Contingency Planning
- Approval to operate the system

Developing the Algorithmic Impact Assessment

Algorithmic Impact Assessment

Link to GitHub project repository

Save | Choose File | No file chosen

Start Again

Algorithmic Impact Assessment v0.7

Page 8 of 13

Impact Assessment

Will the system only be used to assist a decision-maker?

- (•) Yes
- ( ) No

Will the system be replacing a decision that would otherwise be made by a human?

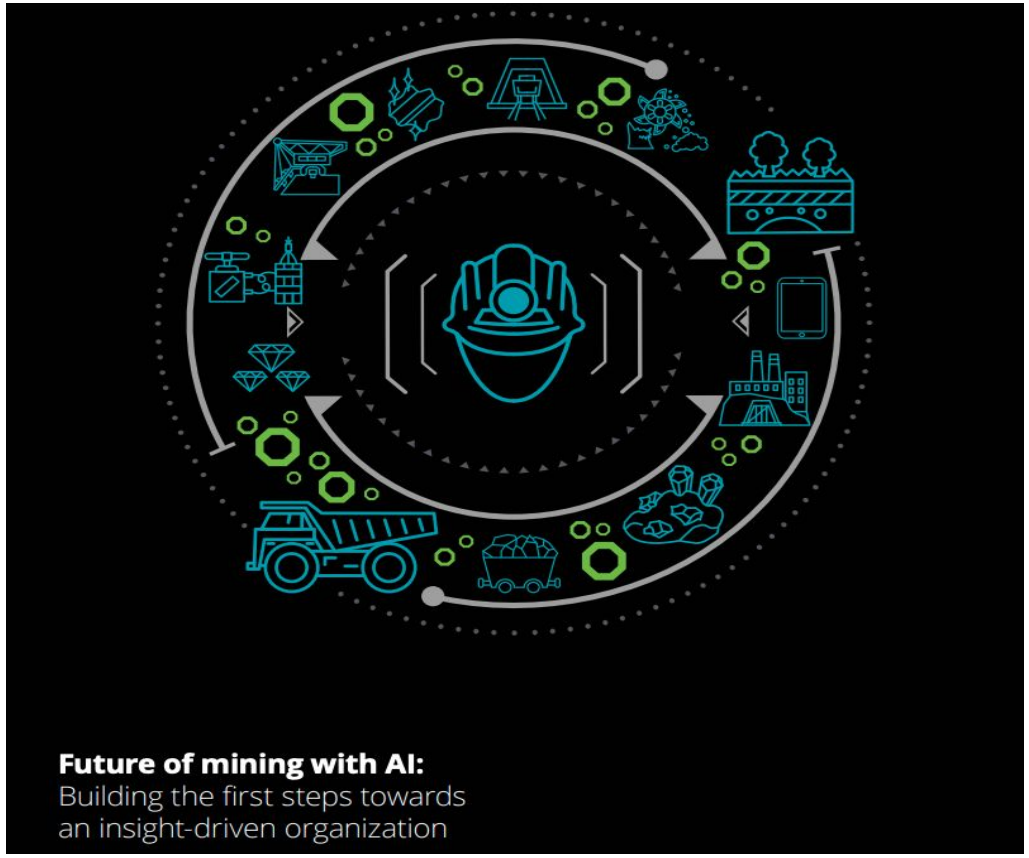
- (•) Yes
- ( ) No

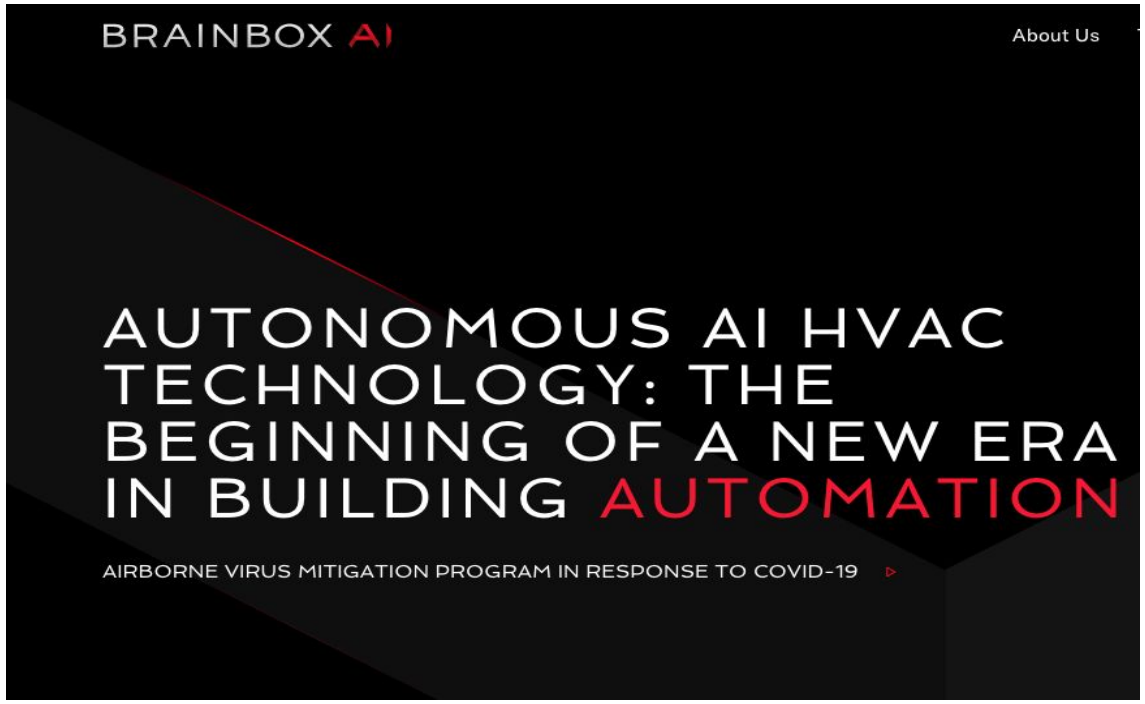
Will the system be replacing human decisions that require judgement or discretion?

- (•) Yes
- ( ) No

Please describe the decision(s) that will be automated

All Companies are Becoming Technology Companies





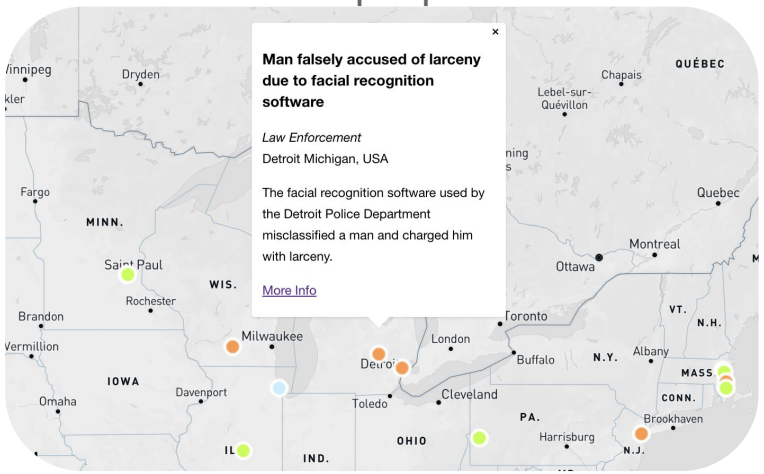
We love technology, but recognize...

Sometimes it hurts people

- Incorrect identification through facial recognition
- Misdiagnosis of health issue
- Biased prediction for:
  - jail sentence
  - insurance rate
  - access to credit
  - health treatment

Sometimes it doesn't get us (or see us)

- Gender and other minority bias for hiring practices
- Lack of recognition by computer for:
  - people with disabilities
  - people of colour



Sometimes it collects information about us without our explicit knowledge or permission

- Social networks
- Contact tracing apps for health
- Complex terms of service agreements
- Automated purchases through voice recognition systems

Why do we need Responsible AI?

- Trust in AI systems is at an all time low
- Clear regulations aren't in place yet
- There has been a significant response from companies, academics, and governments to better understand how AI can be responsible.
- These responses have come in the forms of reports, research, principles documents, and tools.
- While lots of good advice is readily available, it is difficult to navigate what to do.
- Finally, having a single framework would allow for a common understanding of what approaches should be taken by AI practitioners.

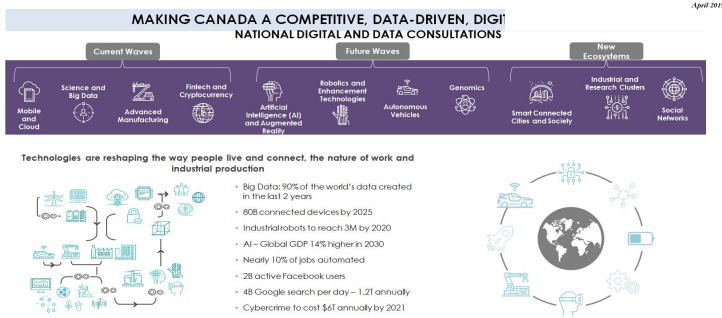

Politics

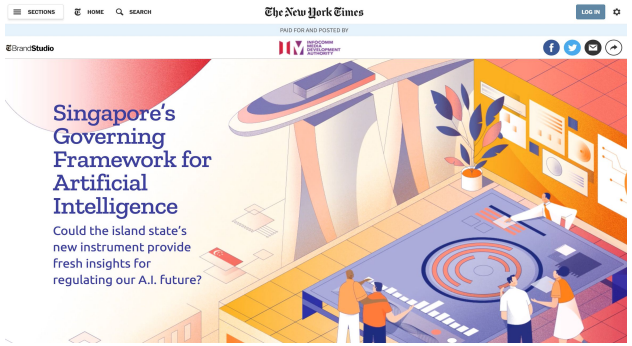
Robinhood Rise Brings Setbacks of Irate Traders, U.S. Probes

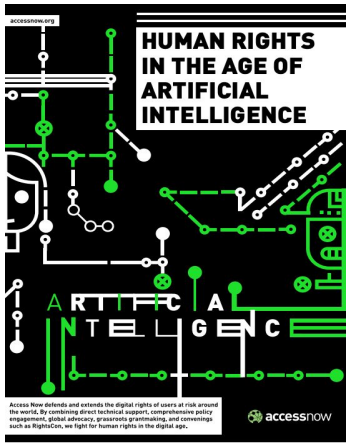
By Robert Schmidt and Benjamin Bain
August 31, 2020, 6:00 AM EDT Updated on August 31, 2020, 12:37 PM EDT

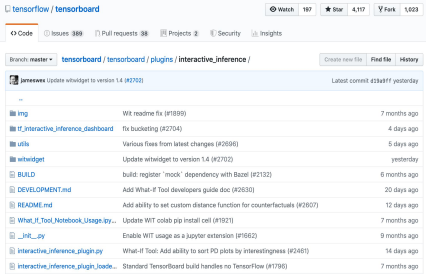
- Agencies flooded with complaints; SEC examining March outage
- Robinhood says it's committed to improving customer service

The Response









G7 FRANCE BIARRITZ 2019

AI for Good Global Summit

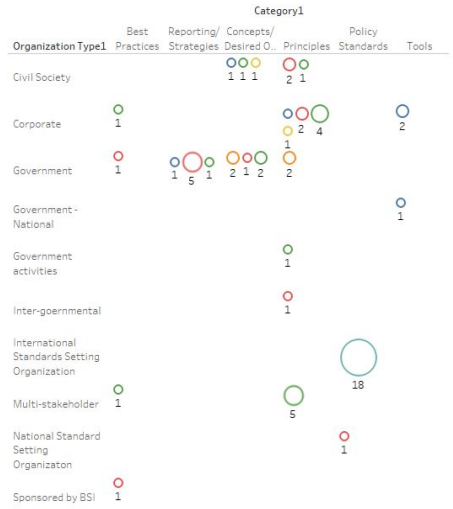
An ITU experience

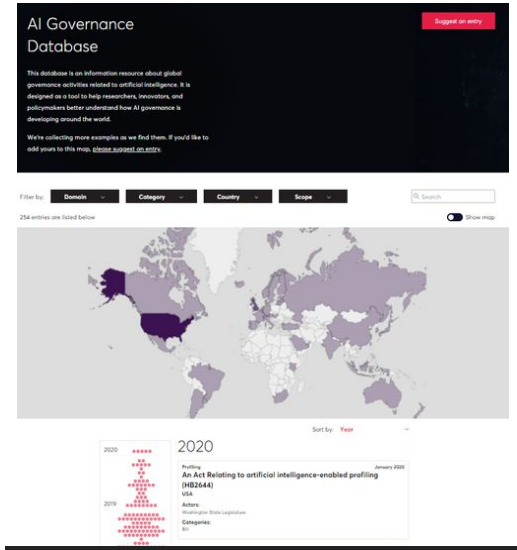
future of life INSTITUTE

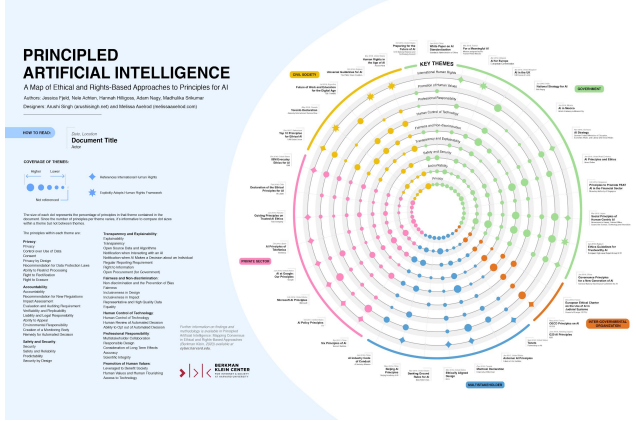
ASILOMAR AI PRINCIPLES

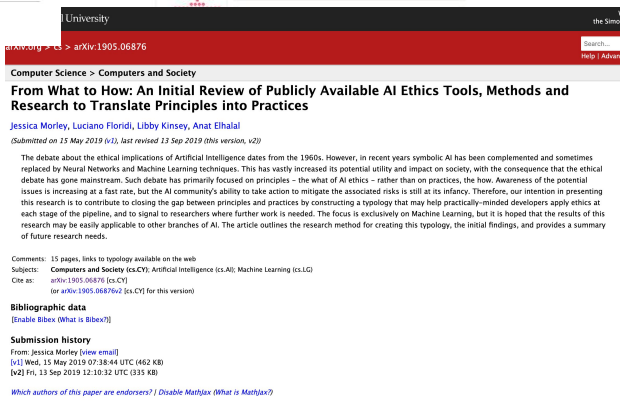
A Plethora of AI Principals

- AI Global mapped over 90 standards, policies, whitepapers, and frameworks
- Nesta is tracking AI Governance Tools, they currently have 256 contributions
- Oxford Internet Institute and Digital Catapult looked to create a common typology based on over 100 documents
- Berkman Klein mapped 35 of the key policies and principles to find commonalities









Responsible AI Landscape

OECD Principles

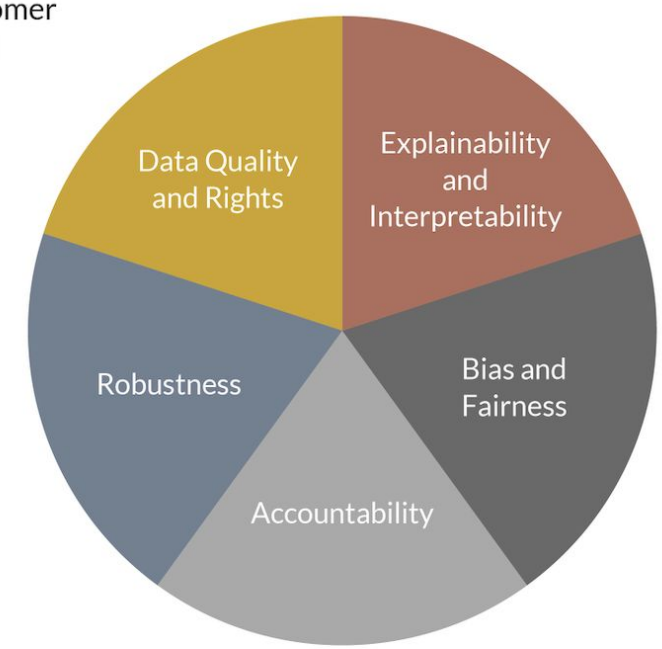
The Recommendation identifies five complementary values-based principles for the responsible stewardship of trustworthy AI:

- AI should **benefit people and the planet** by driving inclusive growth, sustainable development and well-being.
- AI systems **should be designed in a way that respects the rule of law, human rights, democratic values and diversity**, and they should include appropriate safeguards – for example, enabling human intervention where necessary – to ensure a fair and just society.
- There **should be transparency and responsible disclosure** around AI systems to ensure that people understand AI-based outcomes and can challenge them.
- AI systems must function in a robust, secure and safe way throughout their life cycles and potential risks should be continually assessed and managed.
- Organisations and individuals developing, deploying or operating AI systems should be held accountable for their proper functioning in line with the above principles.

Responsible AI Trust Index - operationalizing the principles to protect the public

AI systems should respect data privacy and avoid using customer data beyond its intended and stated use.

AI systems should be explainable to ensure that people understand AI-based outcomes and can challenge them.



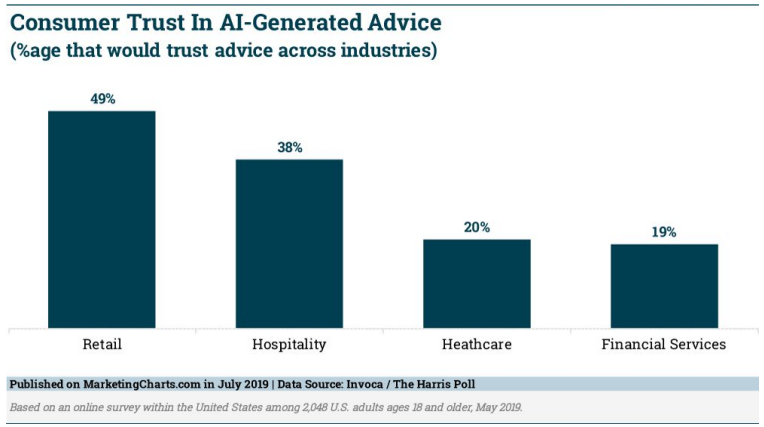
AI systems must function in a robust, secure and safe way throughout their life cycles

AI systems should mitigate unwanted bias and drive inclusive growth to benefit people and the planet

Organizations developing, or operating AI systems should be held accountable for their proper functioning.

Accredited Certification Program for AI

Many claims are made by companies and their systems are ethical and responsible



However consumers still don't trust AI.

Similar to other industries, verifiable accredited certification programs have helped consumers know what can be trusted.

Certification Approach

Build

A comprehensive community validated framework

Accredit

Framework assesed by National accreditation bodies

Audit

Audits performed by trained and independent auditors

Certify

Four levels of RAI Certification are awarded

RAI Certification Framework dimensions

- Accountability
- Bias and Fairness
- Data quality
- Explainability and Interpretability
- Robustness

RAI Leading the Charge

Responsible AI Trust Index

Responsible AI Design Assistant

Public open source version

Responsible AI Certification Mark

Licensed to accredited auditors

Building from existing work

- Fostered in an open source environment, we're not interested in re-inventing any wheels. Our key priority is to make it as easy as possible for practitioners to build their AI systems in a responsible and ethical way, from the start.
- Where possible, the responsible AI certification framework references existing policies, regulations, principles, standards, tools, and industry best practices.
- We also think it's important to know how to be responsible given your context and your region, so rules and best practices may change based on the environment.

OECD.AI Policy Observatory

IBM
AI Fairness 360 Toolkit (AIF360)

< >
Déclaration de Montréal
IA responsable_

Google AI

AI

Canada

OpenAI

PARTNERSHIP ON AI

How the Responsible Trust Index is Used

System: Automated Colon Cancer Screening System Organization: National Health Care Medical Lab

Dimension	Improvements because of Responsible AI Trust Index
Accountability	• Improved governance to provide review of system trade-offs • Robust training of the system was implemented for users before deployment • Ongoing monitoring for unintended outcomes including financial loss (eg. unnecessary lab time, false positives, etc.) and reputation issues (eg. delay of particular ethnic group) was integrated into system deployment
Bias and Fairness	• Above mentioned ongoing monitoring will look to see if there are anomalies in prediction system which the system will learn from. • Challenge function has been integrated for potentially wrong diagnosis
Data Quality	• Data collection, use, and distribution practices were drastically improved, for example, previously data wasn't tested for bias, accuracy, et
Explainability and Interpretability	• Counterfactual analysis was integrated into system review to have an improved understanding of how to the system makes decisions. • Heat maps and other visuals were developed to demonstrate how the system operates. • Clear consent for system users was developed.
Robustness	• Edge cases were considered to mitigate disruption of service

How the Responsible Trust Index is Used

System: Automated Triage of Claims Organization: Regional Insurance Company

Dimension	Improvements because of Responsible AI Trust Index
Accountability	• Improved risk review of current process to understand issues that could be propagated in automated decision. • Integration of key processes like integration of logs built into the system to track activity for future review and audits • Implementation of a contingency plan when system isn't functional
Bias and Fairness	• Simulation testing was done to understand unintended outcomes in various implementation scenarios. • Consideration of use with different users was integrated into development.
Data Quality	• Since historical data was used to train the model this was analysed for accuracy and unintended biases to ensure system wasn't exasperating issues. • Chose to limit testing of bias to reduce collection of privacy information.
Explainability and Interpretability	• System owner has decided to keep information private for proprietary purposes, however, would be able to explain system operations for audit if required.
Robustness	• Edge cases were considered to mitigate disruption of service

Responsible AI Certification Mark

The concept of certification marks for the responsible and ethical use of AI has been raised by many organizations. There are several ways such a program can manifest. Our approach is:

Scope

- AI systems (data, algorithms, processes)
- Complements doesn't duplicate other governance, legal, and risk evaluations

Key Audiences

- Companies committed to responsible AI
- Government procurement
- Small and medium businesses
- Regulators and international organizations

Guiding Principles

Robustness
Accountability
Bias & Fairness
Data Quality
Explainability & Interpretability

“Building a comprehensive and independent certification program that is grounded in accepted principles, is practical & measurable, is internationally recognized, and is built with trust & transparency.”

Dimensions of the Responsible AI Certification Mark

Assessment Components

- Data
- Models
- Processes

Assessment Categories

Robustness	Accountability	Bias and Fairness	Data Quality	Explainability & Interpretability
Ensuring the efficacy of the underlying technology in the system.	Oversight measures and procedures required for AI systems recognizing that they are complex systems that have the capacity to change or fluctuate.	Ensuring equity in the implementation of these systems.	Data is a foundational aspect of all AI systems. As such, the efficacy and operation of these systems is dependent on good data.	Methods and techniques used for the AI system to be understood by humans.

Areas of Focus

Domain

Ensures compliance with relevant regulations and industry best practices for healthcare, insurance, finance, privacy, and more.

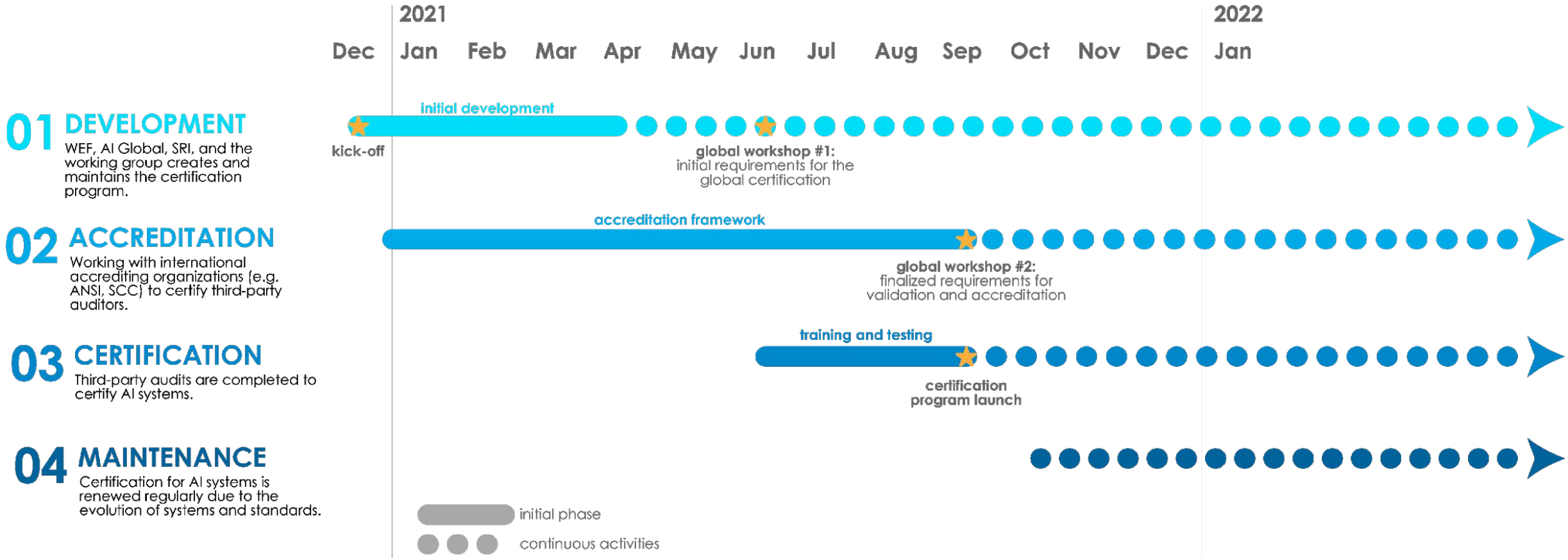
Region

Applicable in the US, UK, Europe, Singapore, NZ, and Canada, with greater adoption to come.

System type

Assessed by type of AI system to ensure the greatest effectiveness, like chatbots, facial recognition, robotic processing automation, and more.

Implementation Plan



Early Participant Organizations

Interest generated from **80 survey responses**:

- 22 organizations want to have their products certified
- 24 organizations want to become certification partners
- 74 individuals and organizations are interested in contributing their time and or resources

Responses **came from people in**:

- North America (US, Canada)
- Asia (India, Japan, Singapore)
- EU (UK, Germany, Netherlands, etc)
- Africa

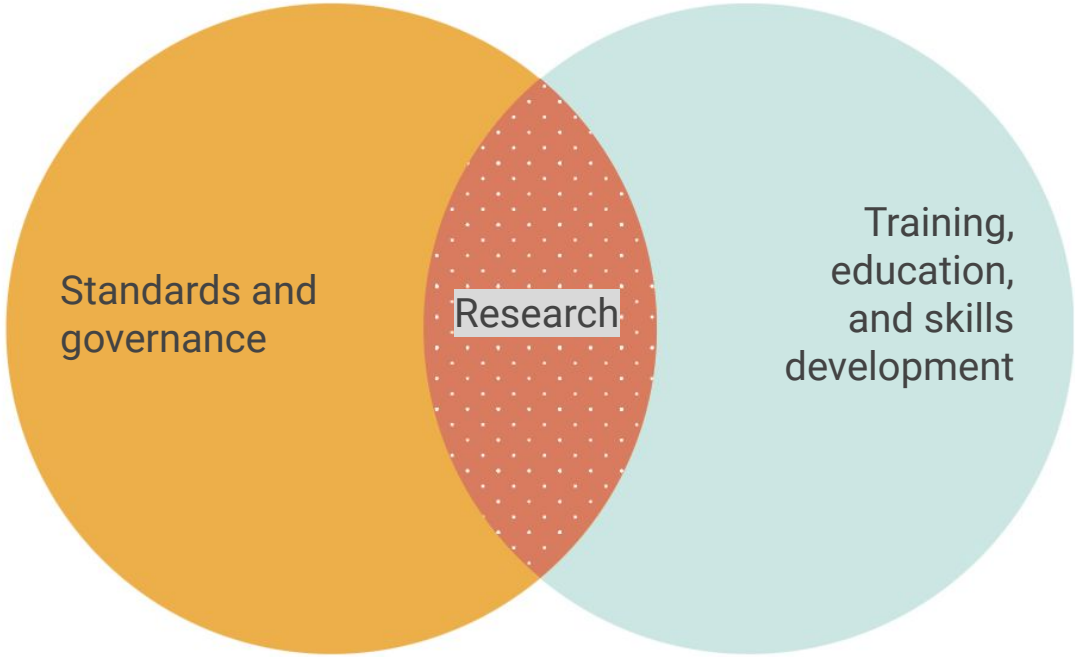
Respondents have expertise in the following **key focus areas**:

- Finance
- Technology companies (large and small)
- Media and networks
- Labour
- Regulation
- Government

Part of the Responsible AI ecosystem

We know that rules don't solve everything, but they do help us navigate what and what not to do.

A comprehensive certification program will provide guardrails to mitigate harm, however, we continue to work with practitioners and researchers to inform and mature our work.



Resources

Website

- RAI Overview

Responsible AI Design Assistant

- Responsible AI Design Assistant Tool
- Launch of Responsible AI Certification Working Group

Responsible AI Certification Mark

- White paper - Creating a Responsible AI Certification Mark

Responsible AI Community Portal

- Currently under construction but here are some useful AI Global documents
  - Where AI has Gone Wrong map and dataset
  - Responsible AI Documents visual and dataset
  - Responsible AI Landscape Review
  - AI Standards Review

Public Consultations

- UNESCO AI Recommendations Consultation report

Responsible AI Toolkit

- Draft Guidelines: Independent Review for Responsible AI Systems

Questions

Do you want to learn more about AI Global?

Follow us:

@responsible-ai-institute

@responsible.ai

@responsibleai

Email me: ashley@responsible.ai

Join us in building a better world with Responsible AI

responsible.ai

Responsible
Artificial Intelligence
Institute

Thank you