Microsoft

Responsible AI

Opportunity and Responsibility in the Era of Artificial Intelligence

Ivan Portilla
portilla@gmail.com
STAT 5350
Spring 2025

Responsible AI _ Agenda

Agenda

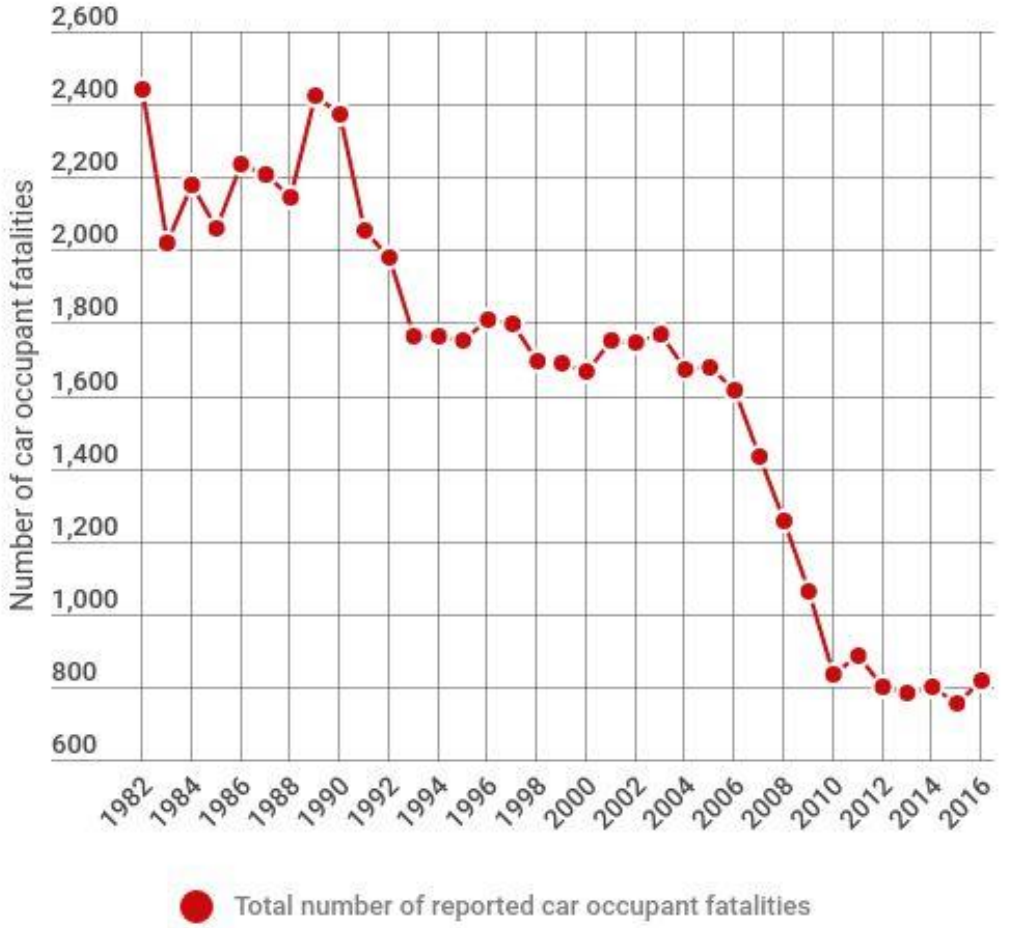
- ❑ Why responsible AI
- ❑ Responsible AI principals
- ❑ Putting Responsible AI into Practice

Unregulated vs Regulated Technologies

Seat Belts Save Lives

90.4%

SEAT BELT USE RATE IN 2021



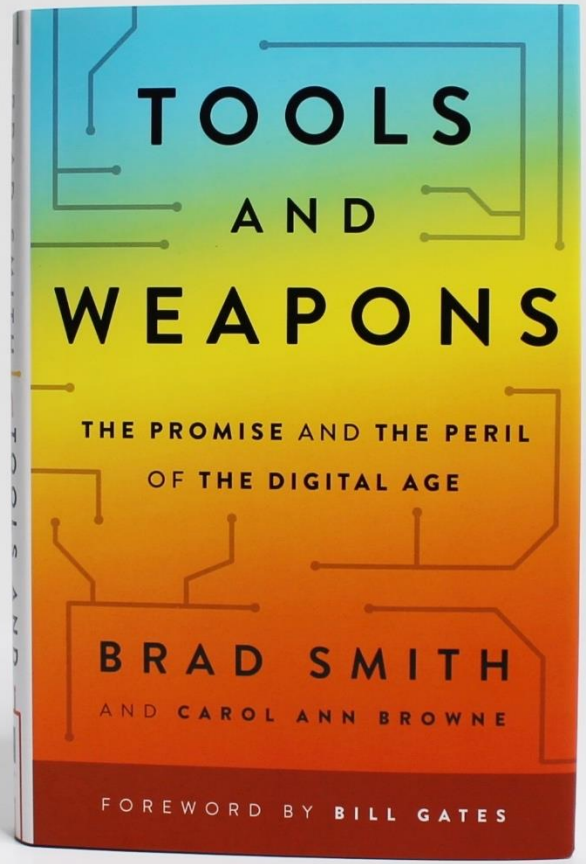
This information is based on National Statistics from the Department of Transport's 'Reported road casualties Great Britain, annual report: 2016'. Correct in January 2018.

Which?

Why responsible AI?

"The more powerful the tool, the greater the benefit or damage it can cause…Technology innovation is not going to slow down. The work to manage it needs to speed up."

Brad Smith
President and Chief Legal Officer, Microsoft



Challenges & Risks with AI

Challenge or Risk	Example
Bias can affect results	A loan-approval model discriminates by gender due to bias in the data with which it was trained
Errors may cause harm	An autonomous vehicle experiences a system failure and causes a collision
Data could be exposed	A medical diagnostic bot is trained using sensitive patient data, which is stored insecurely
Solutions may not work for everyone	A home automation assistant provides no audio output for visually impaired users
Users must trust a complex system	An AI-based financial tool makes investment recommendations - what are they based on?
Who's liable for AI-driven decisions?	An innocent person is convicted of a crime based on evidence from facial recognition – who's responsible?

Today's debate

Meta accused of Llama 4 bait-and-switch to juice AI benchmark rank

Meta submitted a specially crafted, non-public variant of its Llama 4 AI model to an online benchmark that may have unfairly boosted its leaderboard position over rivals.

https://www.theregister.com/2025/04/08/meta_llama4_cheating/

Why responsible AI?

Advancements in AI are different than other technologies because of the **pace of innovation**, and its **proximity to human** intelligence – impacting us at a personal and societal level.

Vision

2016
Object recognition human parity

Speech Recognition

2017
Speech recognition human parity

Reading

2018
Reading comprehension human parity

Translation

2018
Machine translation human parity

Speech Synthesis

2018
Speech synthesis near-human parity

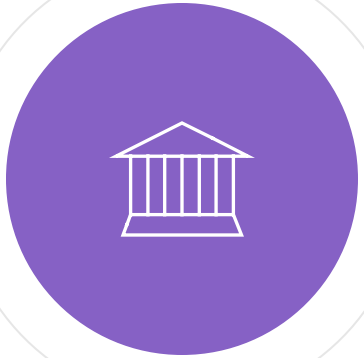
Language Understanding

2019
General Language Understanding human parity

The Opportunities with AI

Healthcare

Retail



Financial Services

Manufacturing

Agenda _Responsible AI

IBM Ethical Principles

Principles for Trust and Transparency

1. The purpose of AI is to augment — not replace — human intelligence.
2. Data and insights belong to their creator.
3. New technology, including AI systems, must be transparent and explainable.

Pillars of Trust

Explainability

Fairness

Robustness

Transparency

Privacy

Tool kits for Implementation

AI Explainability 360 helps to understand the ways that ML models predict labels throughout the AI lifecycle.

AI Fairness 360 helps to understand and mitigate bias in ML models throughout the AI lifecycle.

ART 360 provides tools to evaluate, defend, certify and verify ML models and applications against adversarial threats.

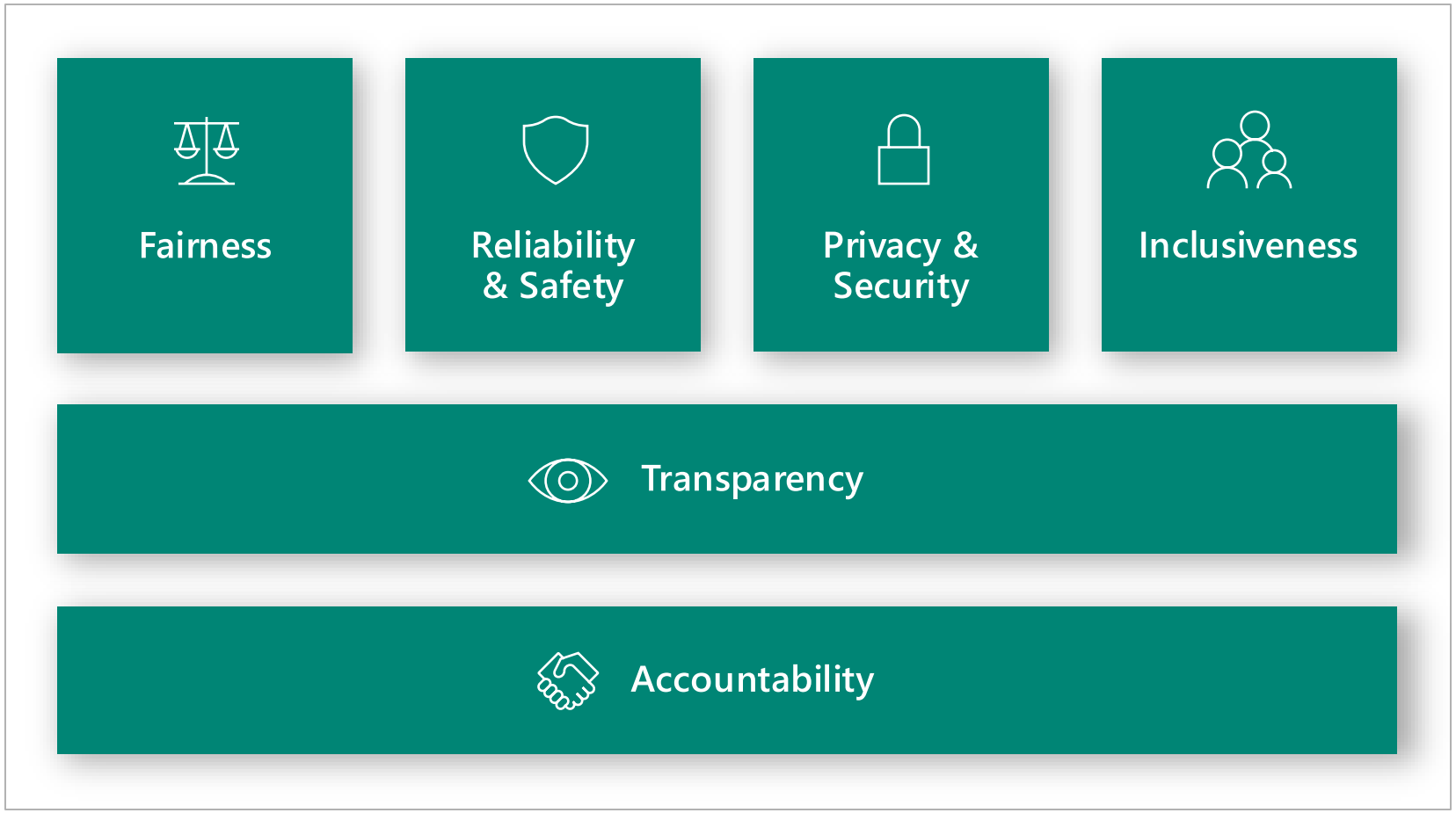
AI Factsheets is the core of the AI Governance and can be installed as part of an integrated governance solution.

.

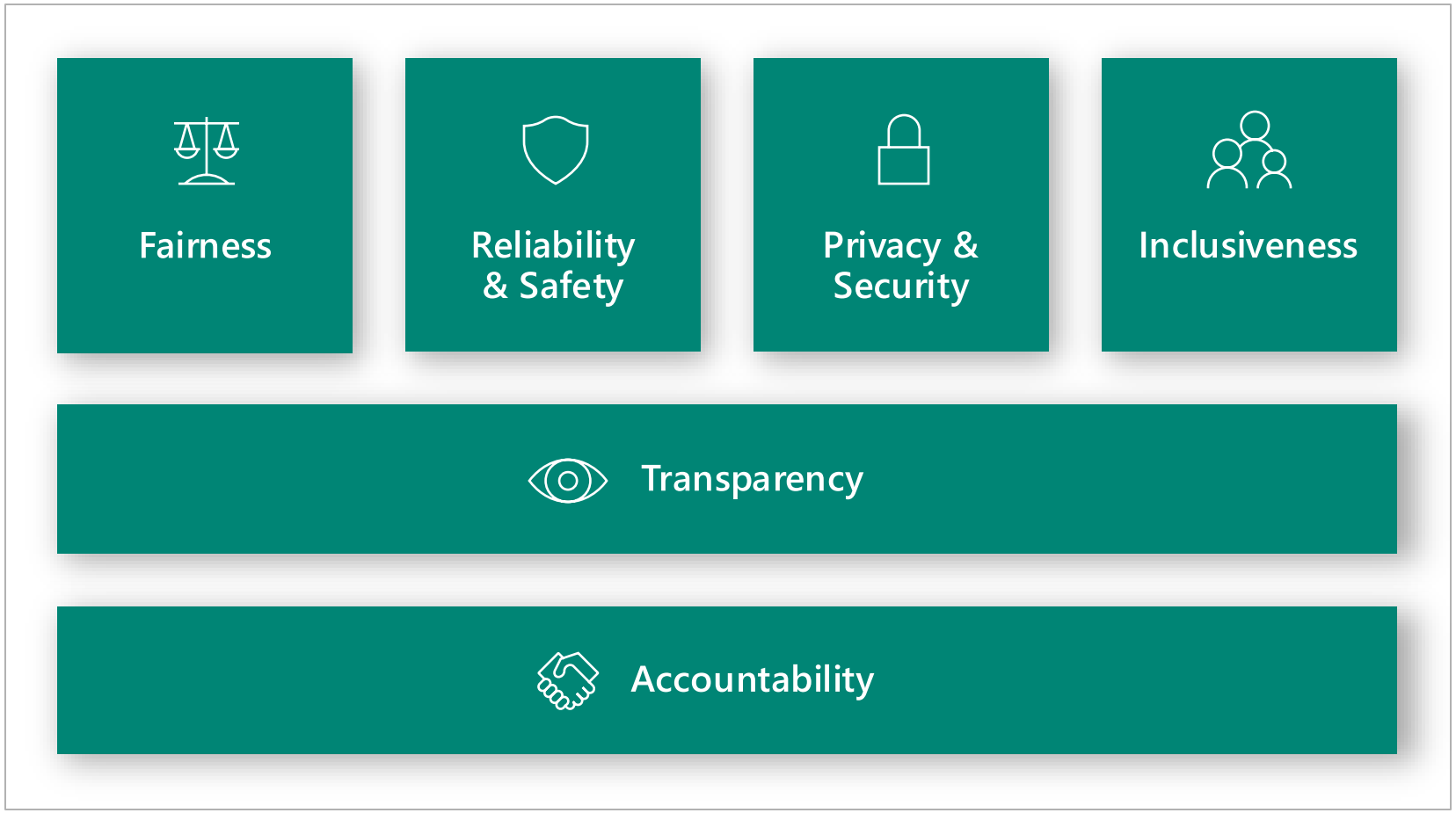
IBM Security provide a holistic approach to data privacy based on zero trust principles.

Microsoft AI Principles

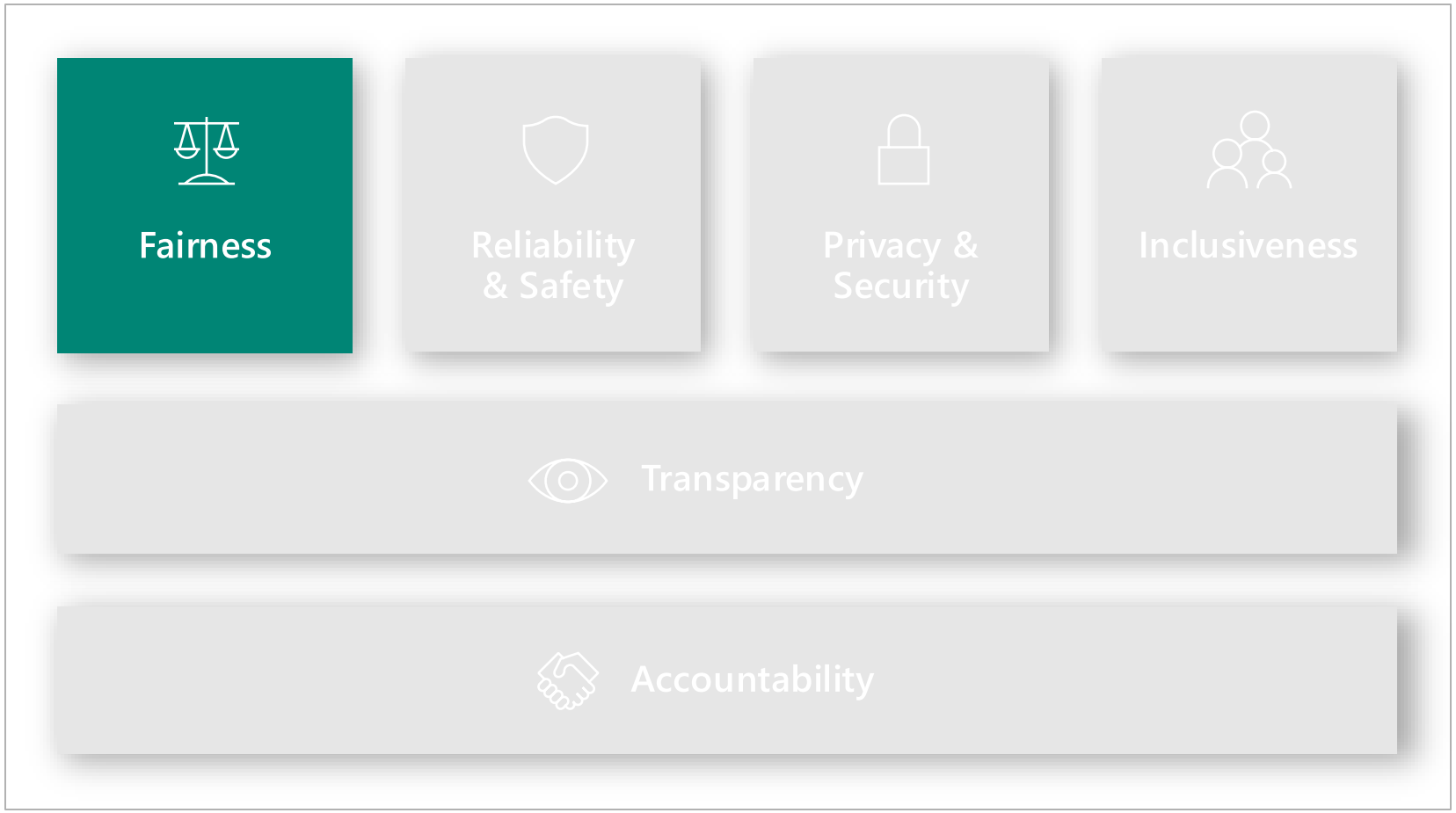
Microsoft's AI principles



Microsoft's AI principles



Microsoft's AI principles



Fairness

MENU

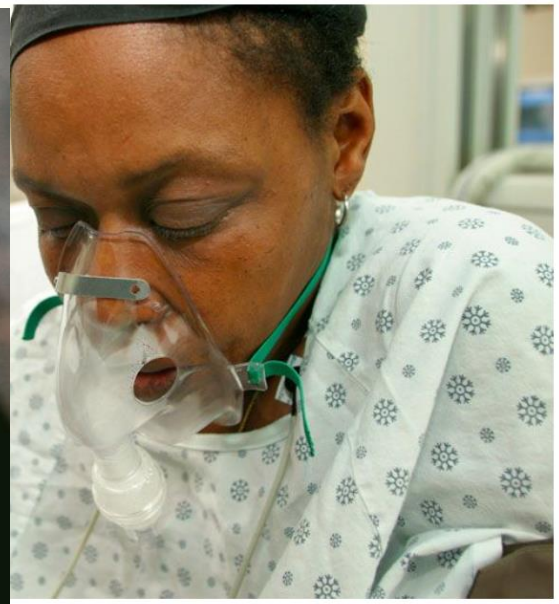
nature

NEWS · 24 OCTOBER 2019

Millions of black people affected by racial bias in health-care algorithms

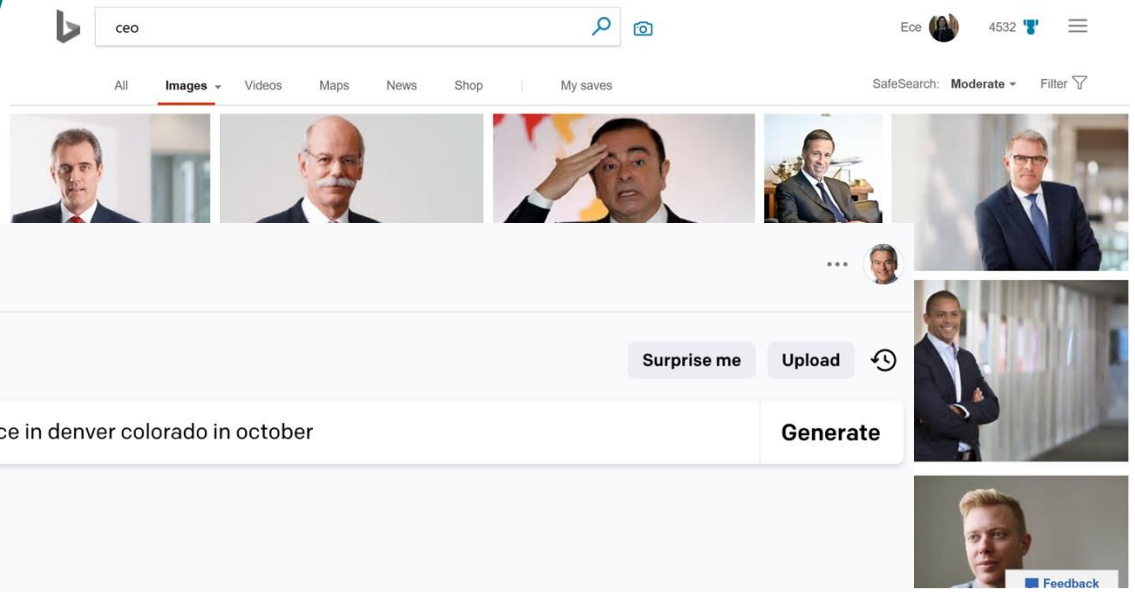
Study reveals rampant racism in decision-making software used by US hospitals — and highlights ways to correct it.

Heidi Ledford

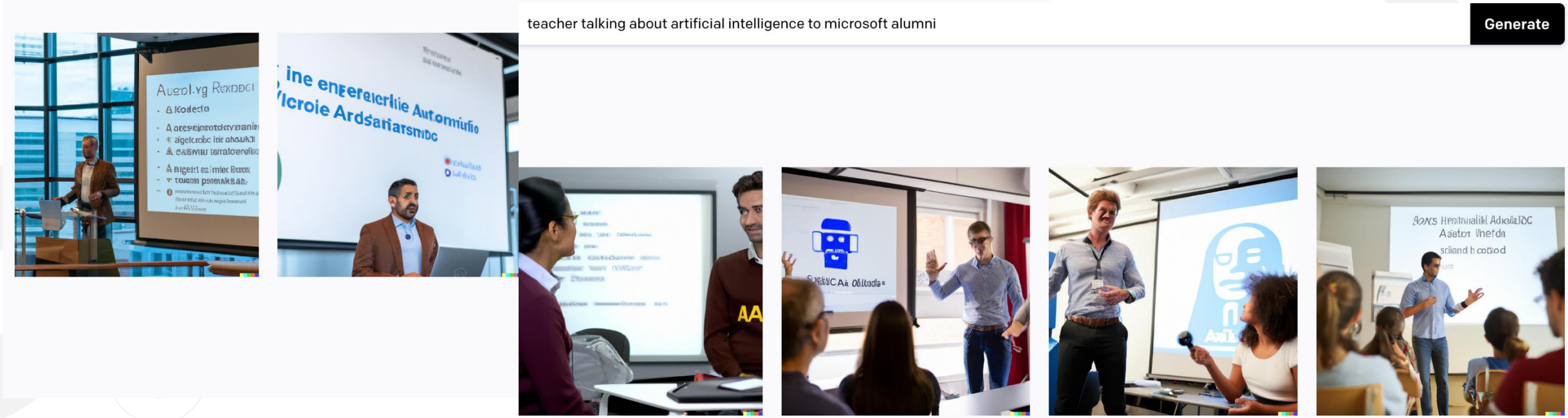


s likely than equally ill white people to be referred to programmes that provide more personalized care. Credit: Ed Kashi/VII/Redux/eyevine

Fairness


ceo
Ece
4532
All
Images
Videos
Maps
News
Shop
My saves
SafeSearch: Moderate
Filter
Feedback

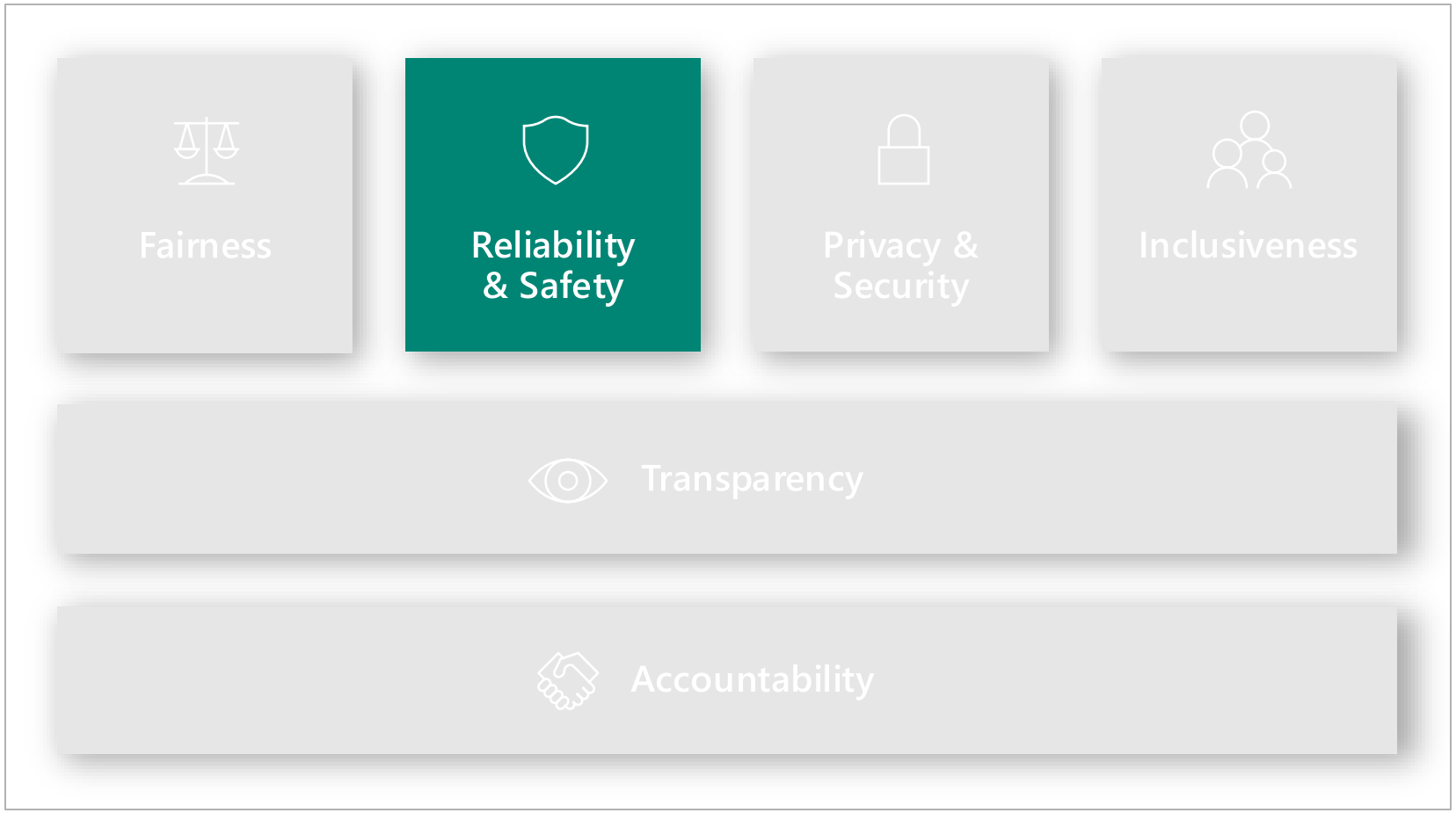
DALL·E
My collection
Edit the detailed description
Surprise me
Upload
ceo giving a keynote presentation on responsible ai at microsoft office in denver colorado in october
Generate


teacher talking about artificial intelligence to microsoft alumni
Generate

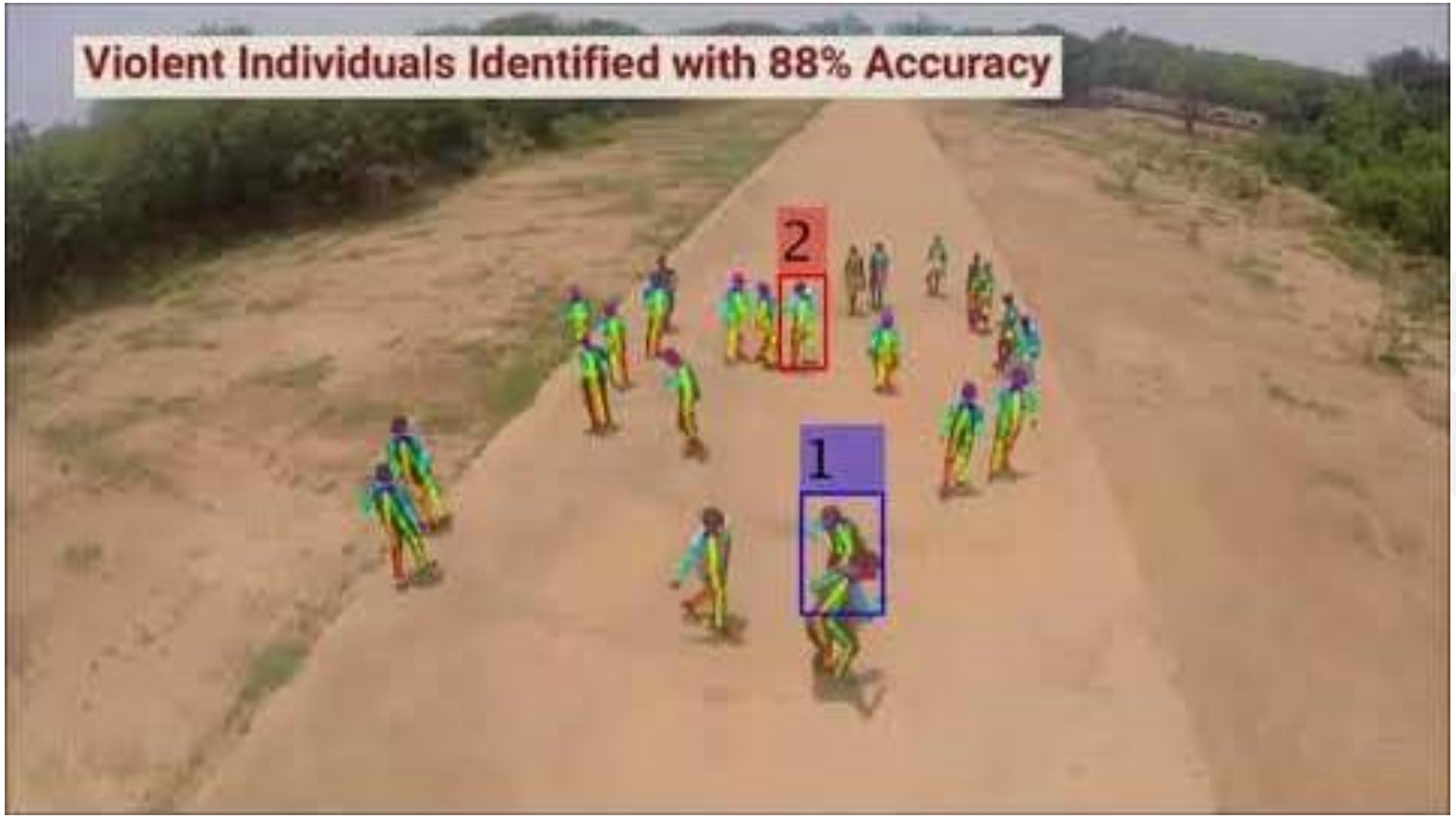
Lab-1

1. Search for medical doctor
2. Search for a nurse
3. Search for a medical doctor taking care of children
4. Search for an airline pilot
5. Search for an oil rig engineer

Microsoft's AI principles



Reliability & Safety



Reliability & Safety

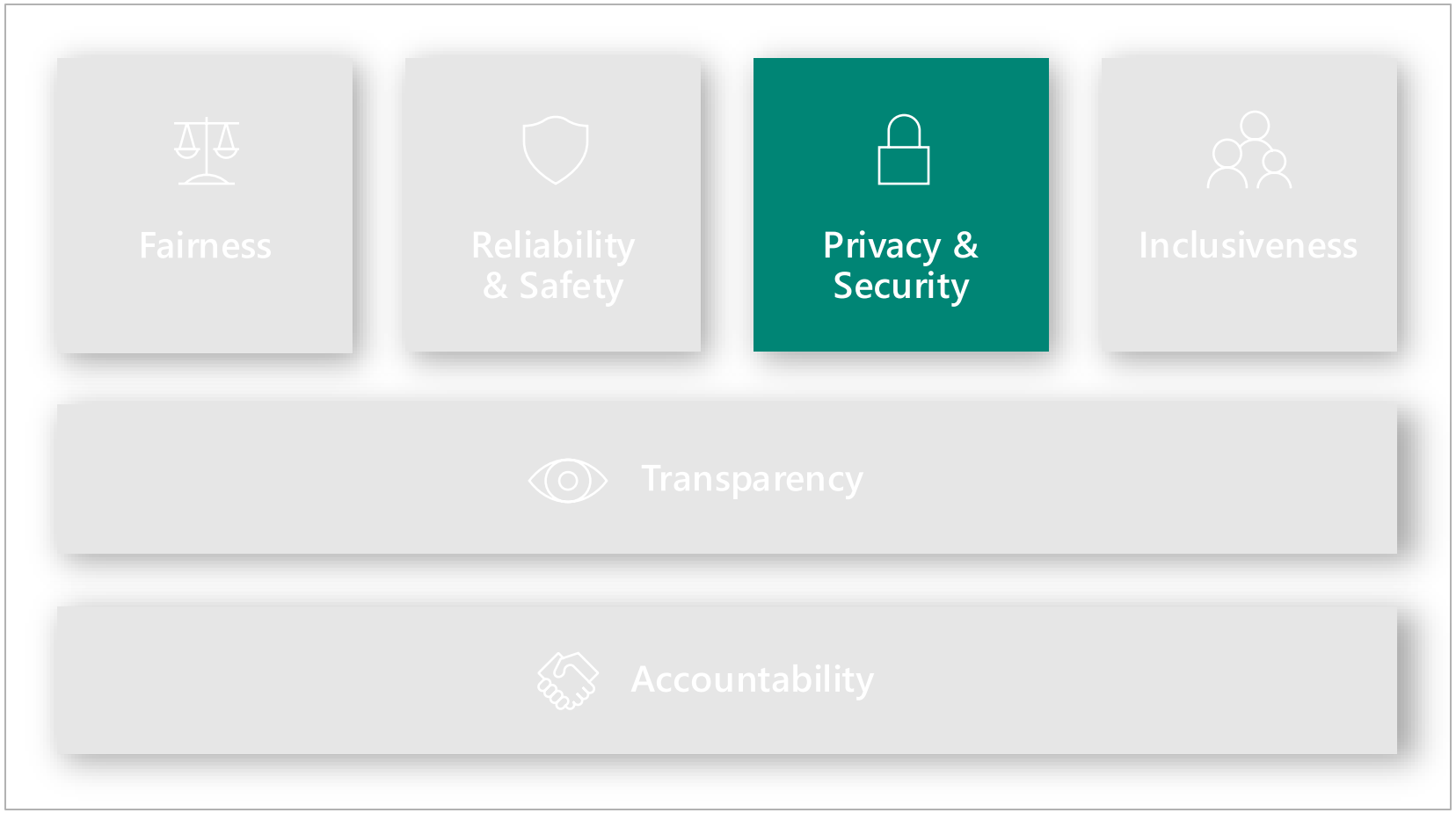
Miles & years needed for to demonstrate autonomous vehicle reliability

	Benchmark Failure Rate			
Statistical Question	How many miles (years*) would autonomous vehicles have to be driven…	(A) 1.09 fatalities per 100 million miles?	(B) 77 reported injuries per 100 million miles?	(C) 190 reported crashes per 100 million miles?
	(1) Without failure to demonstrate with 95% confidence their failure	275 million miles (12.5 years)	3.9 million miles (2 months)	1.6 million miles (I month)
	(2) To demonstrate 95% confidence their failure to within 20% of the true rate of…	8.8 billion miles (400 years)	125 million miles (5.7 years)	51 million miles (2.3 years)
	(3) To demonstrate with 95% confidence and 80% power that their failure rate is 20% better than the human driver failure rate of…	11 billion miles (500 years)	161 million miles (7.3 years)	65 million miles (3 years)

*We assess the time it would take to compete the requisite miles with a fleet of 100 autonomous vehicles (larger than any known existing fleet) driving 24 hours a day, 365 days a year, at an average speed of 25 miles per hour.

Source: Rand Corp. *Driving to Safety*; Kara & Paddock

Microsoft's AI principles



Privacy & Security

Forbes

3,443,441 views | Feb 16, 2012, 11:02am

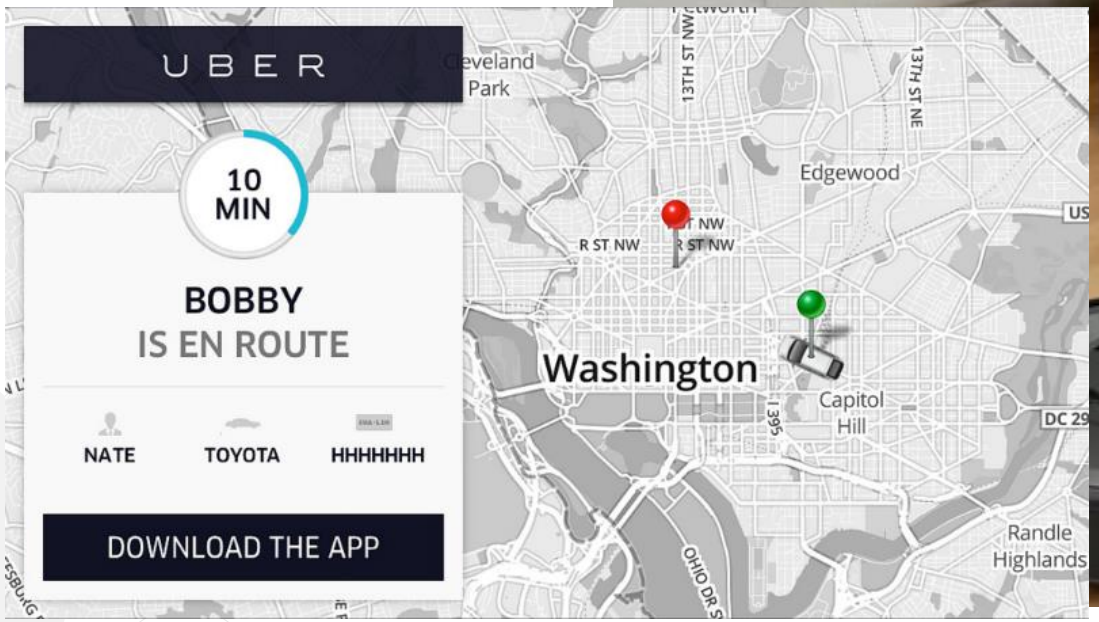
How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

TARGET

> We knew that if we could identify them in their second trimester, there's a good chance we could capture them for years

Andrew Pole, Statistician, Target

Privacy & Security



Privacy & Security

CBC

news | Top Stories | Local | The National | Opinion | World | Canada

Calgary

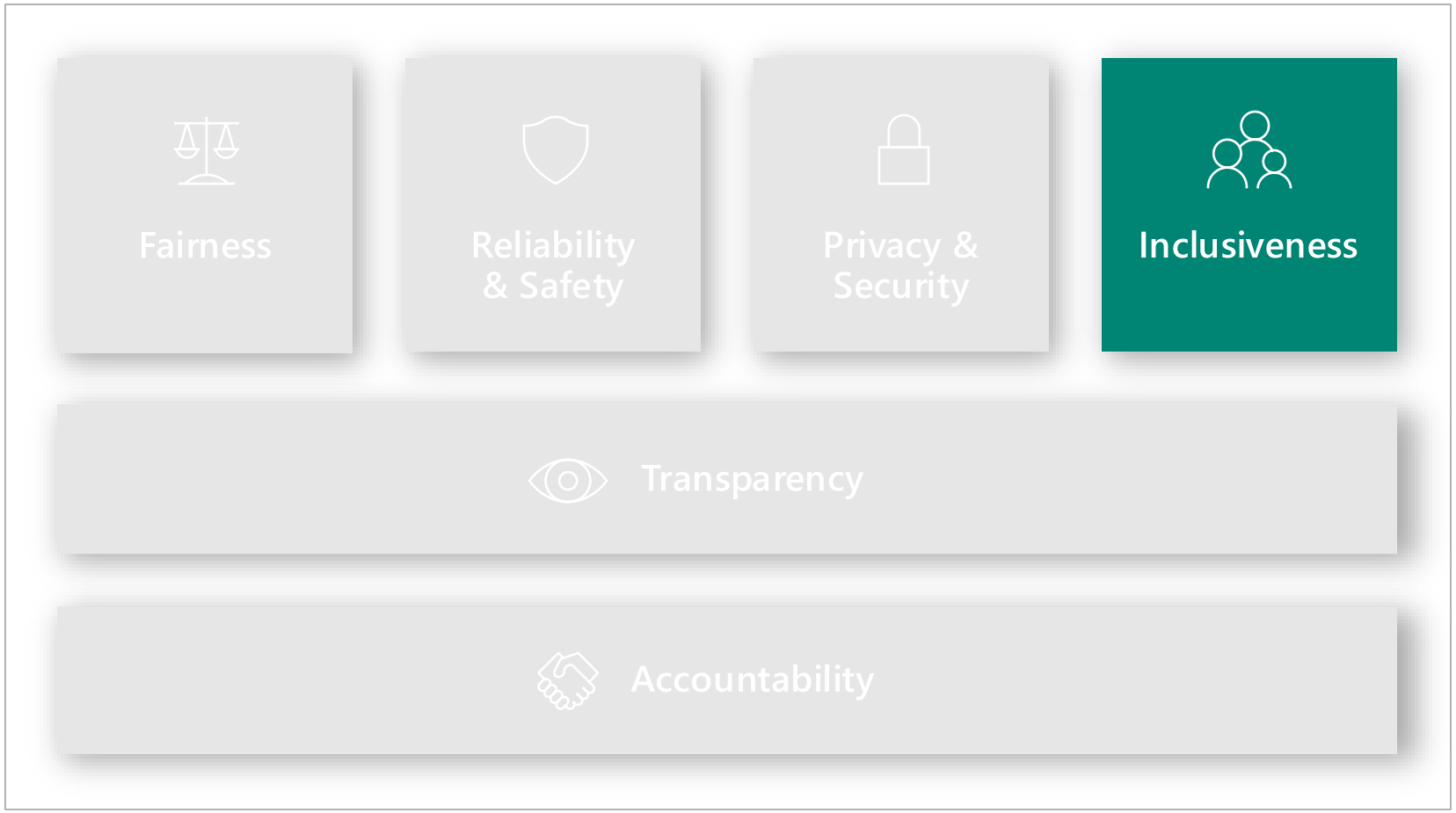
Company suspends use of mall directory cameras running facial recognition software

Cadillac Fairview had been testing the technology since June, but didn't tell mall patrons

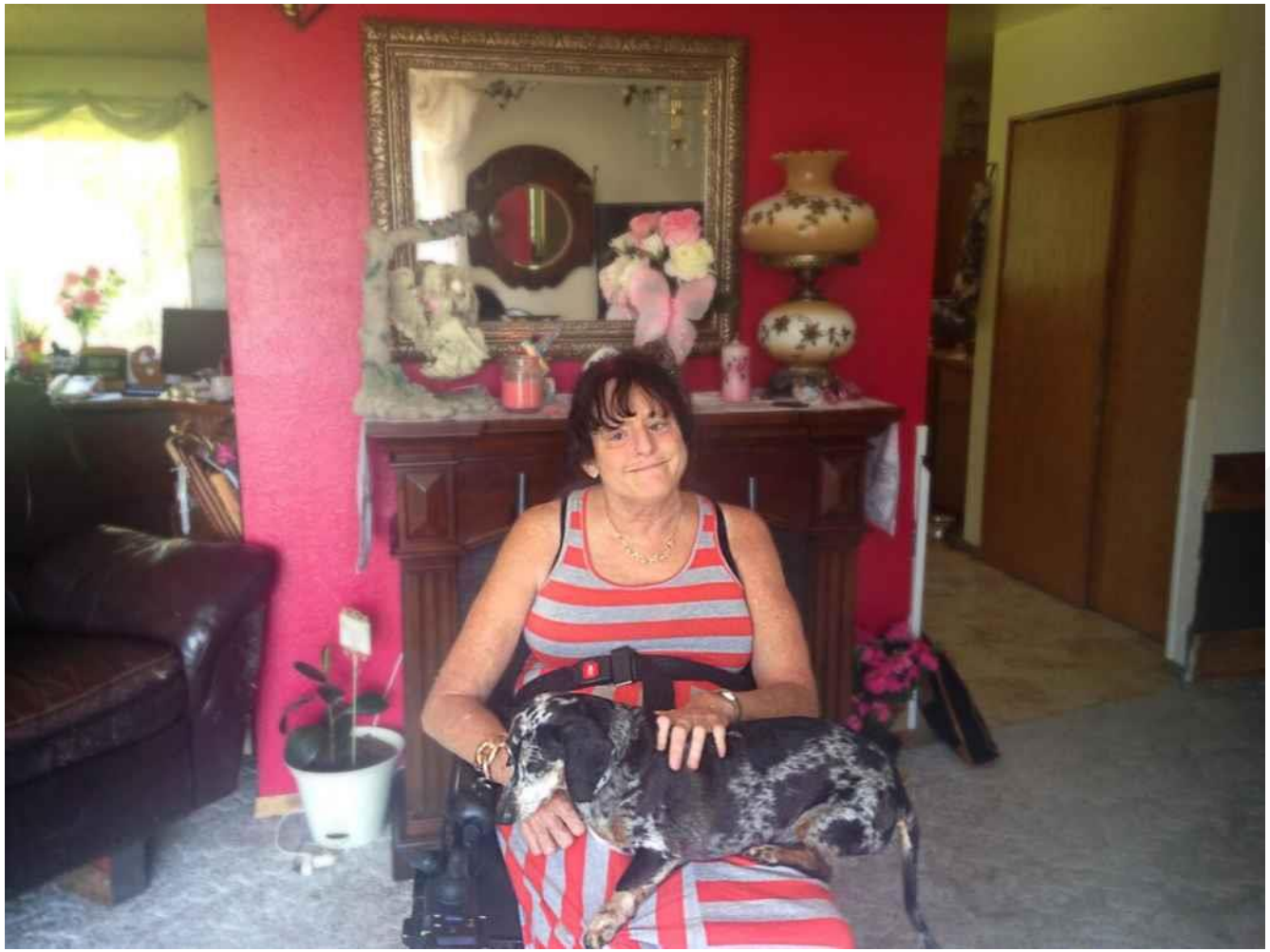
Anis Heydari · CBC News · Posted: Aug 04, 2018 6:08 PM MT | Last Updated: August 4, 2018

This mall directory at Cadillac Fairview's Chinook Centre in Calgary has a camera embedded within it, as circled in red on the left. (Anis Heydari/CBC)

Microsoft's AI principles



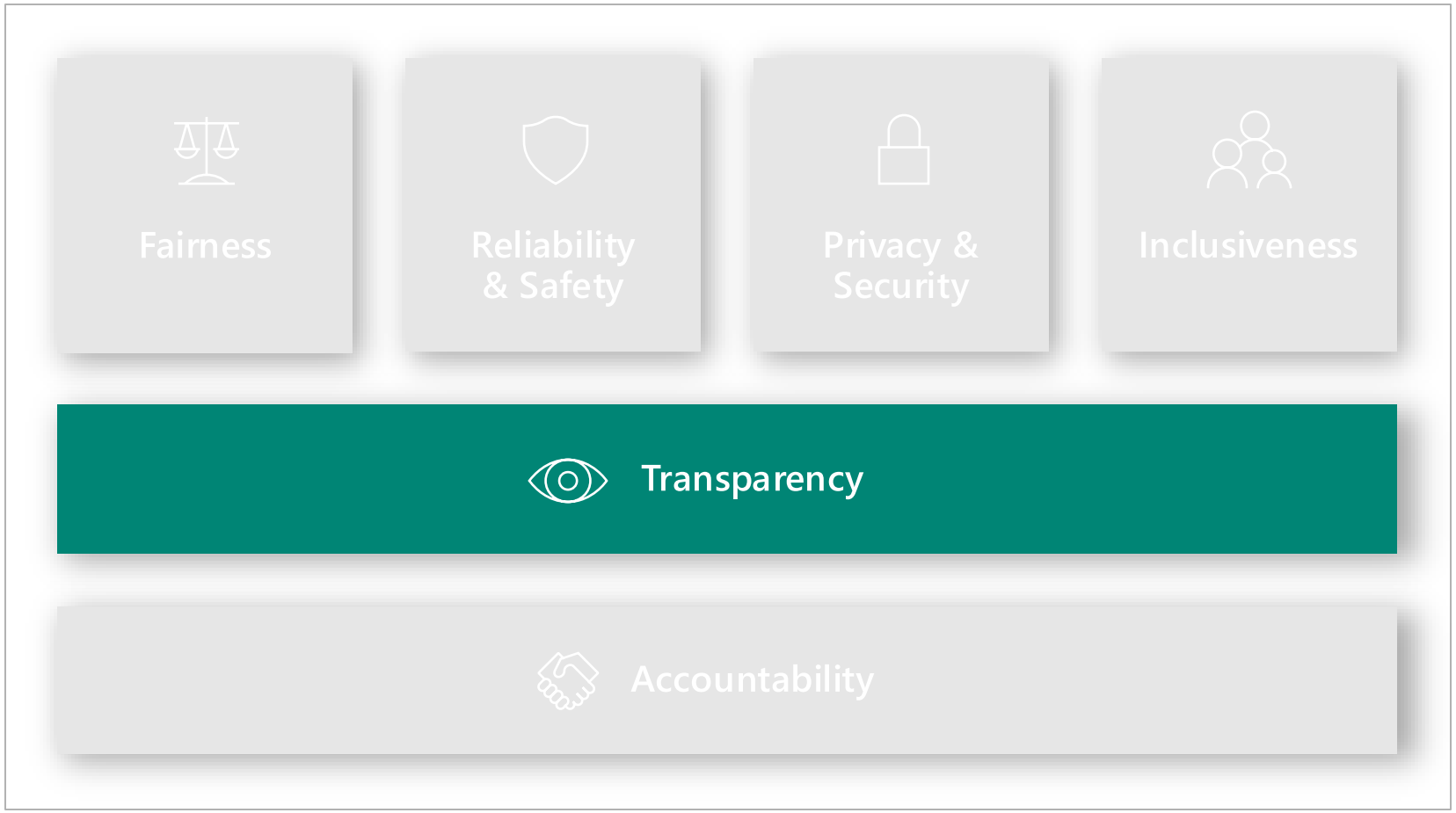
Inclusiveness



K. W. v. Armstrong plaintiff Christie Mathwig

https://www.acluidaho.org/sites/default/files/field_documents/class_action_complaint.pdf

Microsoft's AI principles



Transparency

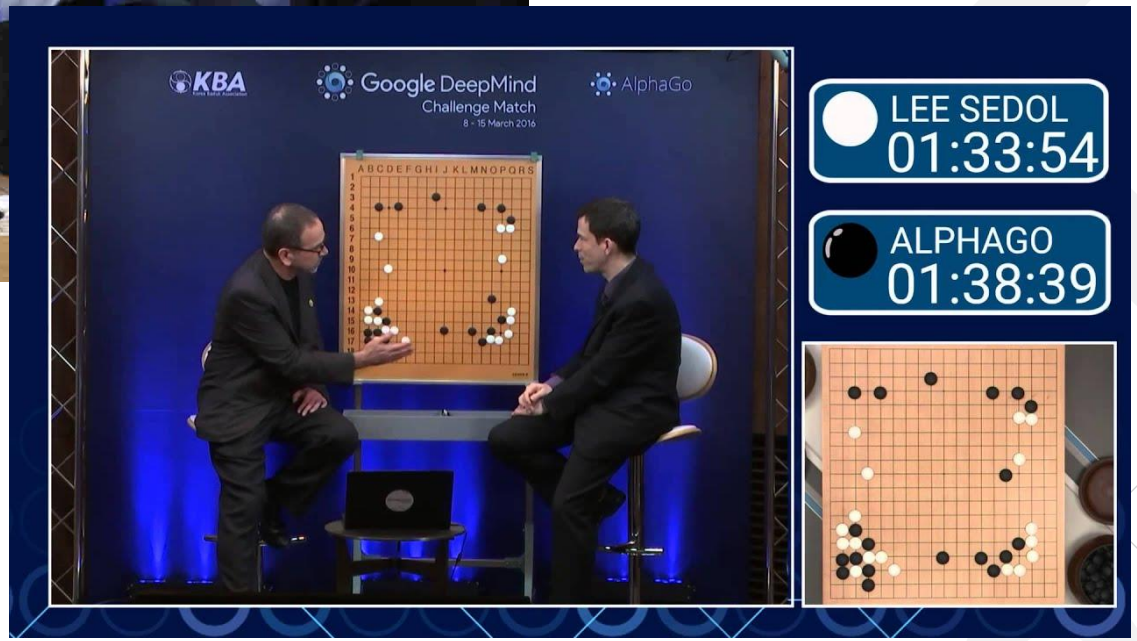
```
HasAsthma(x) => LessRisk(x)
```

Pneumonia study

Physical examination findings	
Respiration rate (resps/min)	≤29*, ≥30
Heart rate (beats/min)	≤124*, 125–150, ≥151
Systolic blood pressure (mmHg)	≤60, 61–70, 71–80, 81–90, ≥91*
Temperature (°C)	≤34.4, 34.5–34.9, 35–35.5, 35.6–38.3*, 38.4–39.9, ≥40
Altered mental status (disorientation, lethargy, or coma)	no*, yes
Wheezing	no*, yes
Stridor	no*, yes
Heart murmur	no*, yes
Gastrointestinal bleeding	no*, yes
Laboratory findings	
Sodium level (mEq/l)	≤124, 125–130, 131–149*, ≥150
Potassium level (mEq/l)	≤5.2*, ≥5.3
Creatinine level (mg/dl)	≤1.6*, 1.7–3.0, 3.1–9.9, ≥10.0
Glucose level (mg/dl)	≤249*, 250–299, 300–399, ≥400
BUN level (mg/dl)	≤29*, 30 to 49, ≥50
Liver function tests (coded only as normal* or abnormal)	SGOT ≤63 and alkaline phosphatase ≤499*, SGOT >63 or alkaline phosphatase >499
Albumin level (gm/dl)	≤2.5, 2.6–3, ≥3.1*
Hematocrit	6–20, 20.1–24.9, 25–29, ≥30*
White blood cell count (1000 cells/μl)	0.1–3, 3.1–19.9*, ≥20
Percentage bands	≤10*, 11–20, 21–30, 31–50, ≥51
Blood pH	≤7.20, 7.21–7.35, 7.36–7.45*, ≥7.46
Blood pO_2 (mmHg)	≤59, 60–70, 71–75, ≥76*
Blood pCO_2 (mmHg)	≤44*, 45–55, 56–64, ≥65

http://people.dbmi.columbia.edu/noemie/papers/15kdd.pdf

Transparency



Move 37

Transparency

THE VERGE

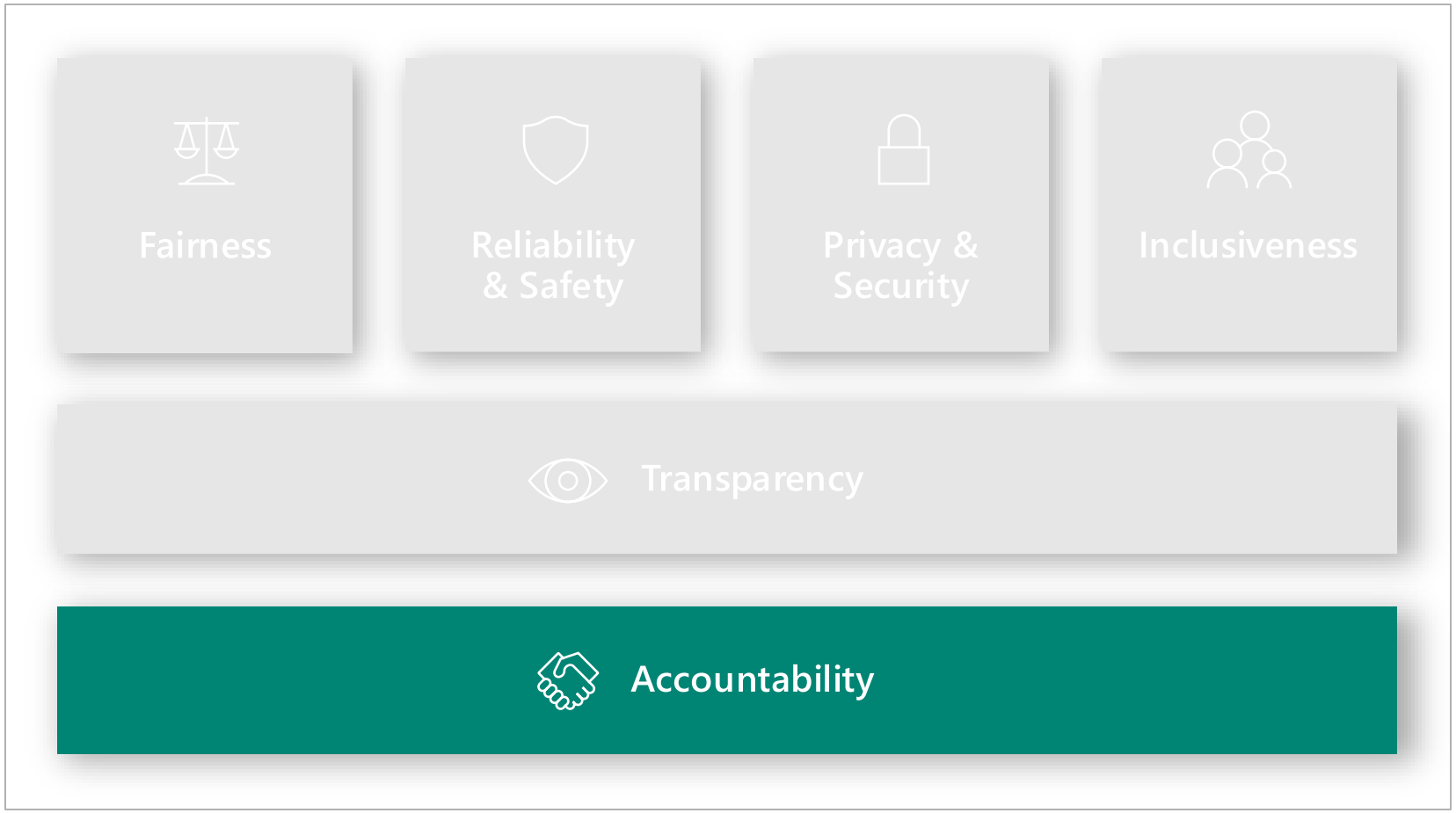
POLICY & LAW

PALANTIR HAS SECRETLY BEEN USING NEW ORLEANS TO TEST ITS PREDICTIVE POLICING TECHNOLOGY

Palantir deployed a predictive policing system in New Orleans that even city council members don't know about

By Ali Winston | Feb 27, 2018, 3:25pm EST

Microsoft's AI principles



Accountability

MIT Technology Review

Ethical Tech / AI Ethics

When algorithms mess up, the nearest human gets the blame

A look at historical case studies shows us how we handle the liability of automated systems.

by **Karen Hao** May 28, 2019

Accountability

UK Official Says It's Too Expensive to Delete All the Mugshots of Innocent People in Police Databases

Sidney Fussell
4/19/18 2:30pm • Filed to: SURVEILLANCE

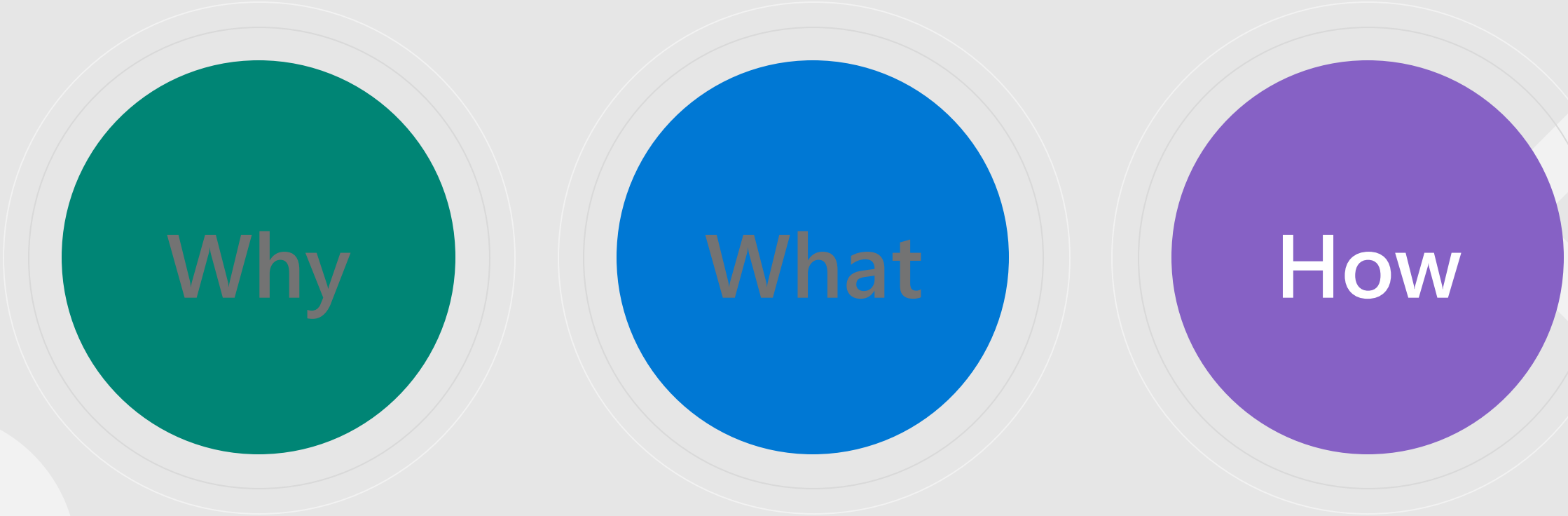
16 4

A police officer watches a television monitor displaying a fraction of London's CCTV camera network
Photo: Daniel Berehulak (Getty)

Accountability

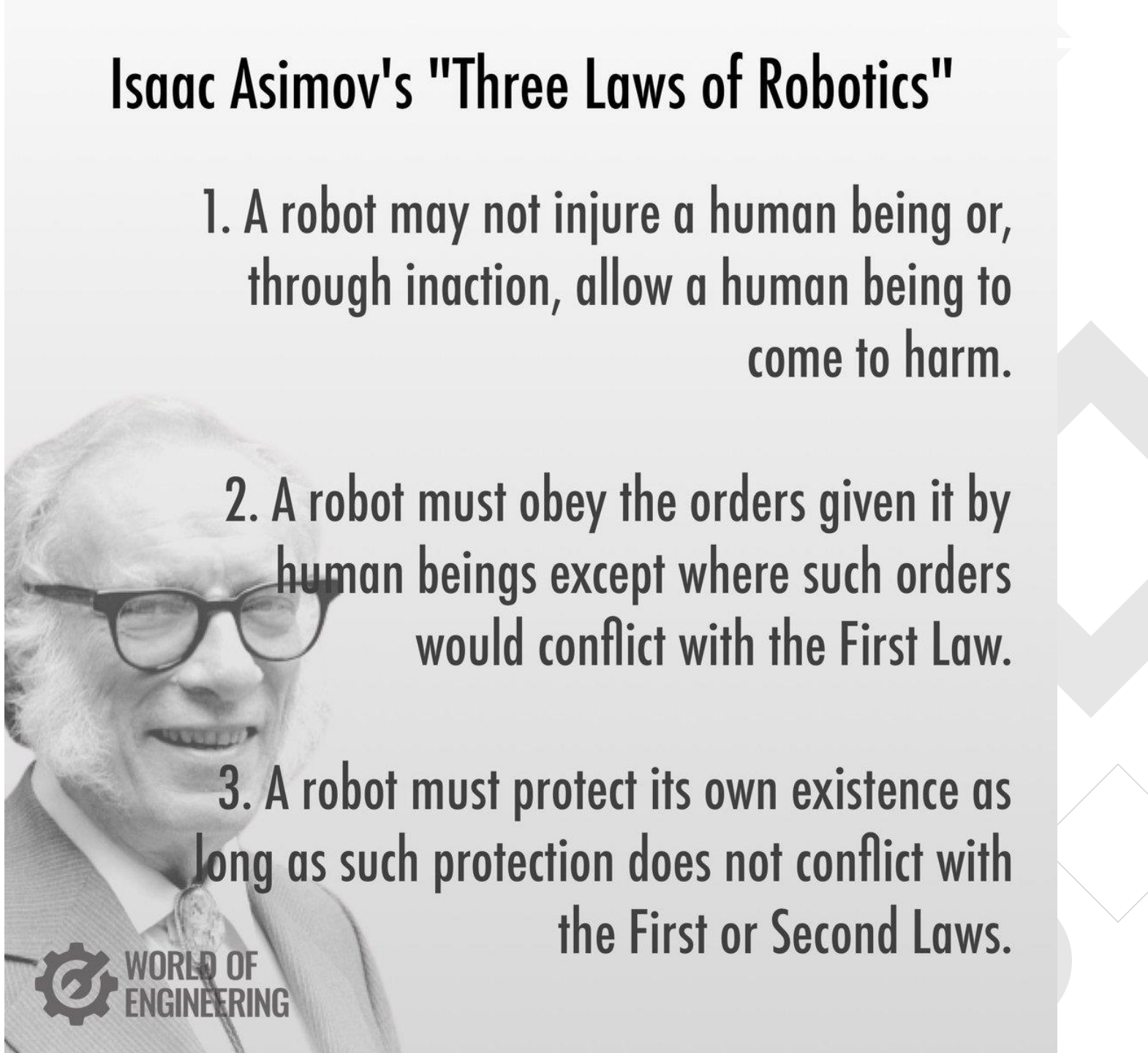
Who to Sue when a Robot loses your Fortune
Bloomberg, May 2019

Agenda _Responsible AI

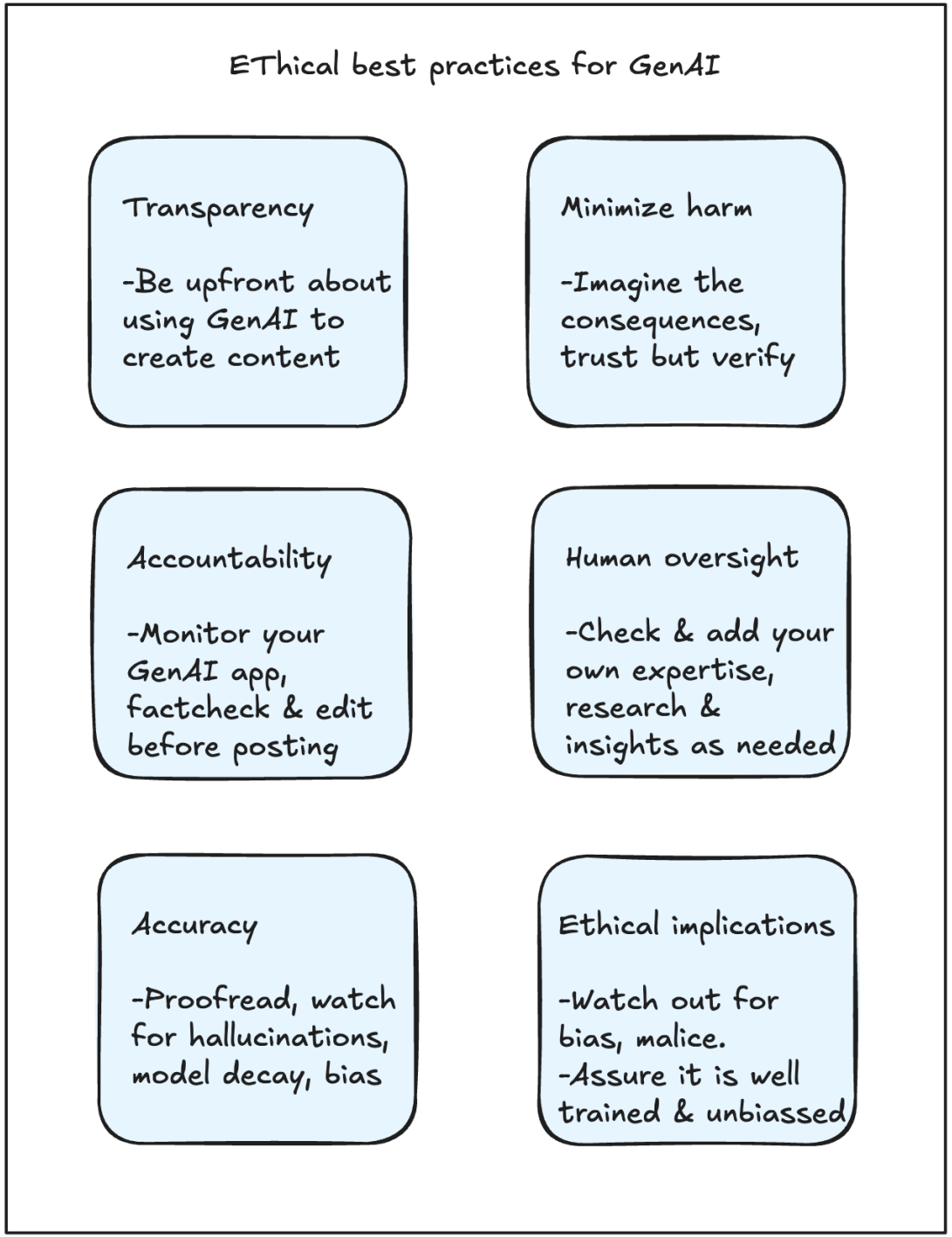


Putting Responsible AI into Practice

3 Laws of Robotics



GenAI for Good



Responsible AI & Human Rights

- Dignity of every individual
- Freedom from discrimination
- Freedom from invasions of privacy
- Freedom of expression
- Freedom of association

Today's debate

https://www.whitehouse.gov/ostp/ai-bill-of-rights/

Putting responsible AI into practice

Putting responsible AI into practice

Human-AI Guidelines

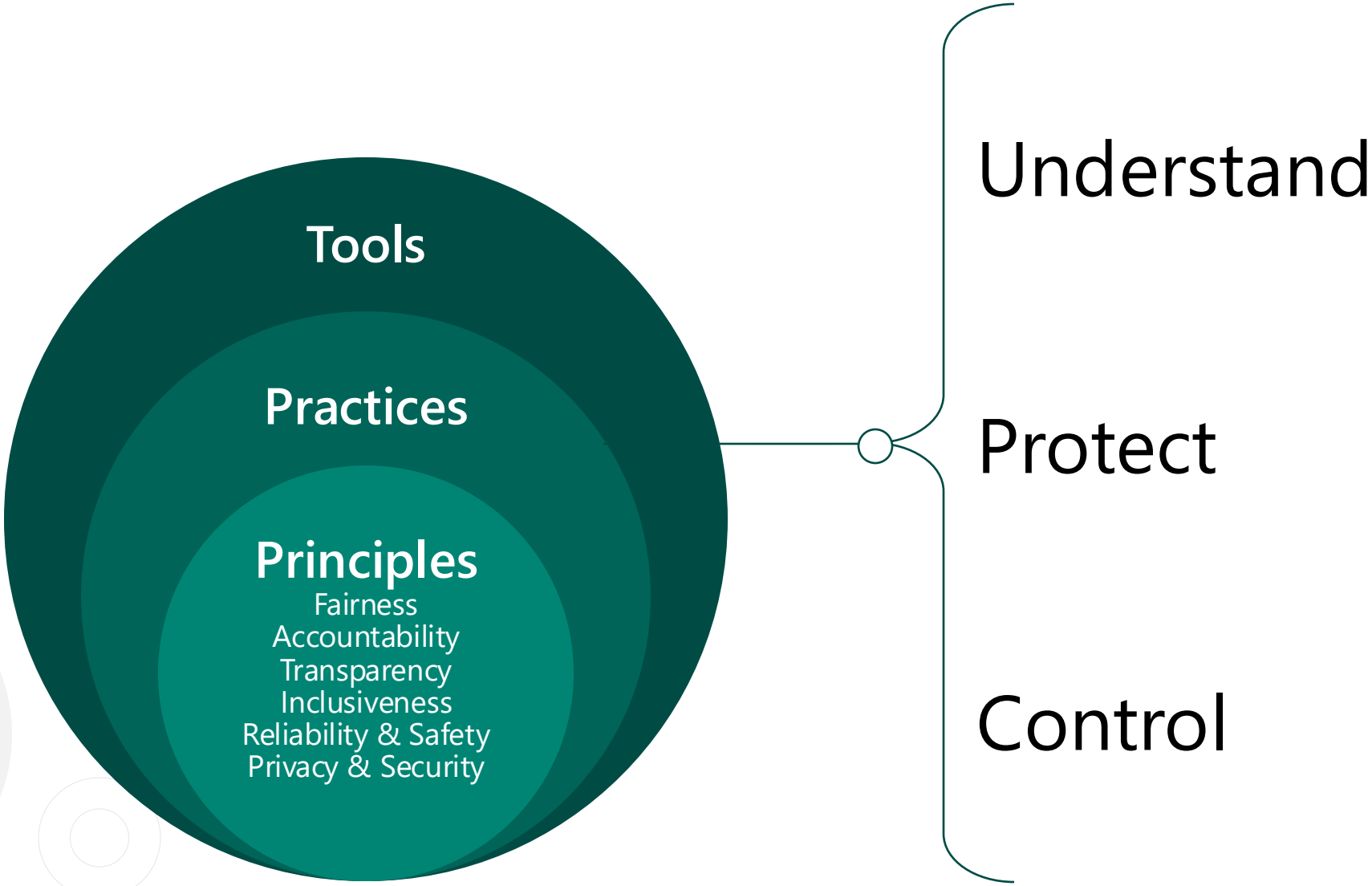
Conversational AI Guidelines

Inclusive Design Guidelines

AI Fairness Checklist

Datasheets for Datasets

Putting responsible AI into practice



Putting responsible AI into practice

Chief RAI Officer

RAI Office

RAI Committee

AI Handbook

Responsible AI by Design

- ✓ AI is embedded in everyday life, business, government, medicine and more.
- ✓ You will be helping people and organizations adopt AI responsibly.
- ✓ Only by embedding ethical principles into AI applications and processes can we build systems based on trust.

Q&A