Microsoft

Responsible AI

Opportunity and Responsibility in the Era of Artificial Intelligence

Ivan Portilla
portilla@gmail.com
AI Leader
Microsoft

Agenda

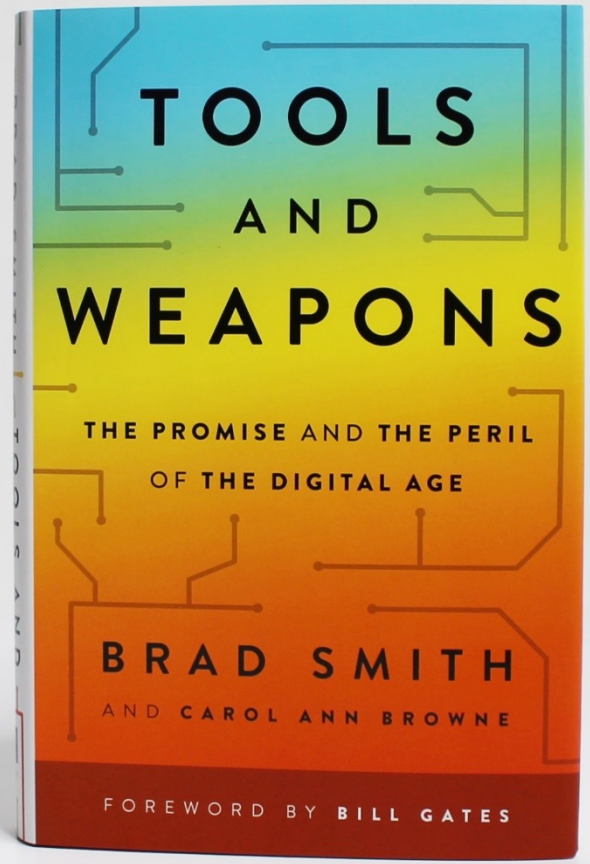
- ✓ Why responsible AI
- ✓ Responsible AI principals
- ✓ Putting Responsible AI into Practice

Agenda

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



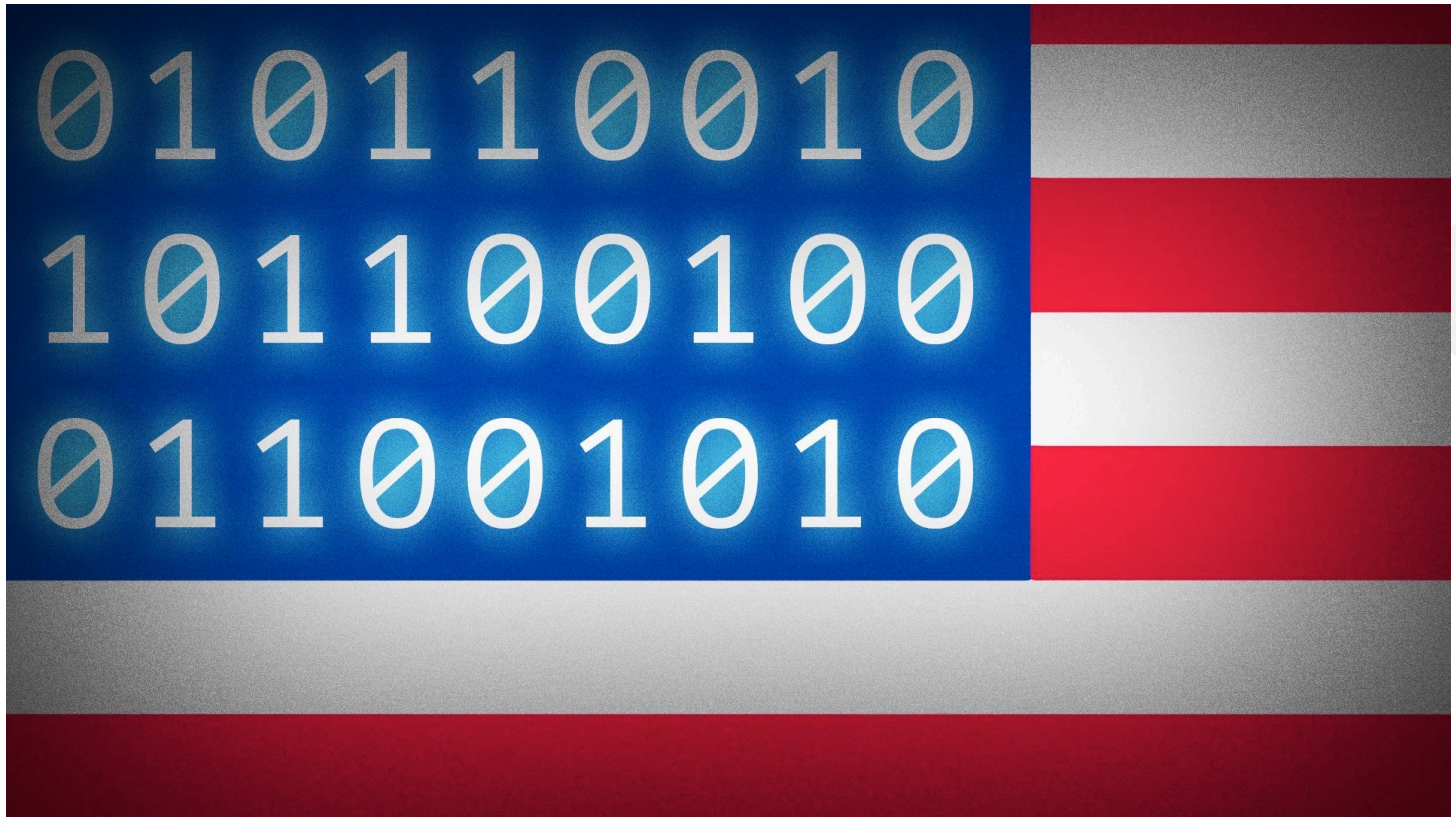
Today's debate

Five takeaways from UK's AI safety summit at Bletchley Park

https://www.theguardian.com/technology/2023/nov/02/five-takeaways-uk-ai-safety-summit-bletchley-park-rishi-sunak

Today’s debate

What's in Biden's AI executive order — and what's not



https://www.axios.com/2023/11/01/unpacking-bidens-ai-executive-order

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

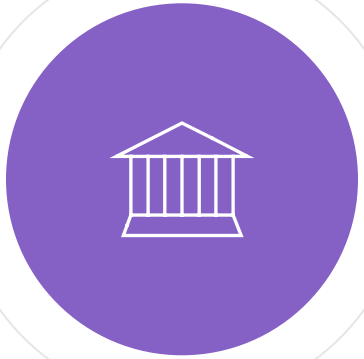
Language Understanding

2019
General Language Understanding human parity

The Opportunities with AI

Healthcare

Retail



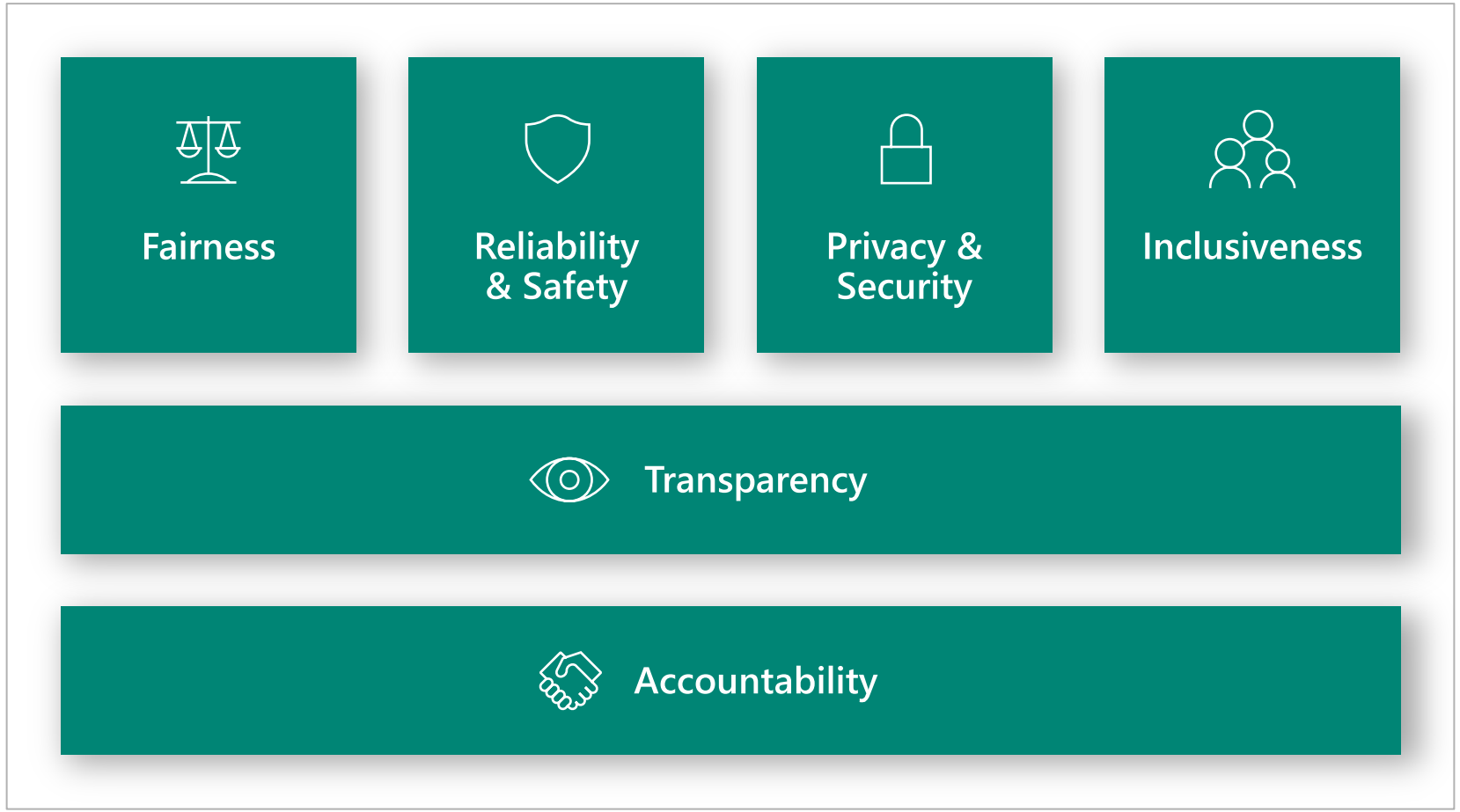
Financial Services

Manufacturing

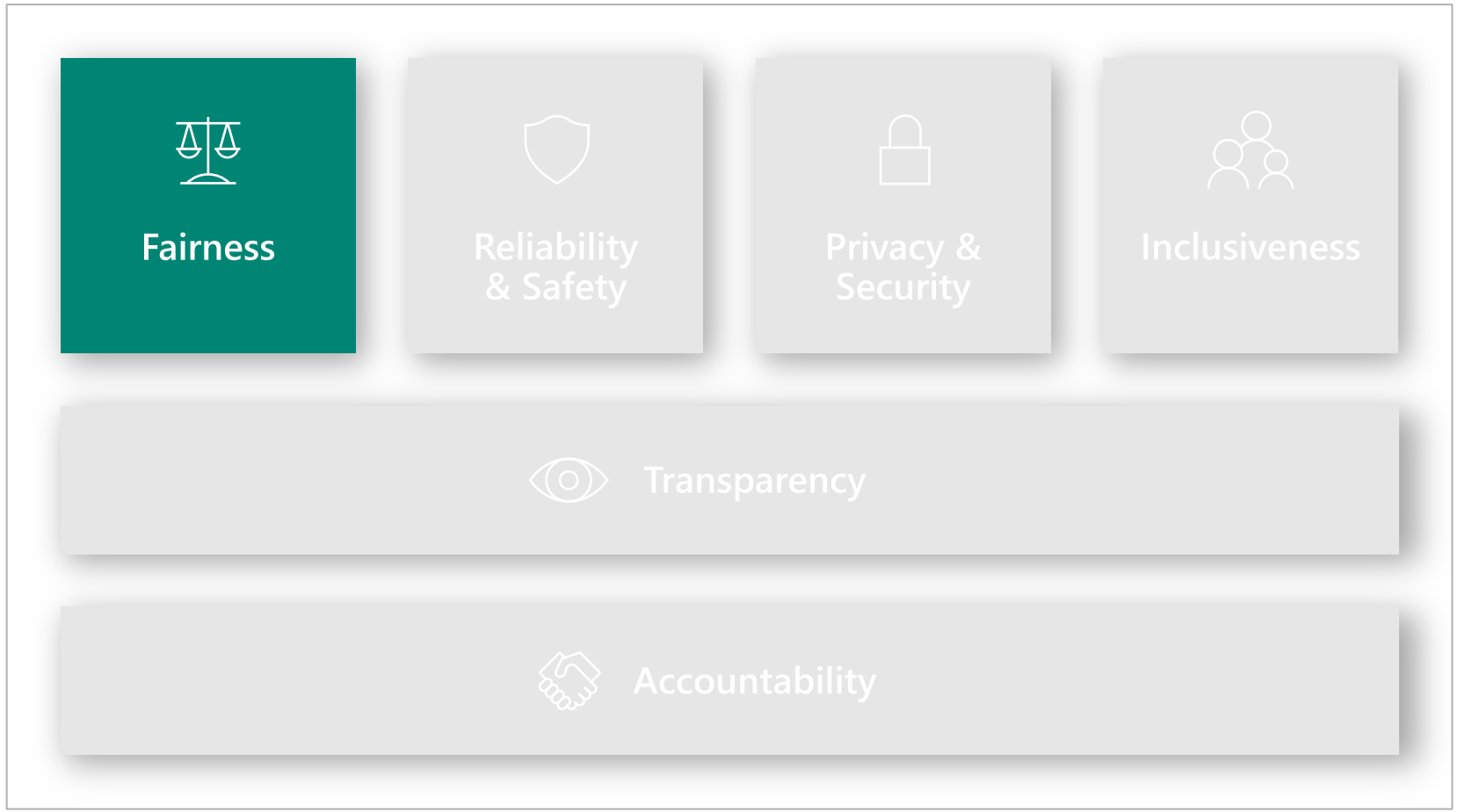
Agenda

Microsoft 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.

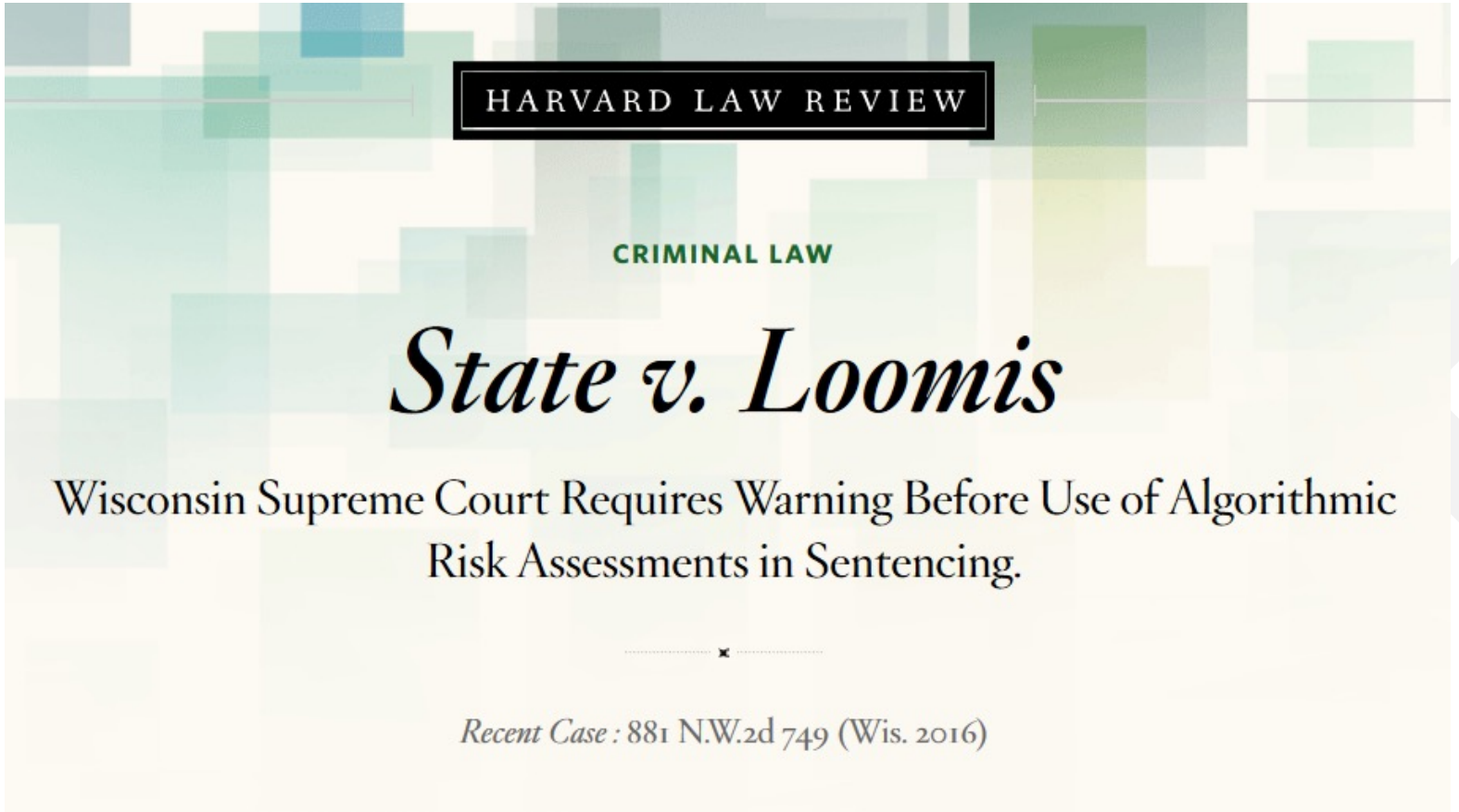
Heidi Ledford

Black people with complex medical needs were less likely than equally ill white people to be referred to programmes that provide more personalized care. Credit: Ed Kashi/VII/Redux/eyevine

Fairness

```
BE_INT
  = f(0.10
          - 0.14 x GENDER_FEMALE
          - 0.13 x AGE_GROUP_30_49
          - 0.70 x AGE_GROUP_50_PLUS
          + 0.16 x NATIONALITY_EU
          - 0.05 x NATIONALITY_THIRD
          + 0.28 x CERTIFICATION_APPRENTICESHIP
          + 0.01 x CERTIFICATION_COMPULSORY_PLUS
          - 0.15 x MANDATORY_CARE
          - 0.34 x RGS_TYP_2
          - 0.18 x RGS_TYP_3
          - 0.83 x RGS_TYP_4
          - 0.82 x RGS_TYP_5
          - 0.67 x DISABLED
          + 0.17 x PROFESSION_PRODUCTION
          - 0.74 x EMPLOYMENT_DAYS_LIMITED
          + 0.65 x FREQUENCY_UNEMPLOYED_1
          + 1.19 x FREQUENCY_UNEMPLOYED_2
          + 0.65 x FREQUENCY_UNEMPLOYED_3_PLUS
          - 0.80 x UNEMPLOYED_LONG
          - 0.57 x MN_PARTICIPATION_1
          - 0.21 x MN_PARTICIPATION_2
          - 0.43 x MN_PARTICIPATION_3)
```

Fairness


HARVARD LAW REVIEW
CRIMINAL LAW
State v. Loomis
Wisconsin Supreme Court Requires Warning Before Use of Algorithmic Risk Assessments in Sentencing.
Recent Case : 881 N.W.2d 749 (Wis. 2016)

Fairness

3 Low Risk

DYLAN FUGETT

Prior Offense
1 attempted burglary

Subsequent Offenses
3 drug possessions

10 High Risk

BERNARD PARKER

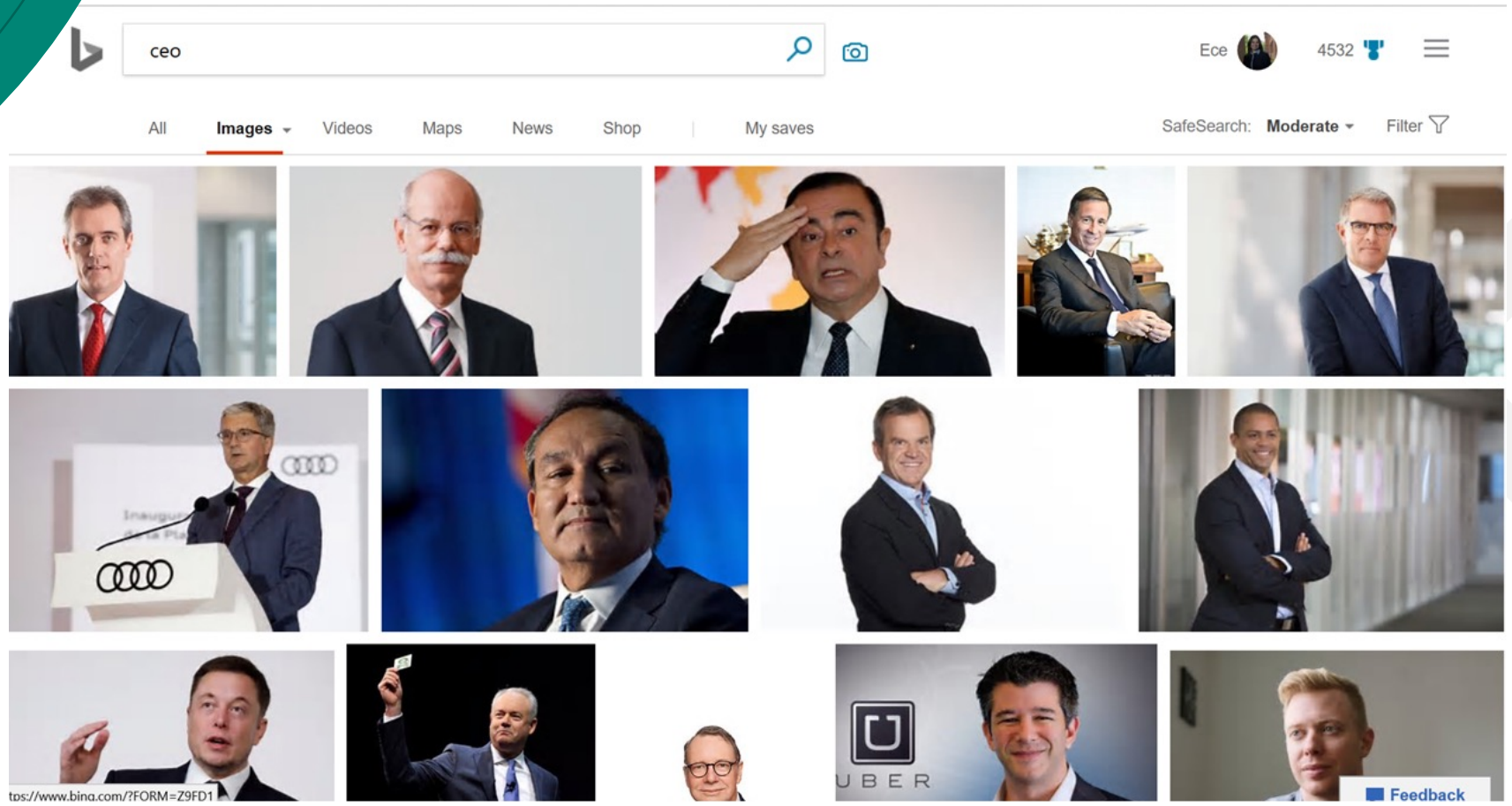
Prior Offense
1 resisting arrest without violence

Subsequent Offenses
None

Fairness


STEPHEN GAINES
A736258/T:1.23
SARA WILLIAMS
S683529/T:0.57
CALCU

Fairness


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

Fairness

Is there gender bias in lending?

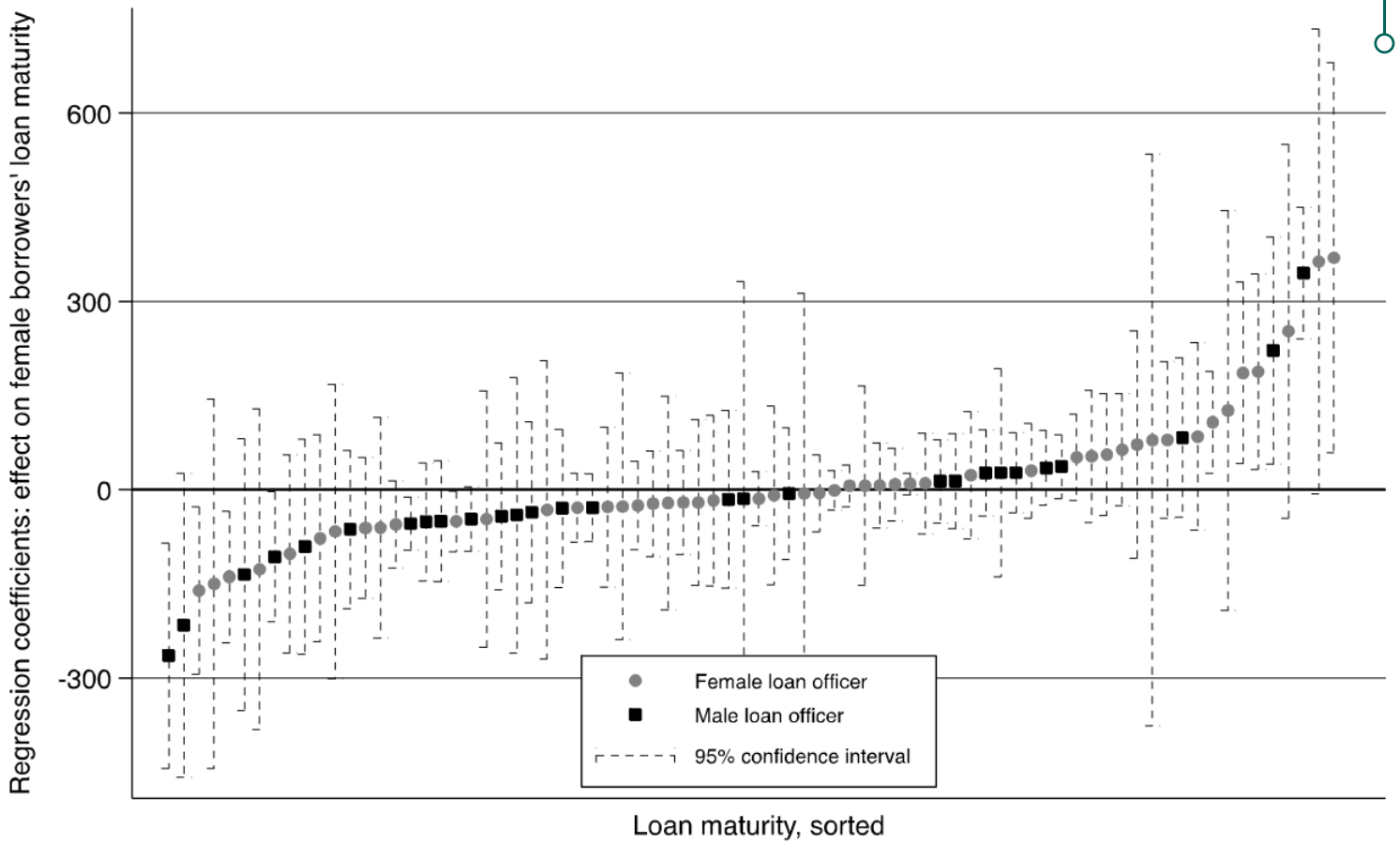


Figure 3. The figure shows the distribution of the bias on loan maturity by officer gender. Each coefficient represents an estimate of the number of extra days of loan maturity an individual officer approves for female versus male borrowers.

Source: SSRN, Sex and Credit: Is there gender bias in lending?, Beck, Behr & Madestom, April 2017

Fairness

The New York Times

Facial Recognition Is Accurate, if You're a White Guy

By STEVE LOHR FEB. 9, 2018

Gender was misidentified in **up to 1 percent of lighter-skinned males** in a set of 385 photos.

Gender was misidentified in **up to 12 percent of darker-skinned males** in a set of 318 photos.

Gender was misidentified in **up to 7 percent of lighter-skinned females** in a set of 296 photos.

Gender was misidentified in **35 percent of darker-skinned females** in a set of 271 photos.

Fairness

✕ The photo you want to upload does not meet our criteria because:

- Subject eyes are closed

Please refer to the technical requirements.
You have 9 attempts left.

Check the photo requirements.

Read more about common photo problems and how to resolve them.

After your tenth attempt you will need to start again and re-enter the CAPTCHA security check.

Reference number: 20161206-81

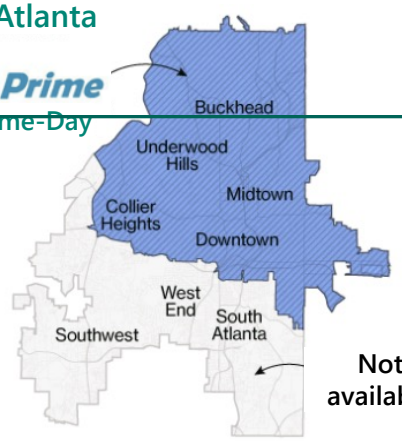
Filename: Untitled.jpg

If you wish to contact us about the photo, you must provide us with the reference number given above.

Fairness

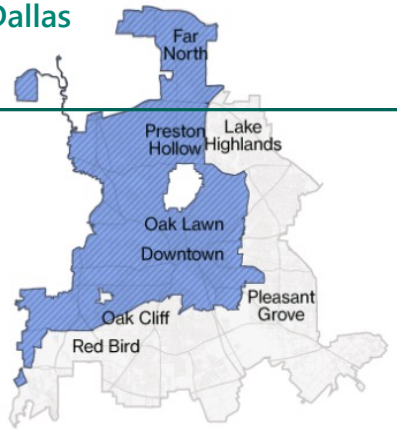
Atlanta

amazon Prime Same-Day



Not available

Dallas

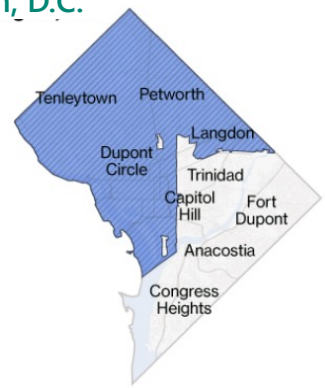


Boston

Chicago

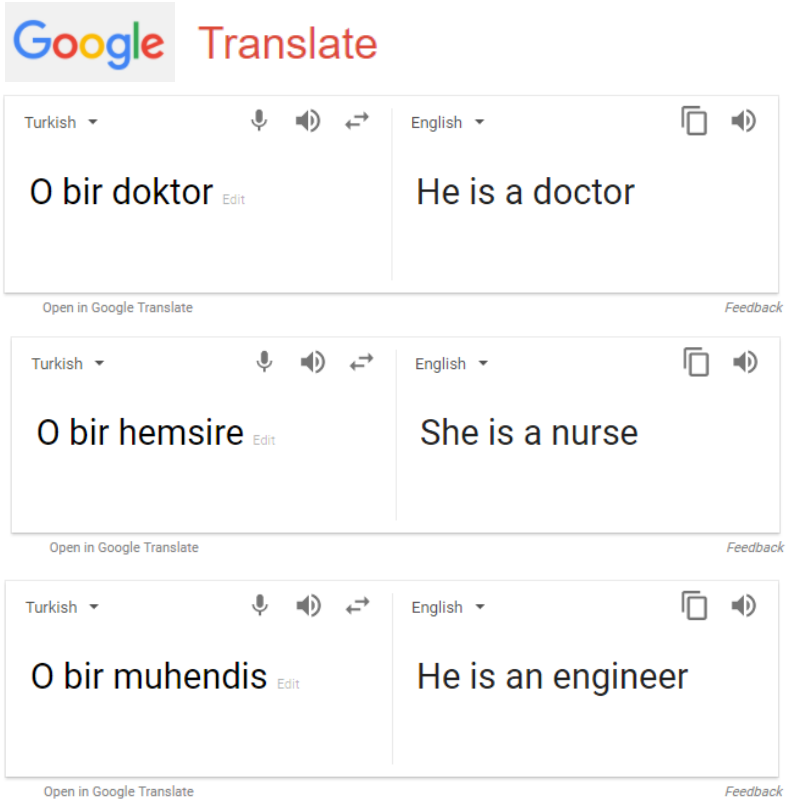
New York City

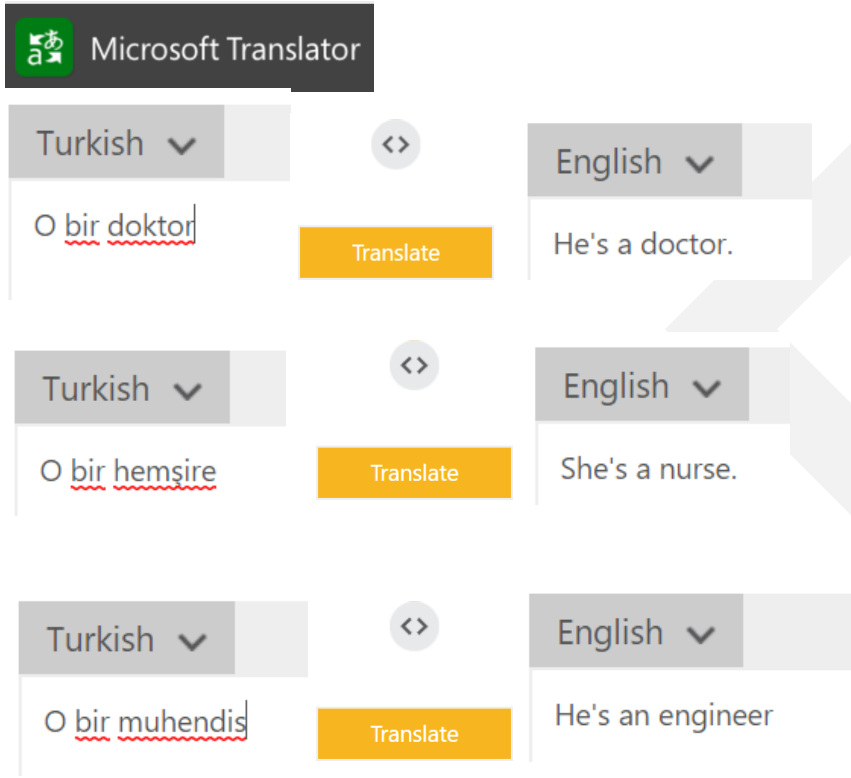
Washington, D.C.



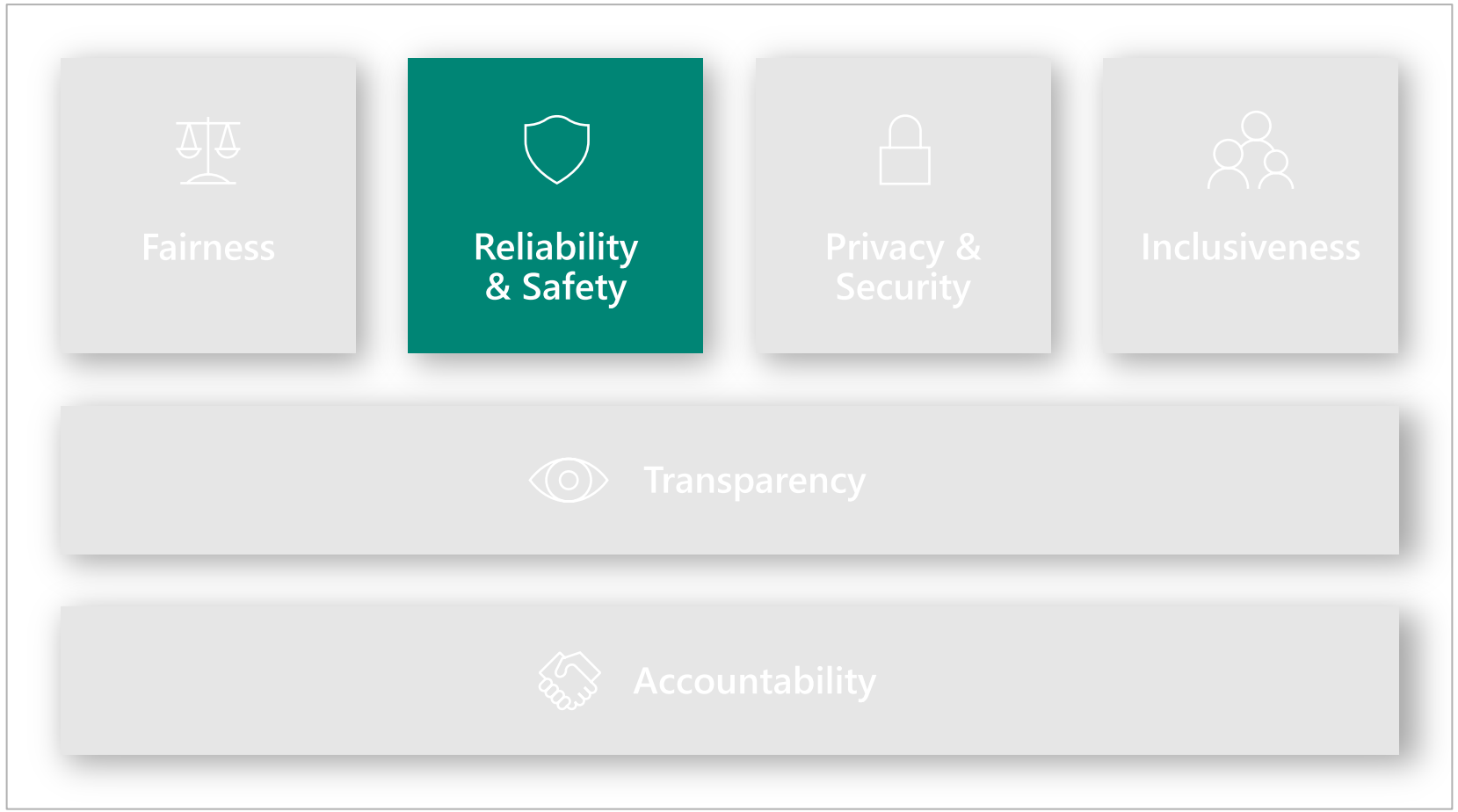
https://www.bloomberg.com/graphics/2016-amazon-same-day/

Fairness

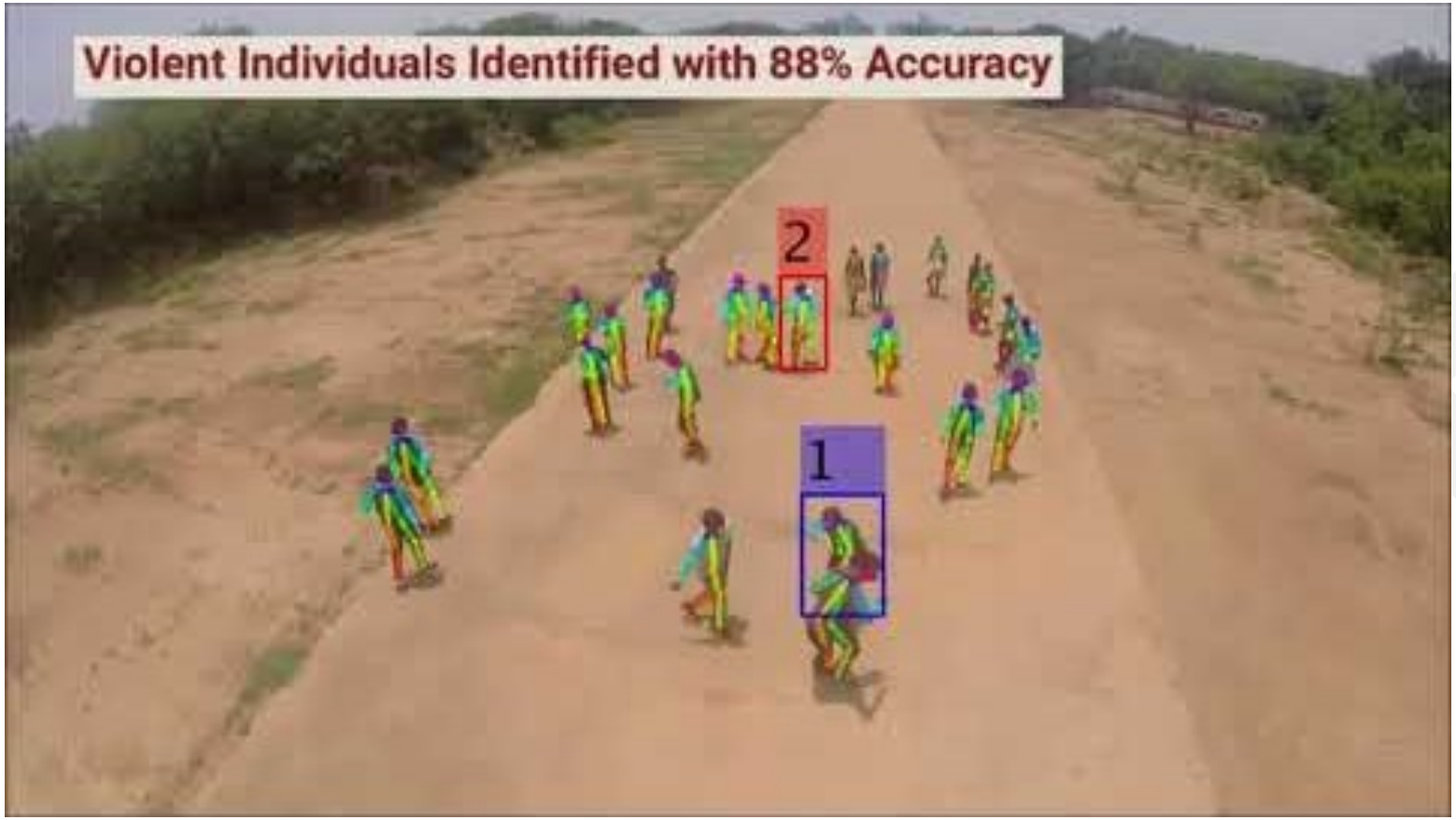

Google Translate
Turkish
English
O bir doktor Edit
He is a doctor
Open in Google Translate
Feedback
Turkish
English
O bir hemsire Edit
She is a nurse
Open in Google Translate
Feedback
Turkish
English
O bir muhendis Edit
He is an engineer
Open in Google Translate
Feedback


Microsoft Translator
Turkish
English
O bir doktor
Translate
He's a doctor.
Turkish
English
O bir hemsire
Translate
She's a nurse.
Turkish
English
O bir muhendis
Translate
He's an engineer

Microsoft's AI principles



Reliability & Safety



Reliability & Safety

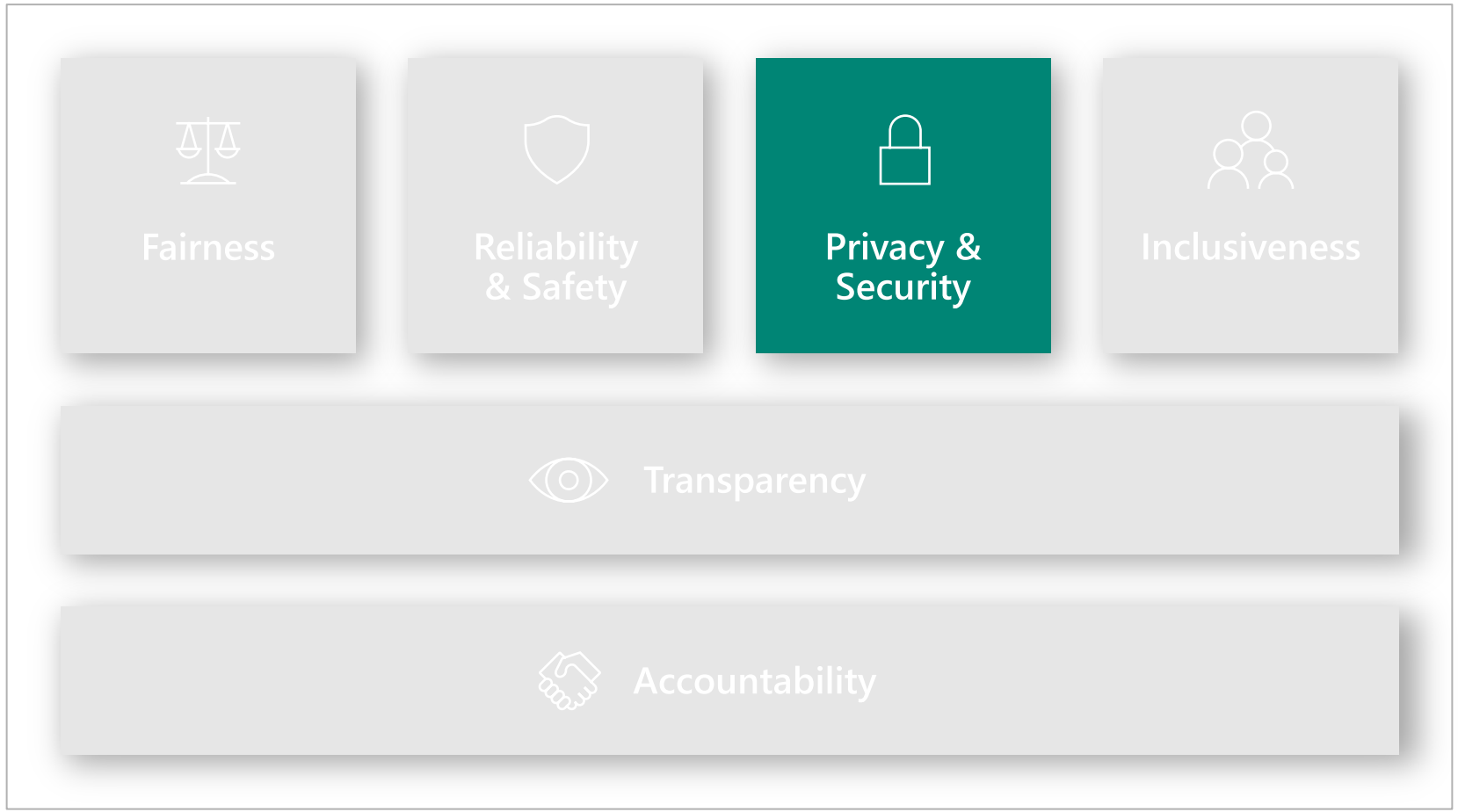
Miles & years needed for to demonstrate autonomous vehicle reliability

	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



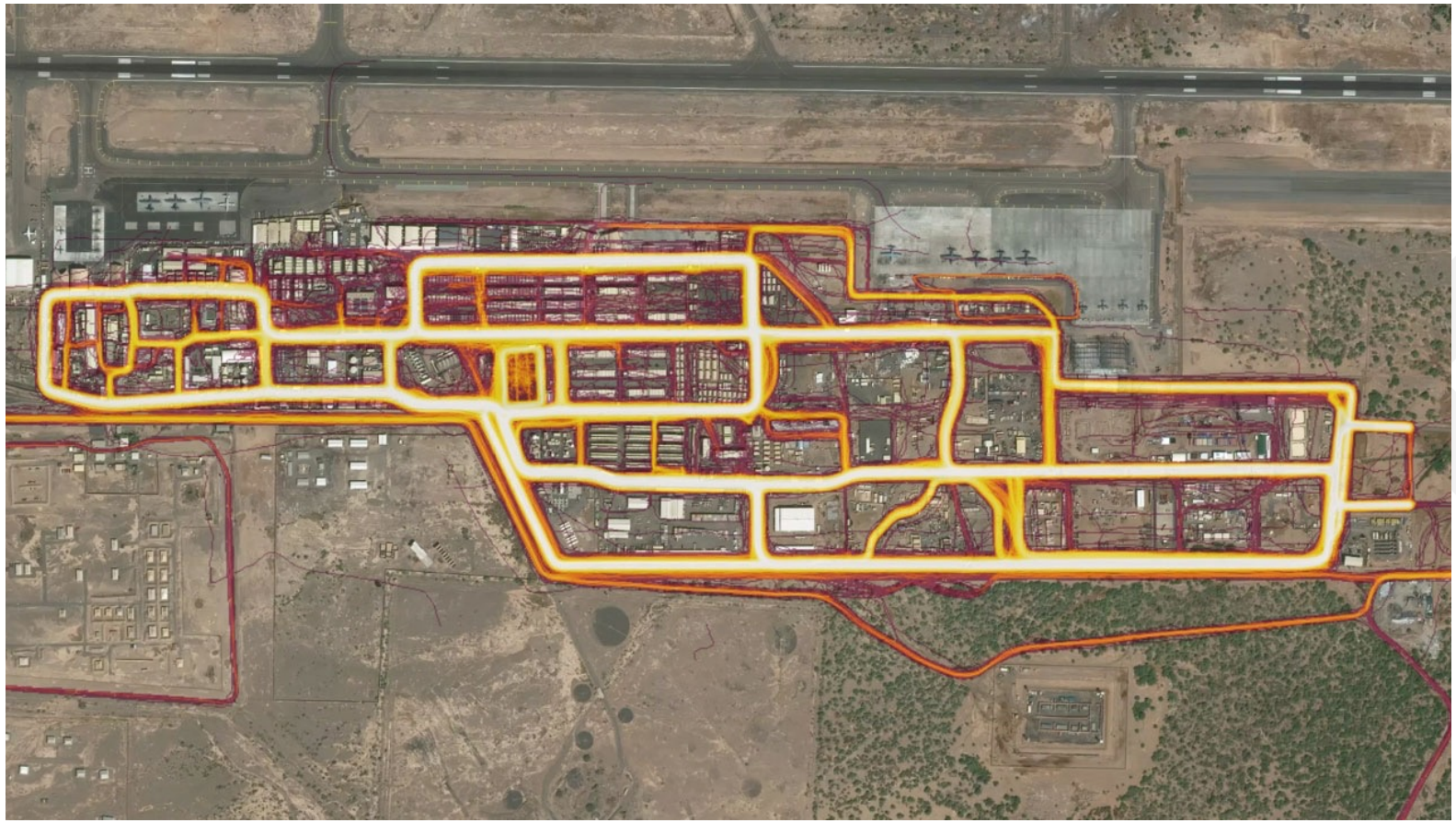
Privacy & Security

> 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



Privacy & Security

NIAGARA GAZETTE
SINCE 1854
Niagara Falls, New York

No start date yet for Lockport schools' facial recognition system

SCHOOLS: Superintendent says privacy matters related to system are still being discussed with state.

LOCKPORT — City School District Superintendent Michelle Bradley said this week that the district does not currently have a date for implementation of its new Aegis facial recognition system.

Lockport schools are using $1.4 million of its allocated $4.2 million SmartSchools Bond Act funding to implement a new security camera system equipped with Aegis facial recognition software, which has been purchased from Ontario-based SNTech.

On Monday, Bradley said district officials are still discussing privacy matters with the state education department. She said the district anticipates that the state Board of Regents will be discussing a section of New York education law that deals with the release of personally identifiable information at its next meeting. The board is scheduled to meet on Jan. 14.

Jan 8, 2019

Privacy & Security

CBC

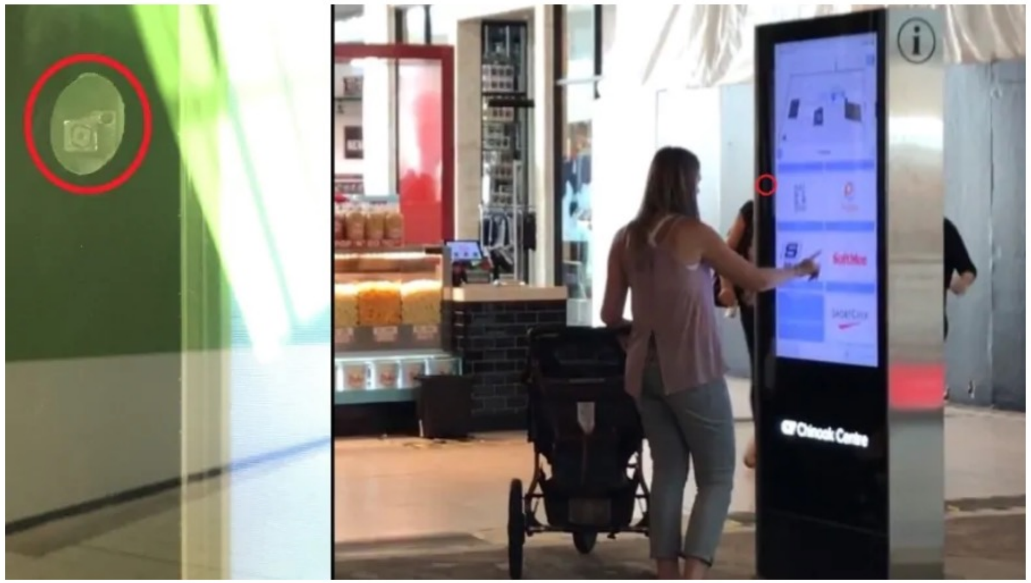
news | Top Stories | Local | The National | Opinion | World | Canada

Calgary

Company suspends use of mall directory cameras running facial recognition software

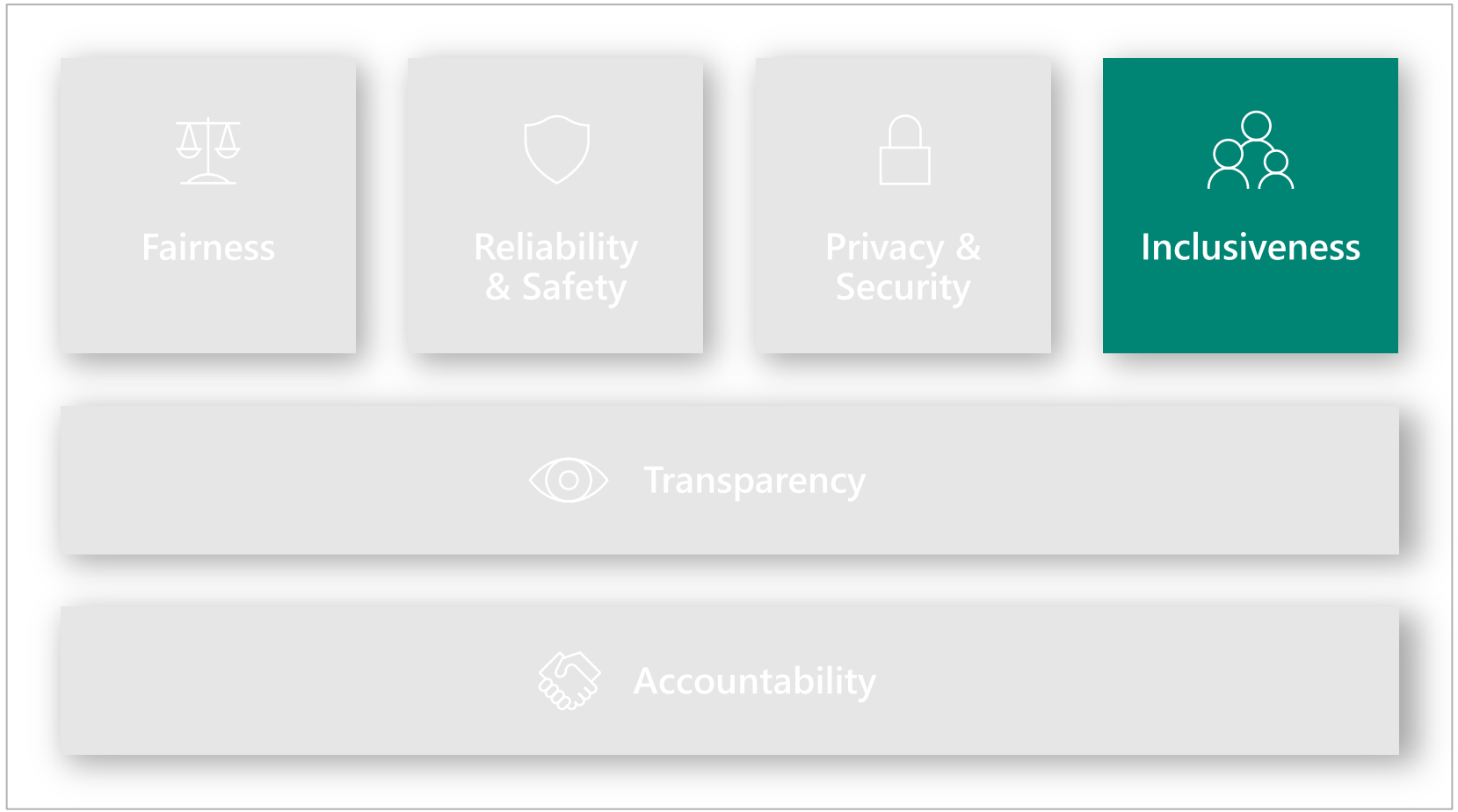
Cadillac Fairview had been testing the technology since June, but didn't tell mall patrons

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

LinkedIn Economic Graph Research

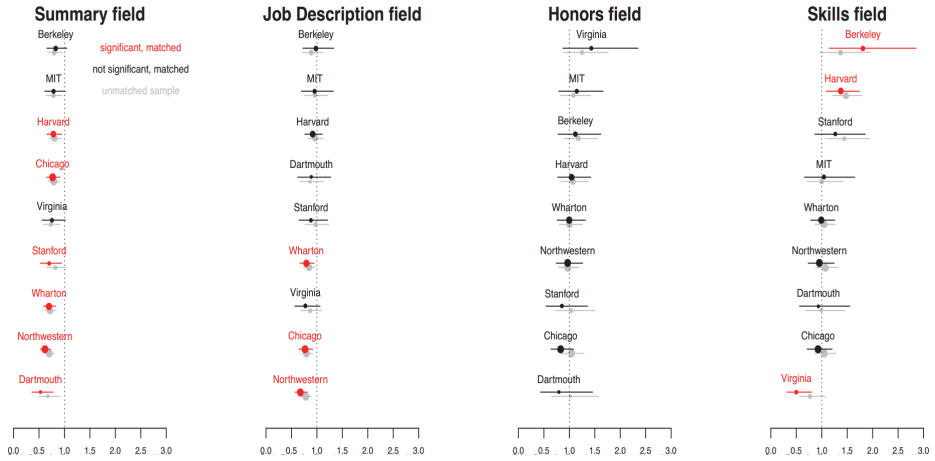
Are There Gender Differences in Professional Self-Promotion?
An Empirical Case Study of LinkedIn Profiles among Recent MBA Graduates

Inclusive design practices to address potential barriers that could unintentionally exclude people

Enhances opportunities for those with disabilities

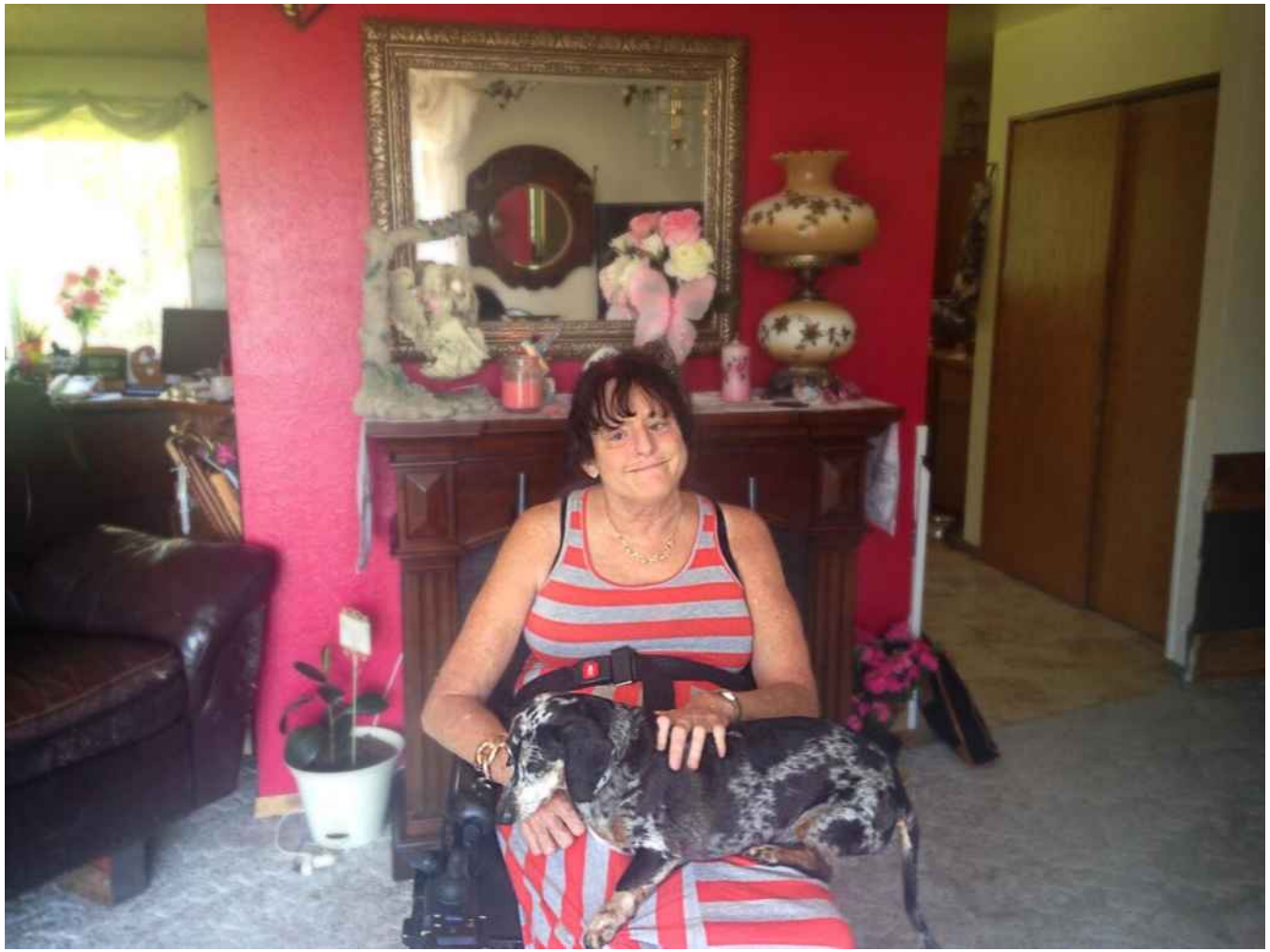
Build trust through contextual interaction

EQ in addition to IQ



Odds ratio for female user vs. male user to include self-promotion fields

Inclusiveness

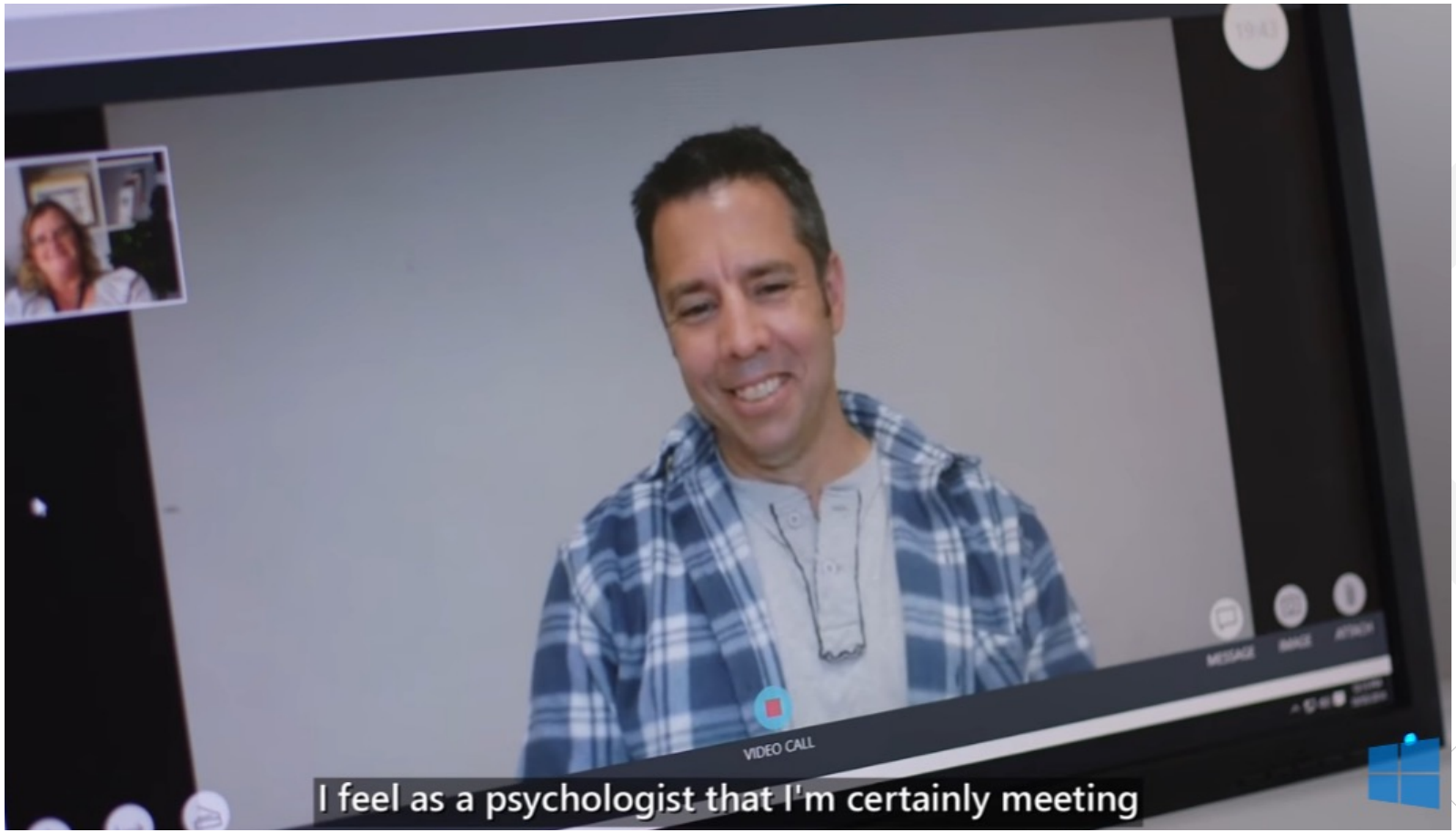


K. W. v. Armstrong plaintiff Christie Mathwig

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

Inclusiveness

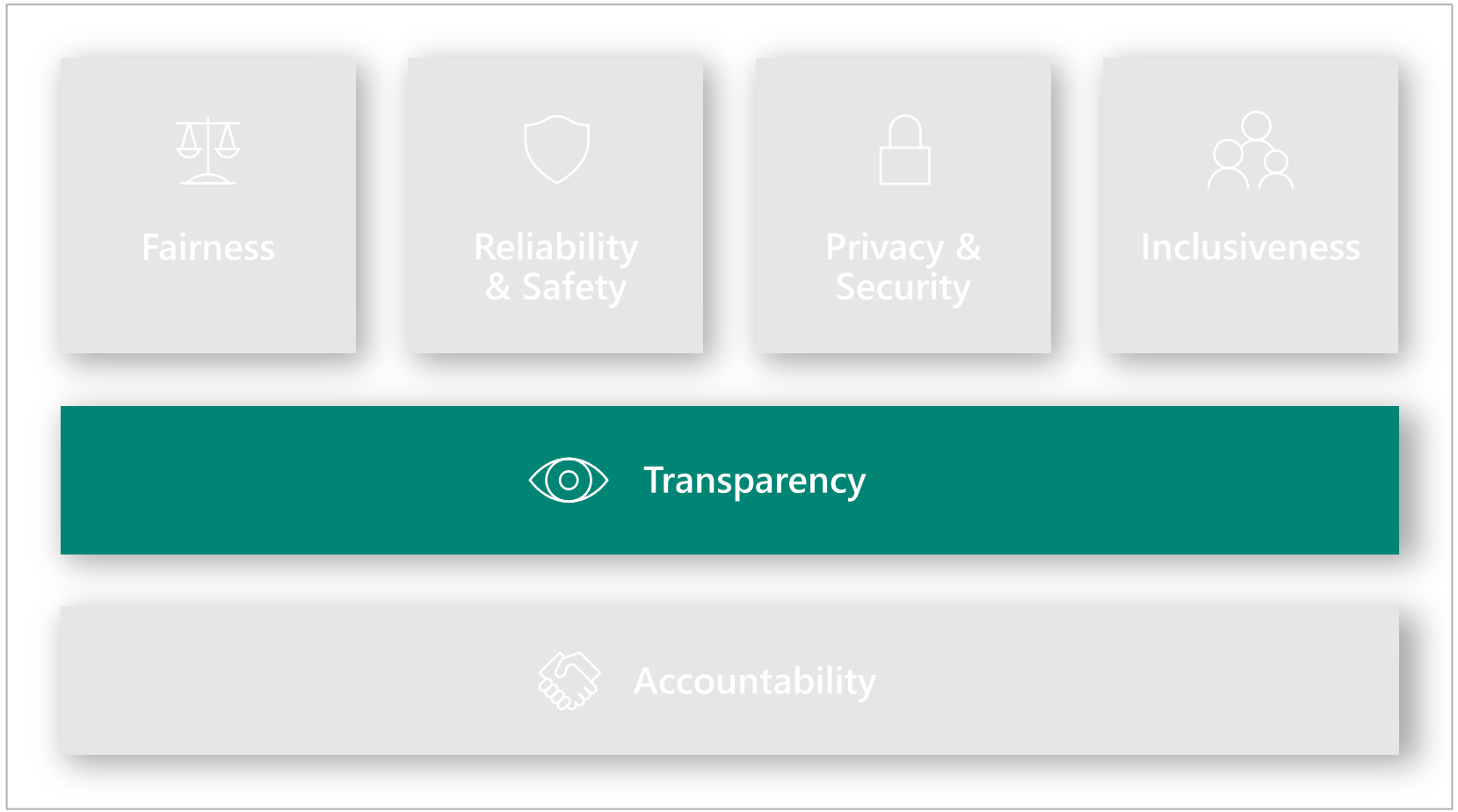
Artificial Intelligence transforms even the most human services



https://news.microsoft.com/en-au/features/artificial-intelligence-transforms-even-human-services/

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

Microsoft's AI principles



Transparency

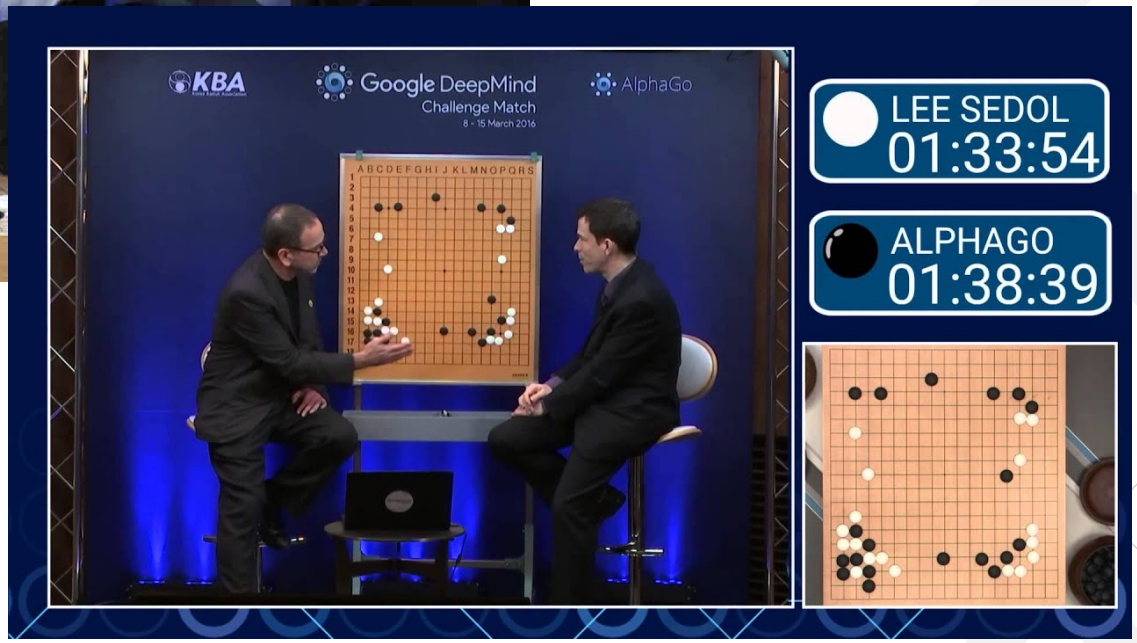
```
HasAsthma(x) => LessRisk(x)
```

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



Transparency

THE VERGE

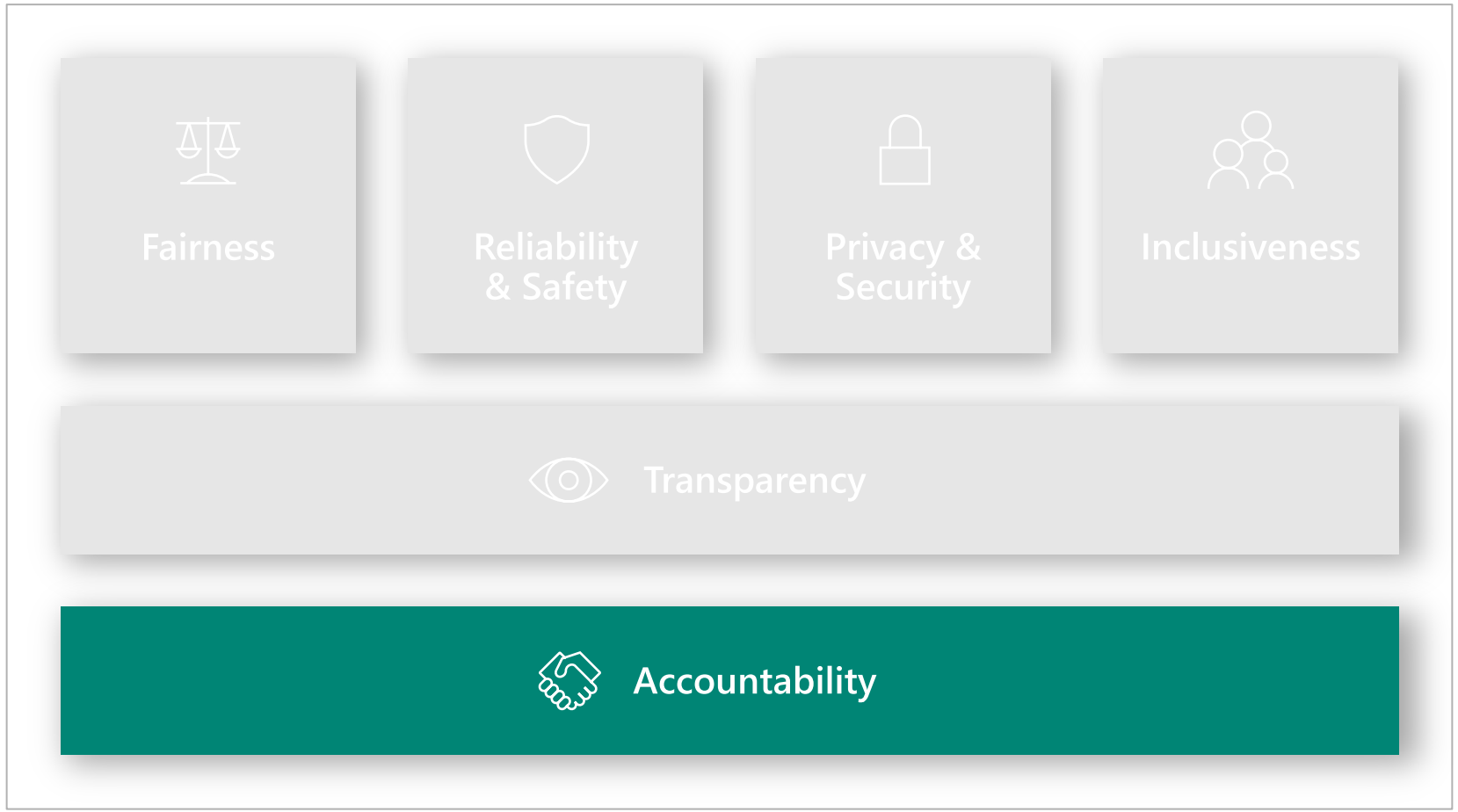
POLICY & LAW

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

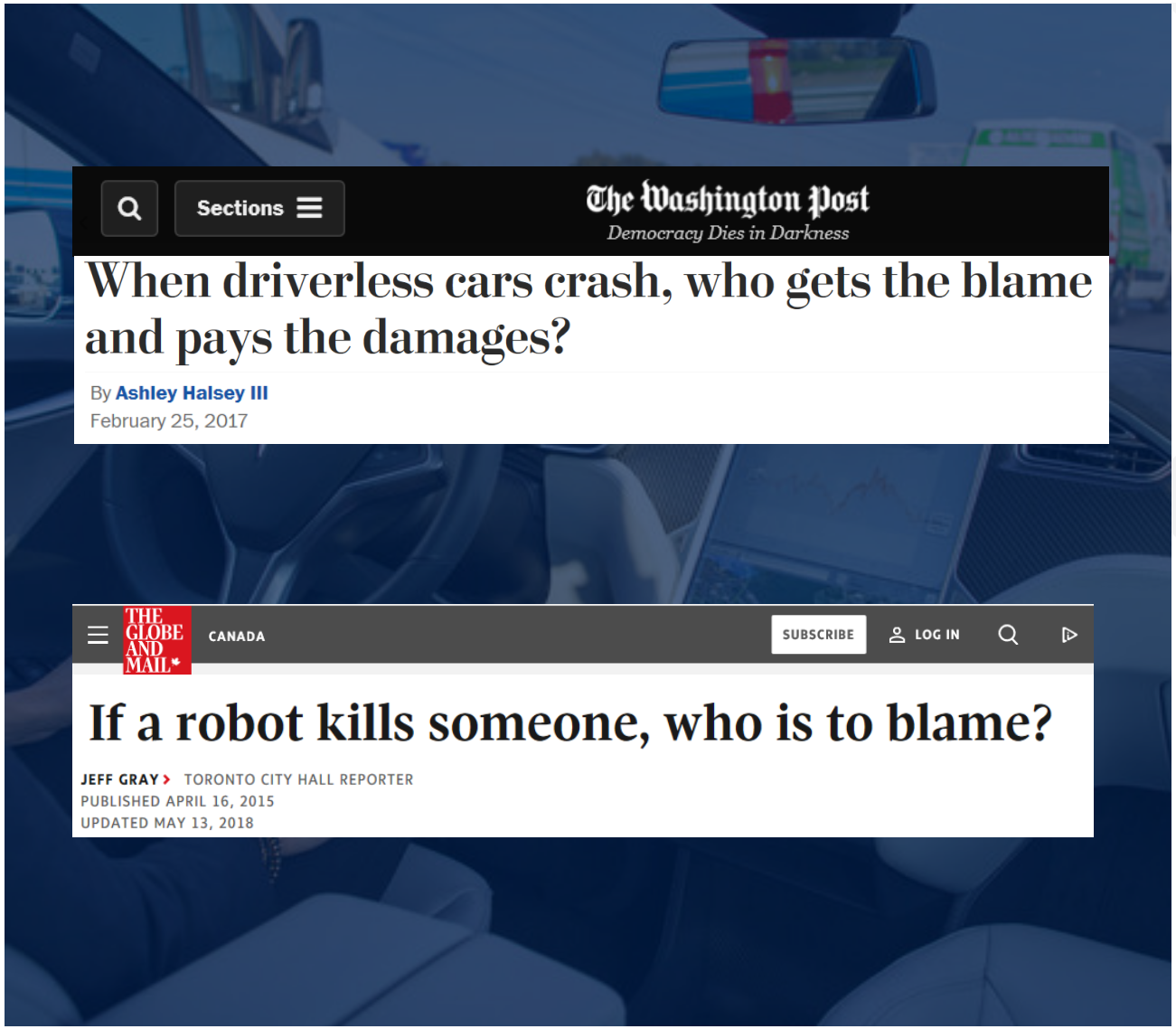
Palantir deployed a predictive policing system in New Orleans that even city council members don't know about

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

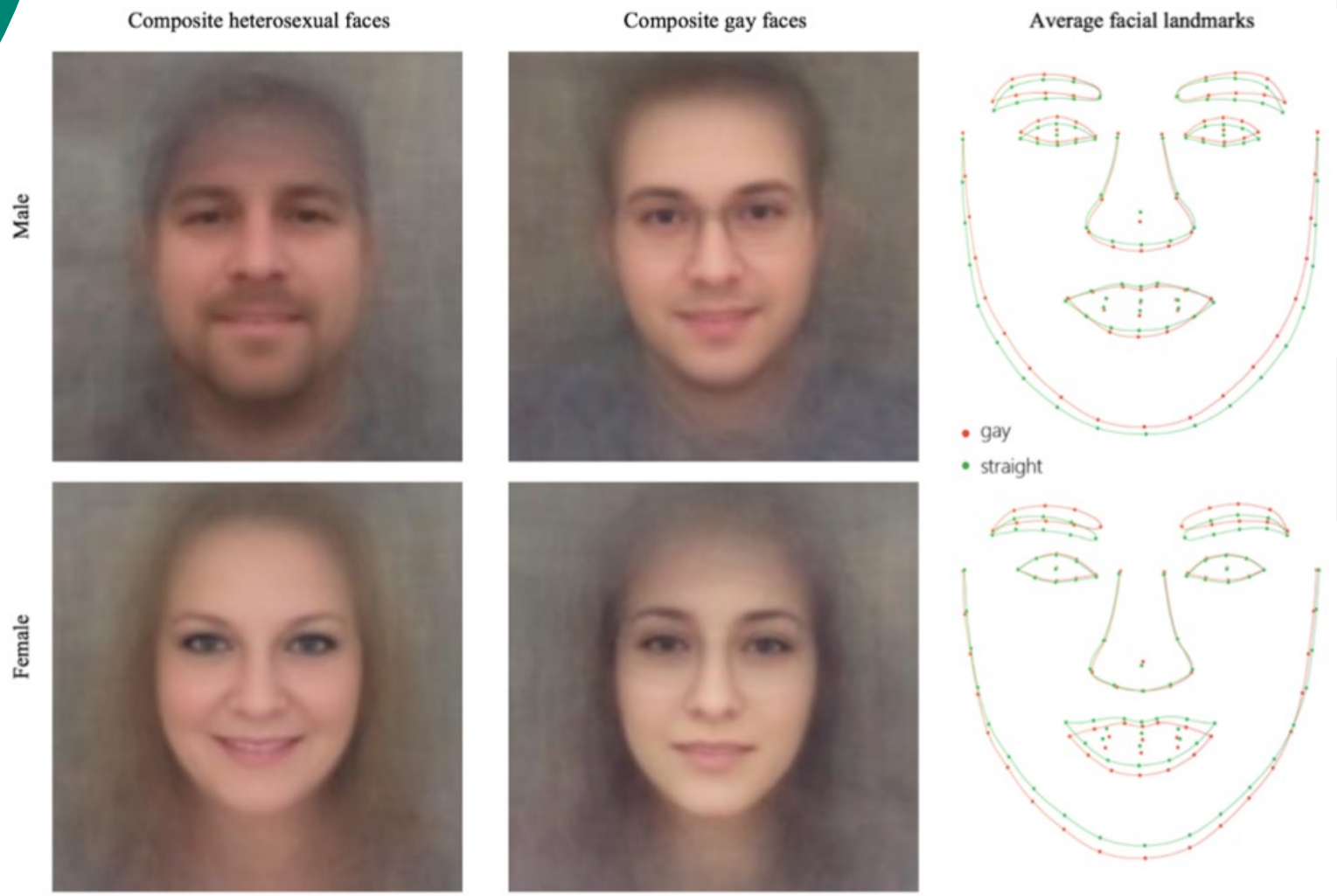
Microsoft's AI principles



Accountability



Accountability



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

Accountability

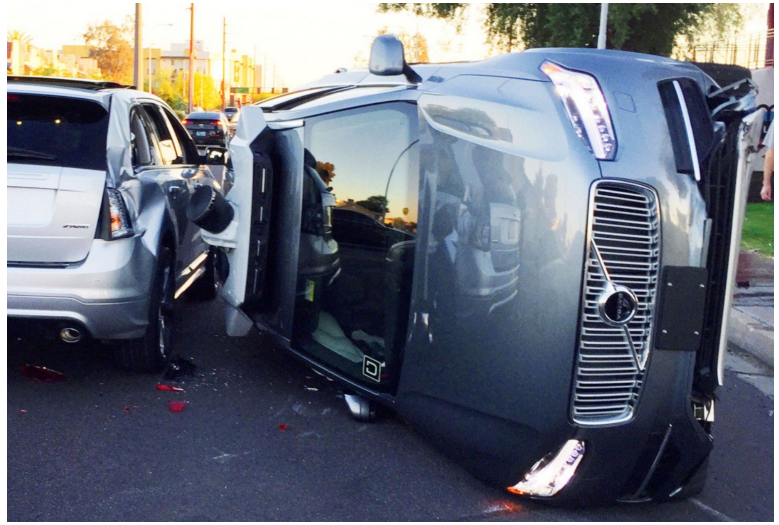
MIT Technology Review

Ethical Tech / AI Ethics

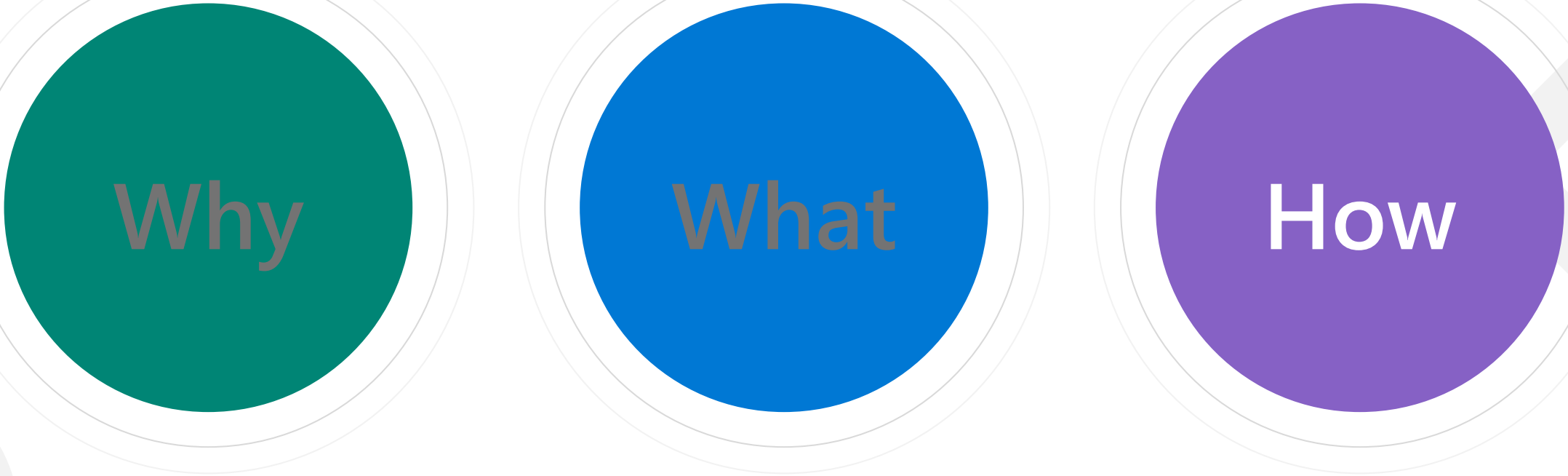
When algorithms mess up, the nearest human gets the blame

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

by **Karen Hao** May 28, 2019



Agenda



Putting Responsible AI into Practice

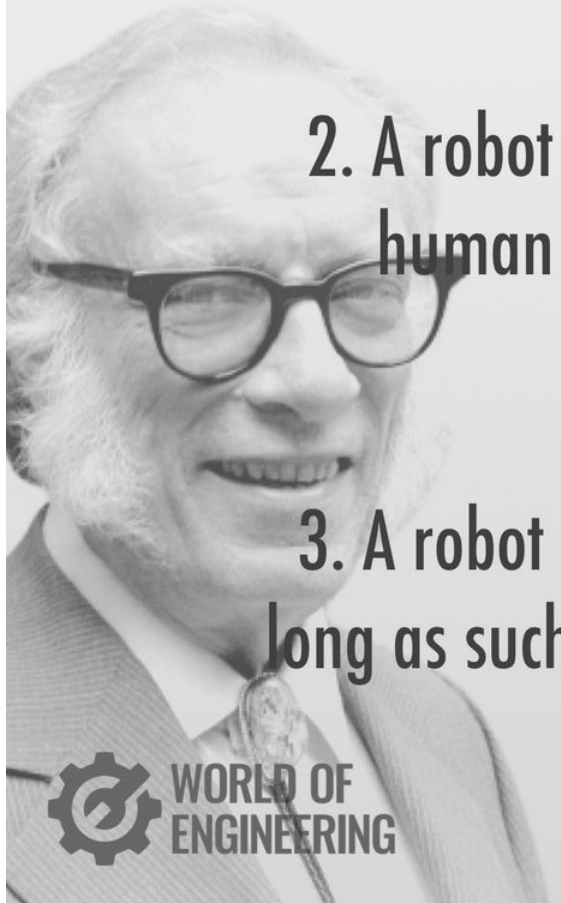
3 Laws of Robotics

Isaac Asimov's "Three Laws of Robotics"

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.

2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.

3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.



WORLD OF ENGINEERING

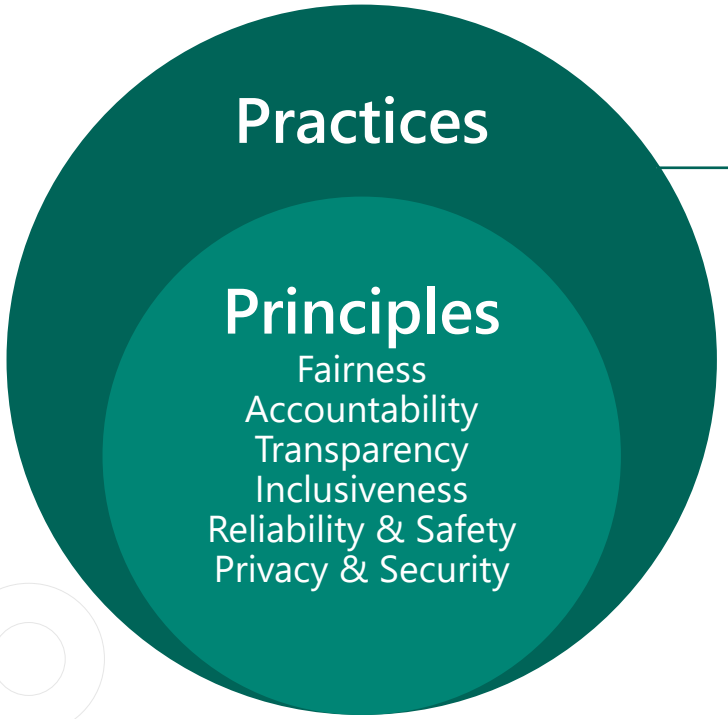
Case Management: Sensitive Uses

Responsible AI & Human Rights

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

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

Understand

Protect

Control

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.