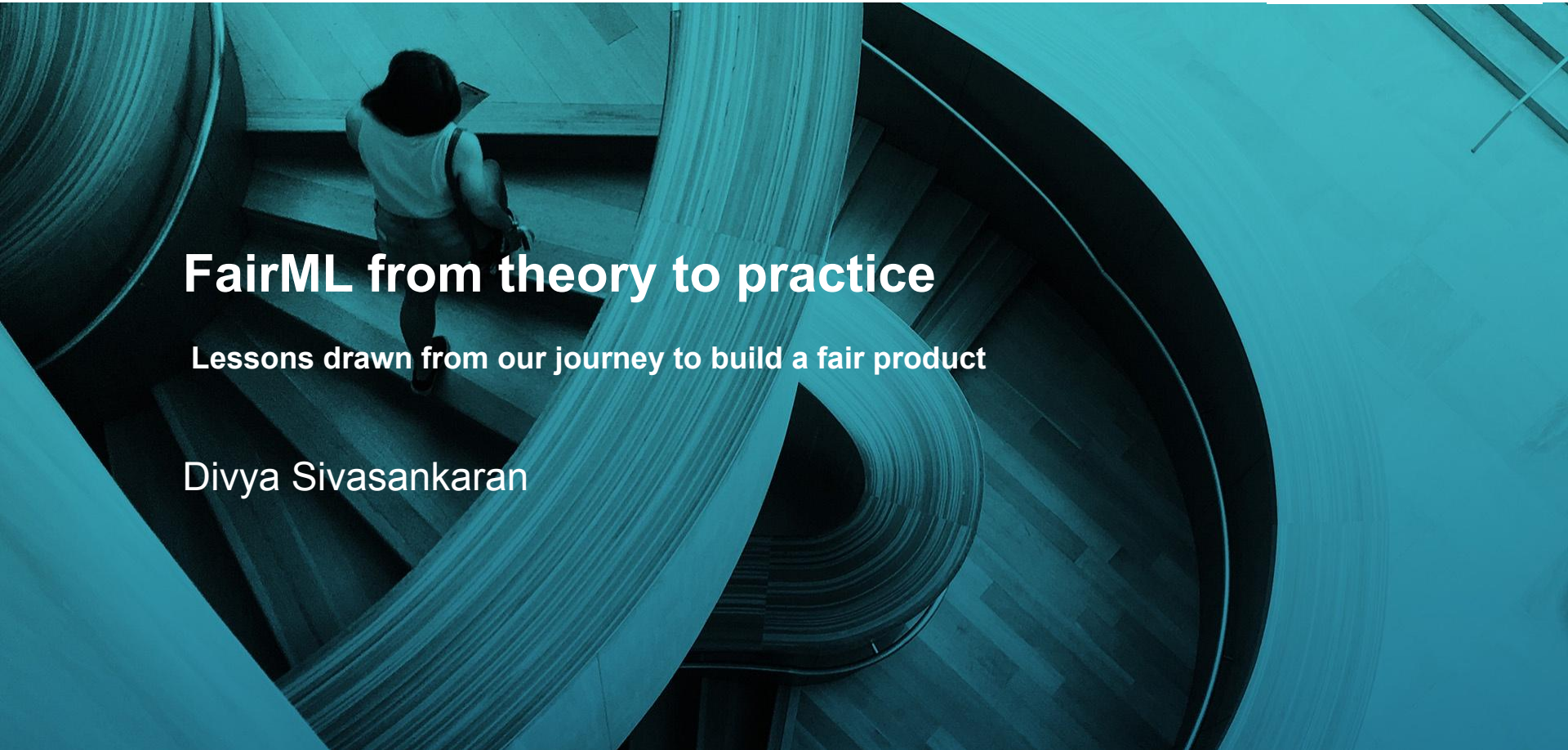
integrate.ai

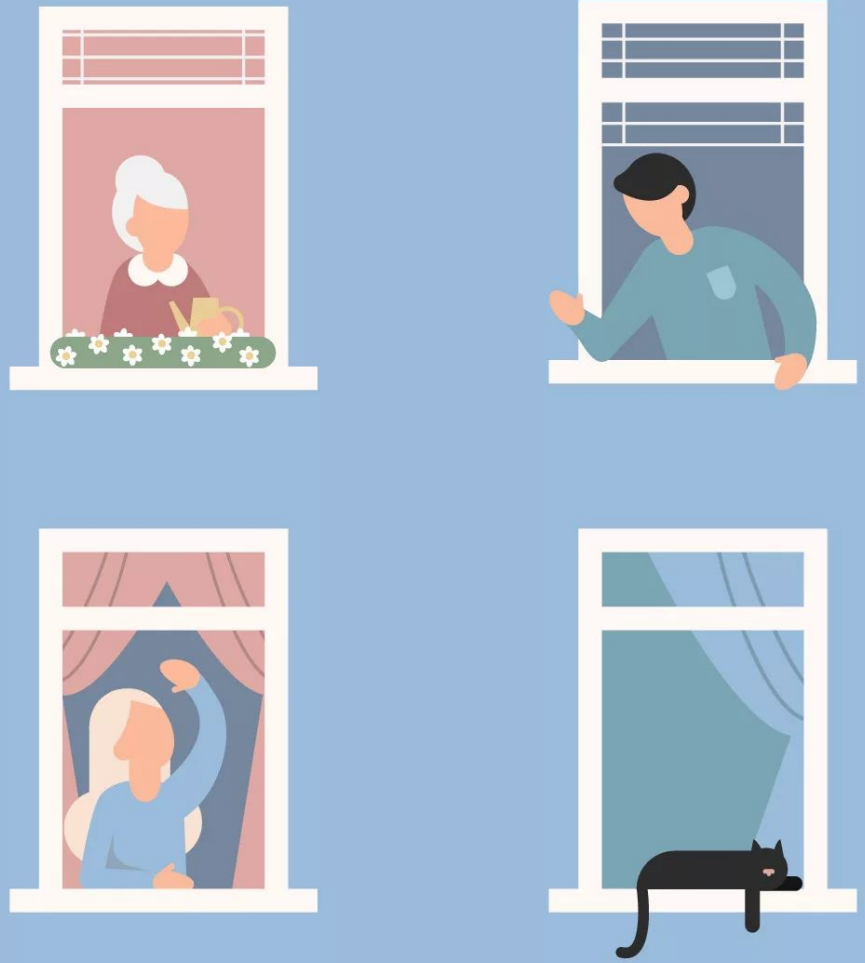
TORONTO WOMXN IN DATA SCIENCE



FairML from theory to practice

Lessons drawn from our journey to build a fair product

Divya Sivasankaran

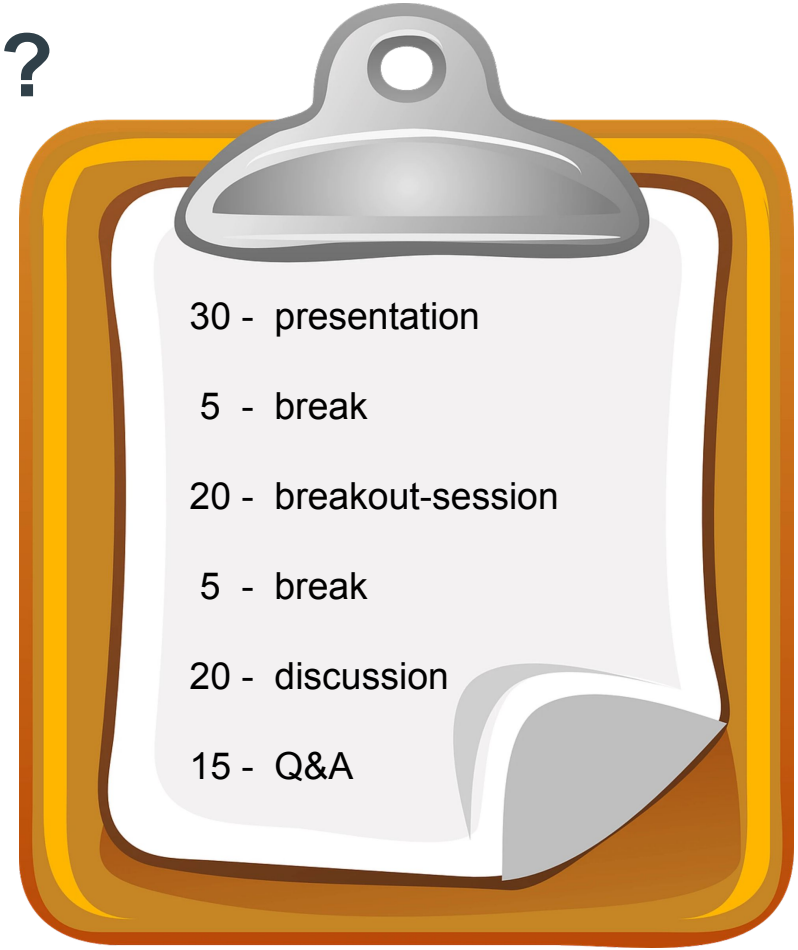


Hello there!

Introduce yourself to your neighbour...

Tell them what you most want to get out of today!

What’s on?

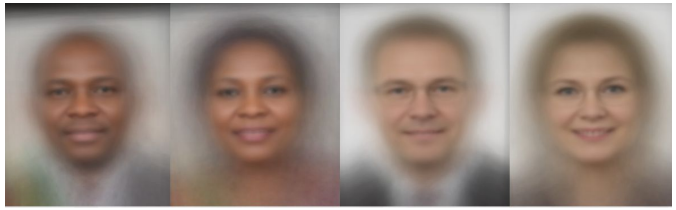

30 - presentation
5 - break
20 - breakout-session
5 - break
20 - discussion
15 - Q&A

integrate.ai

AI as a service to better help businesses understand their customers without violating their privacy and trust.

AI doesn’t work the same for everyone!

Gender Shades



Gender Classifier	Darker Male	Darker Female	Lighter Male	Lighter Female	Largest Gap
Microsoft	94.0%	79.2%	100%	98.3%	20.8%
FACE++	99.3%	65.5%	99.2%	94.0%	33.8%
IBM	88.0%	65.3%	99.7%	92.9%	34.4%

Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency* (pp. 77-91).

Biased AI systems could have severe downstream consequences for historically marginalized communities

theglobeandmail.com

Toronto police chief orders officers to stop using Clearview AI software

Tom CardosoCrime and justice reporter

4-5 minutes

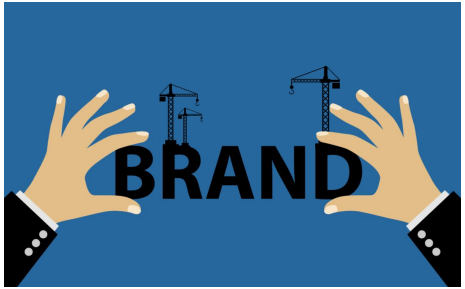
Hoan Ton-That, founder of Clearview AI, shows the results of a search for a photo of himself, in New York, Jan. 10, 2019.

AMR ALFIKY/NYTNS

Canada’s largest municipal police service has acknowledged "informally testing” a powerful surveillance tool, but said it has stopped its use pending reviews.

Why should you care?

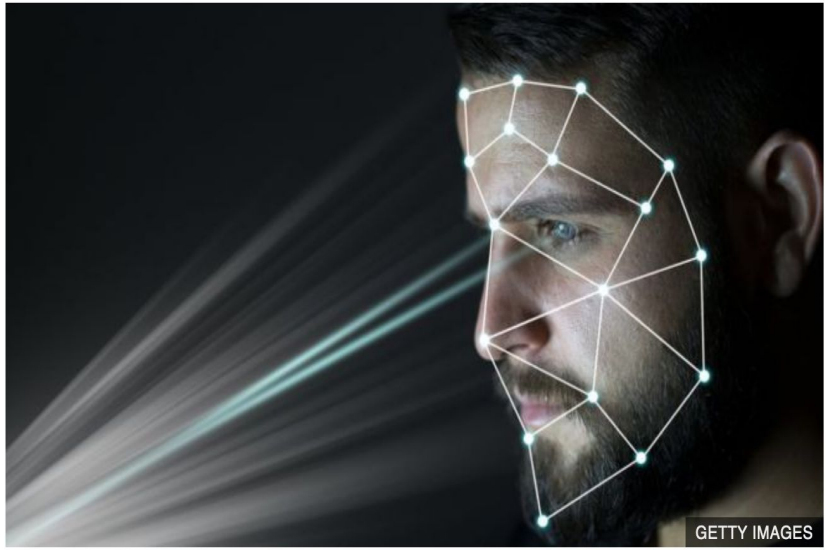
You use them every day!



What are governments doing?

Facial recognition: EU considers ban of up to five years

17 January 2020 Share



The European Commission has revealed it is considering a ban on the use of facial recognition in public areas for up to five years.

EU no longer considering facial recognition ban in public spaces

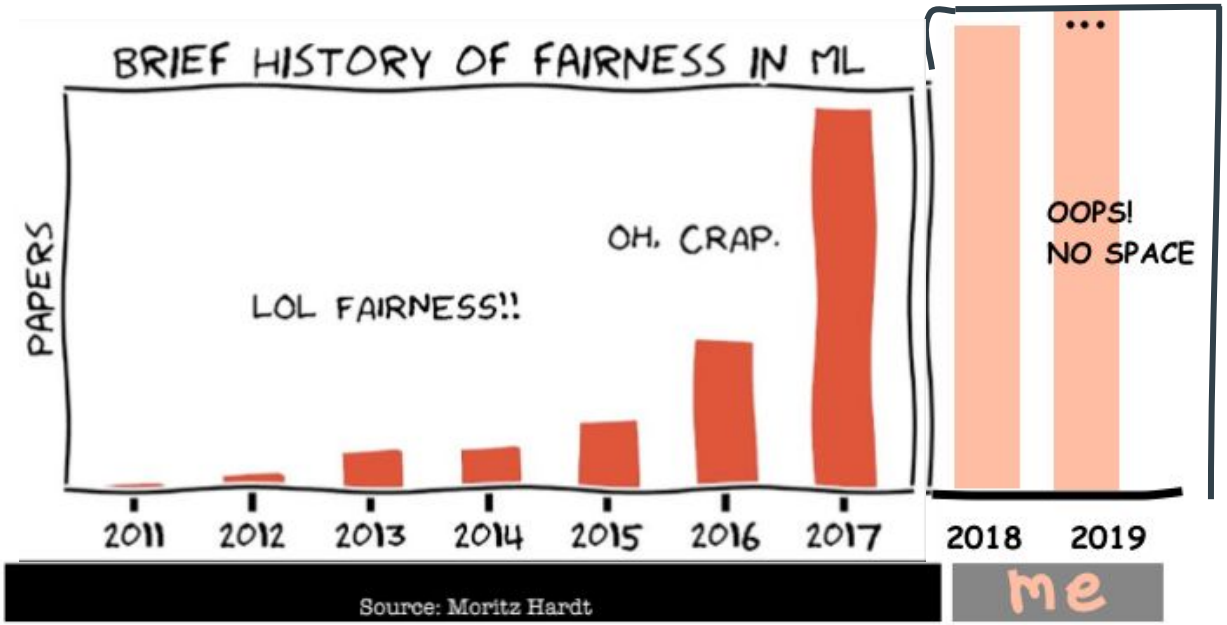
Jan 30, 2020 I Luana Pascu

2 minutes

According to the latest draft of the EU's Artificial Intelligence strategy seen by Reuters and EURACTIV, the European Union is no longer interested in introducing a ban on facial recognition in public spaces however there should be 'clear criteria' in future

What about academia?

Research on fairness and bias in ML



What about the industry?

open source™

THE ETHICS INSTITUTE

A lot of talk, but how exactly do we apply them?

no clear answers or standards

focus on impact

technologyreview.com

In 2020, let's stop AI ethics-washing and actually do something

Karen Hao

5-6 minutes

Last year, just as I was beginning to cover artificial intelligence, the AI world was getting a major wake-up call. There were some incredible advancements in AI research in 2018—from reinforcement learning to generative adversarial networks (GANs) to better natural-language understanding. But the year also saw several high-profile illustrations of the harm these systems can cause when they are deployed too hastily.

https://www.technologyreview.com/s/614992/ai-ethics-washing-time-to-act/?utm_medium=tr_social&utm_campaign=site_visitor.unpaid.engagement&utm_source=LinkedIn#Echobox=1582820020

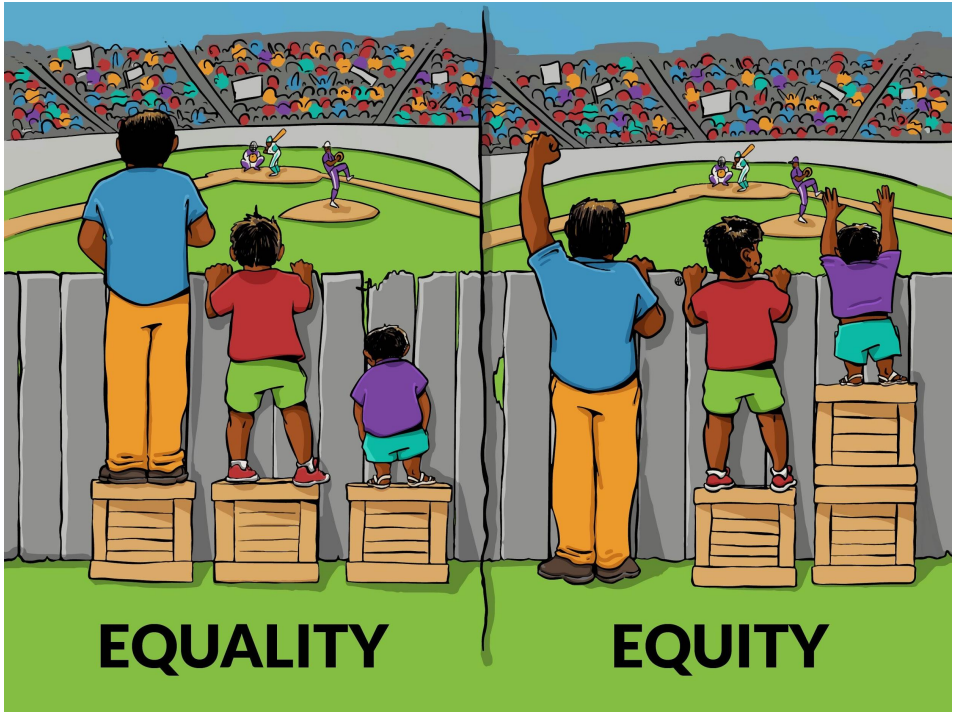
What is fairness?

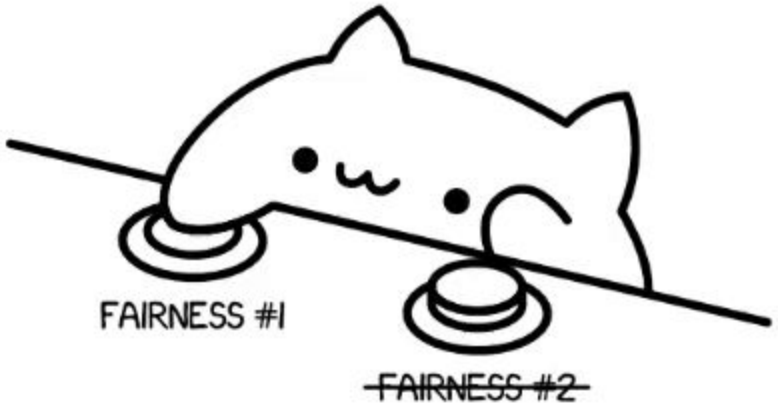
"21 definitions of algorithmic fairness"

Statistical parity
Group fairness
Demographic parity
Conditional statistical parity
Darlington criterion (4)
Equal opportunity
Equalized odds
Conditional procedure accuracy
Avoiding disparate mistreatment
Balance for the negative class
Balance for the positive class
Predictive equality
Equalized correlations
Darlington criterion (3)
Cleary model
Conditional use accuracy
Predictive parity
Calibration within groups
Darlington criterion (1), (2)

Narayanan, A. (2018, February). Translation tutorial: 21 fairness definitions and their politics. In *Proc. Conf. Fairness Accountability Transp., New York, USA*.

What is fairness?



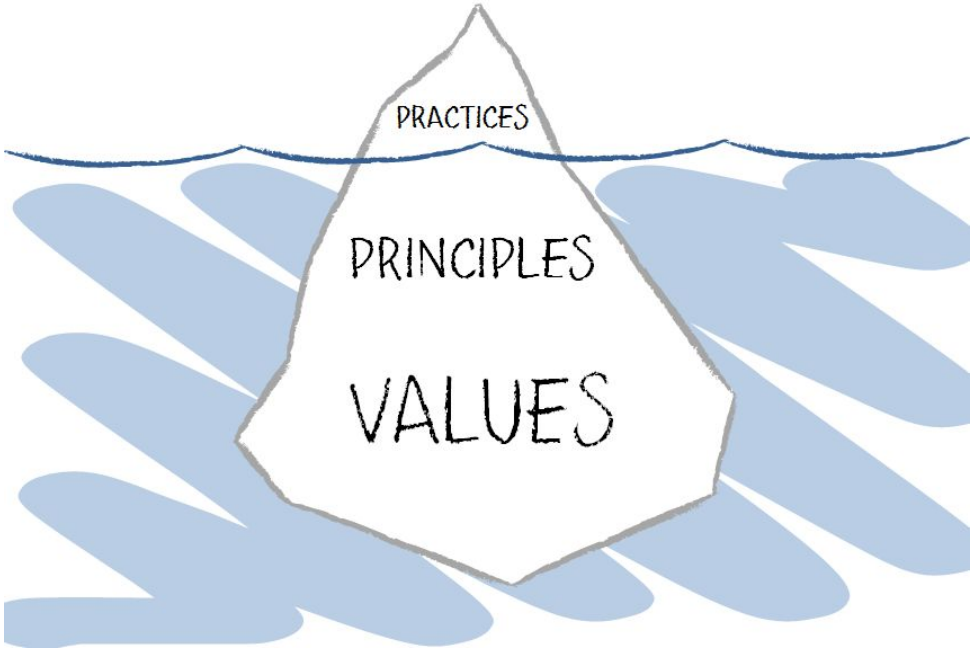


[1] https://machinesgonewrong.com/fairness/
[2] Friedler, S. A., Scheidegger, C., & Venkatasubramanian, S. (2016). On the (im) possibility of fairness. *arXiv preprint arXiv:1609.07236*.

Who gets to decide?

No free lunch y'all! Oh and diversity in your teams couldn't have mattered more!

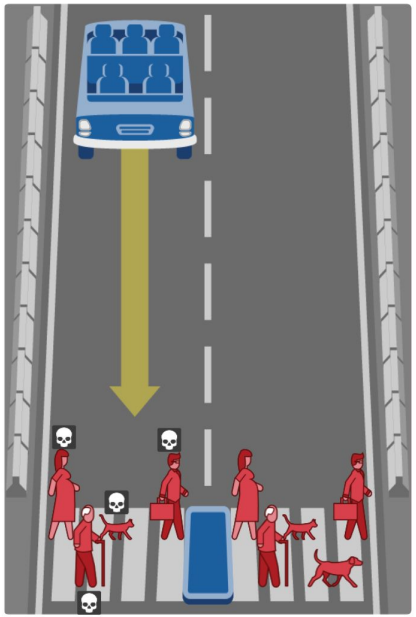
Value driven ethics?

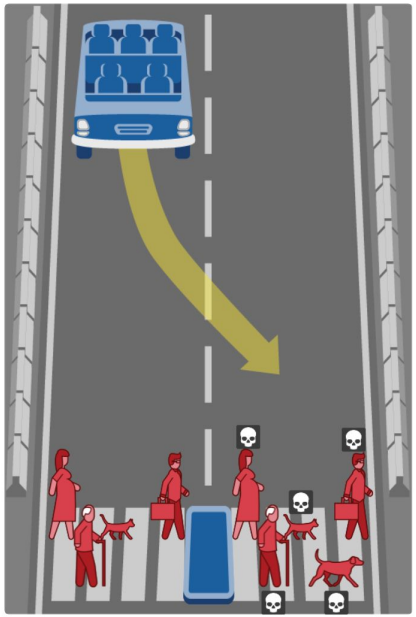


Moral machine

233 countries
and territories

What should the self-driving car do?

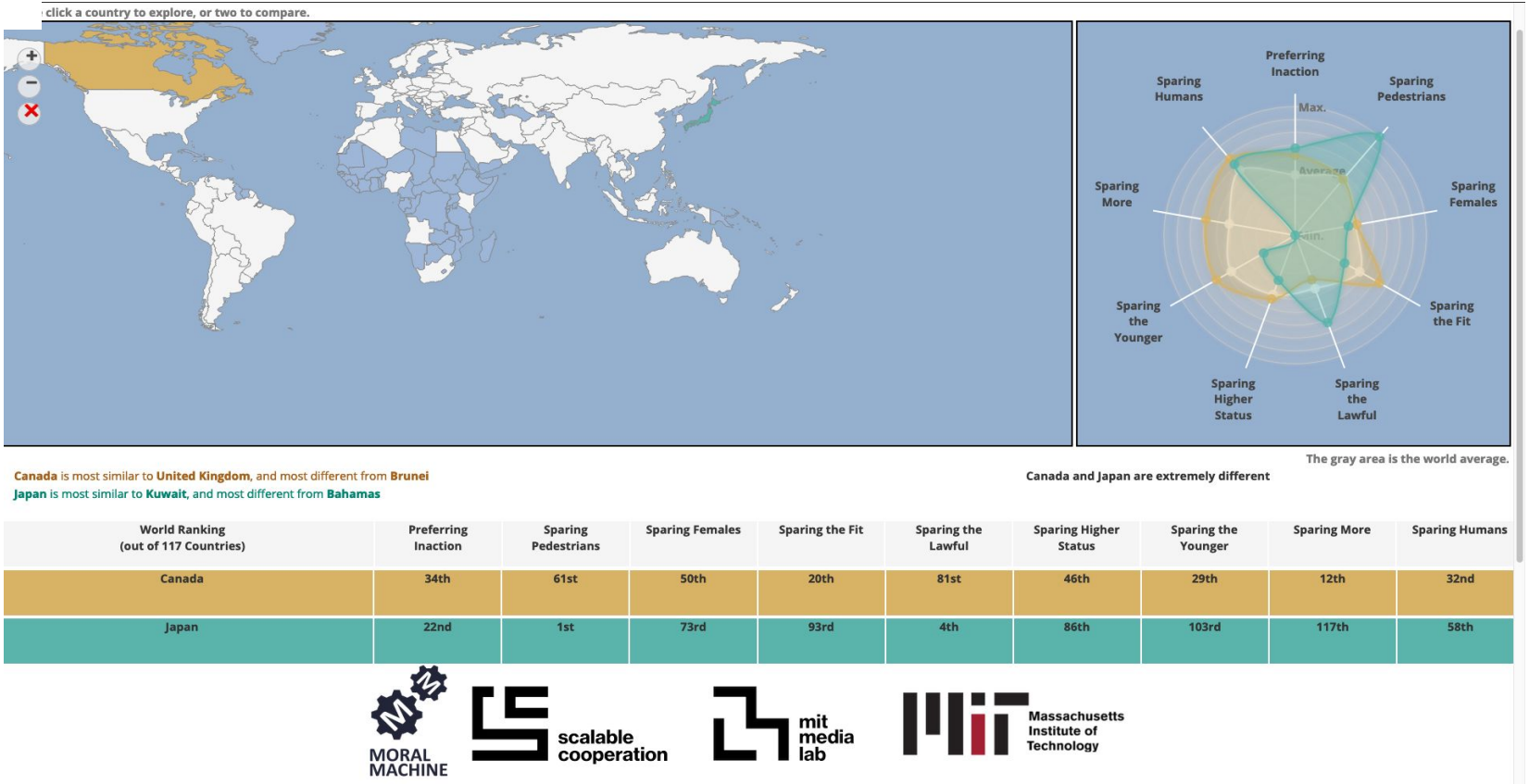




40 million
ethical decisions

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., ... & Rahwan, I. (2018). The moral machine experiment. *Nature*, *563*(7729), 59-64.

Values vary with culture and location



World Ranking (out of 117 Countries)	Preferring Inaction	Sparing Pedestrians	Sparing Females	Sparing the Fit	Sparing the Lawful	Sparing Higher Status	Sparing the Younger	Sparing More	Sparing Humans
Canada	34th	61st	50th	20th	81st	46th	29th	12th	32nd
Japan	22nd	1st	73rd	93rd	4th	86th	103rd	117th	58th

MORAL MACHINE

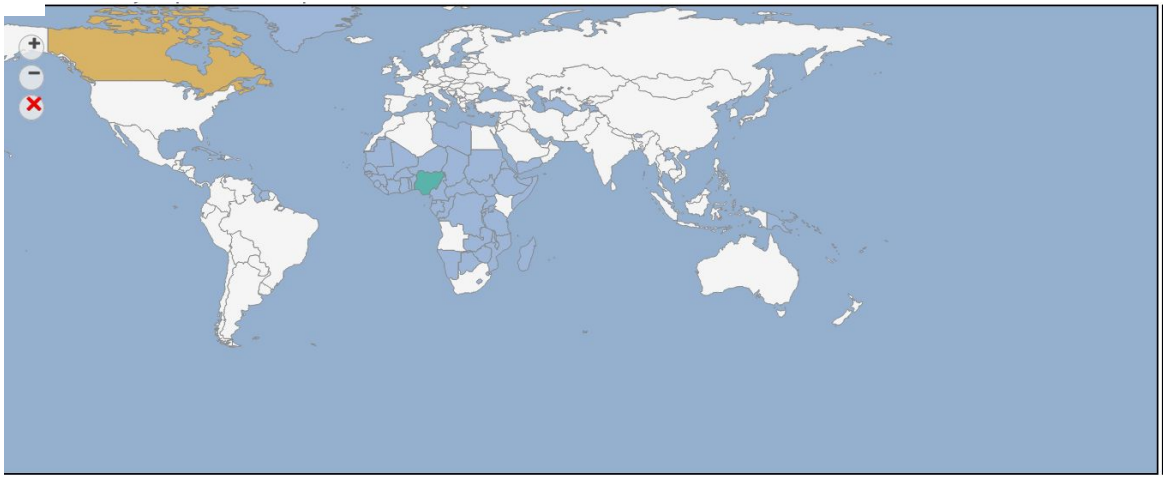
scalable cooperation

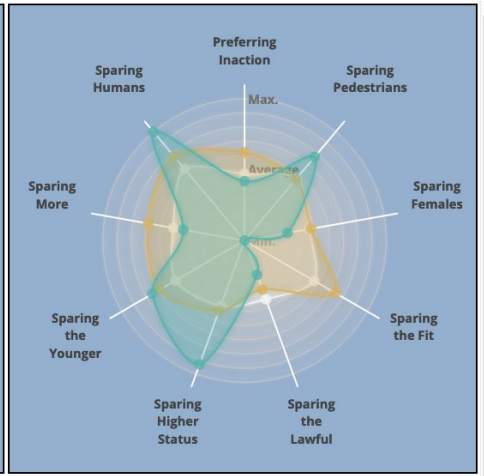
mit media lab

Massachusetts Institute of Technology

http://moralmachineresults.scalablecoop.org/

Values vary with culture and location





The gray area is the world average.

Canada is most similar to **United Kingdom**, and most different from **Brunei**
Nigeria is most similar to **Kenya**, and most different from **Brunei**

Canada and Nigeria are extremely different

World Ranking (out of 117 Countries)	Preferring Inaction	Sparing Pedestrians	Sparing Females	Sparing the Fit	Sparing the Lawful	Sparing Higher Status	Sparing the Younger	Sparing More	Sparing Humans
Canada	34th	61st	50th	20th	81st	46th	29th	12th	32nd
Nigeria	62nd	12th	103rd	117th	102nd	3rd	19th	82nd	1st

MORAL MACHINE

scalable cooperation

mit media lab

Massachusetts Institute of Technology

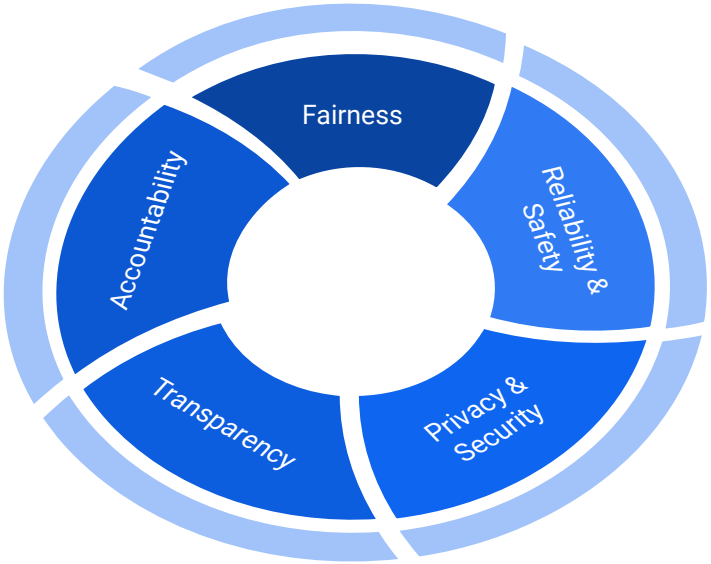
http://moralmachineresults.scalablecoop.org/

How does this apply to your work?

Principles

You're not trying to solve ethics for everyone!

- Identify what matters most to your organization
- Consider time-boxing it
- Don't reinvent the wheel



https://integrate.ai/blog/integrates-ai-principles

So, we've got 'em values

we got 'em code

our models are now fair!

Are we done?

Problems with industry adoption

Data maturity

- most industries are still grappling with onboarding AI
- ethics is not yet a concern or priority

Industry

- highly regulated industries care more

Geography

- GDPR, CCPA

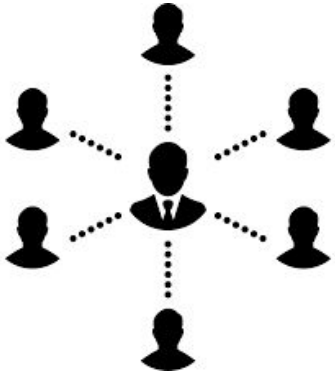
Brandon Purcell, Srividya Sridharan, Fatemeh Khatibloo, TJ Keitt, Jennifer Wise, Christopher McClean, and Christian Austin. The ethics of AI: How to avoid harmful bias and discrimination. Technical report, Forrester, February 2018.

Fairness is not a zero-sum game!

vs

Who are the stakeholders?

Decision maker



Key influencer

Regulator

We can do better!

Government

- Responsibility (deontology)
- What are the driving principles?

Designer

- Who are the stakeholders?
- Who’s left out? What’s the impact?

Consumer

- What do you do when the systems aren’t built to work for you?
- Now imagine if this was a decision being made by an AI?
- Do you get to contest a decision that is made for you by an algorithm?

Breakout

Discussion

5 minutes per group

- Describe the scenario chosen
- Were you able to reach a consensus on anything?
- What did you find most challenging?
- Do you have a learning/observation to share?

Fairness vs privacy

We want want both.. but is it always possible to have both?

Can you think of scenarios when you wouldn’t?

Why/why not?

If there’s one thing you would add/remove from the things we’ve discussed today, what would they be?

Recap

- ➔ It’s only a principle if it costs you something! Identify what matters to you
- ➔ Work together with relevant stakeholders to find a common definition of fairness
- ➔ Design systems with ethical principles in mind from the start
- ➔ Fairness is not a zero-sum game
- ➔ Work doesn’t end once you deploy! Fairness requires continuous effort
- ➔ Build diverse teams. Consider unconscious-bias training/workshops for your team

That’s all folks! Thank you!

divya@integrate.ai
LinkedIn

Tentative structure

- 5 min of setup/situate
- 5 min first to say what session looks like (breakout groups, lecture timing) and key part: introduce yourself to someone near you and introduce yourself and say what you most want to get out of today
- 25 min (???) present all 6 themes, ask what they think is most important /most need to understand (your example hits all/most but focuses on what the audience wants)
    - 5 minute - intro to the problem area, definitions etc
    - 5 minutes - present all key findings + vote
    - 10 minutes - deep dive into one topic (prepare for at least ¾ full topics) OR free-flow discussion with the use-case/example associated
    - 5 minutes - Pro-tips, some more thoughts etc along with Conclusion (or here.. Depending on time)
- 20 min break out groups to look at different scenarios
- 20 min report back (small group leaders report to full workshop)
- 15 min active group participation, gather “new themes”, eliminate one of your 6 themes?
- 10 min conclusion?

2:44

Alternatively - Try 3 topics / breakout / next 3 topics / breakout

Key learnings

What is fairness? -- 21 definitions, who decides what is right? you- developer? consumer? business owner? Impossibility results

okay define your values -- and then you can act… (usually values driven)

Bringing principles/values → action --- how do you do that? it changes with location, case, domain, … political views of the person implementing it?!
(Hint: Diversity in teams is ever more important) Tip: Don't do it youreslf. 5 most common themes + links you can just pick off from

Okay so we said for ourselves → transparency, privacy and this matters most at this stage. Now let's get tlaking to our clients.. to see how they feel and come up with some user-stuff!

stakeholder → revenue → how do you make fairness a non-zero sum game?

Fairness vs privacy -- Open question