The AI/Data Science Professional

Responsible AI/DS

Acknowledgement of Country

The AI/Data Science Professional

Course Coordinator: Flora Salim

What's next...

RMIT
UNIVERSITY

RMIT University acknowledges the people of the Woi wurrung and Boon wurrung language groups of the eastern Kulin Nation on whose unceded lands we conduct the business of the University.

RMIT University respectfully acknowledges their Ancestors and Elders, past and present. RMIT also acknowledges the Traditional Custodians and their Ancestors of the lands and waters across Australia where we conduct our business.

Ngarara Place

The AI/Data Science Professional – Week 2: Responsible AI/DS

Responsible AI/Data Science

Objectives

- Passion about AI / Data Science :)
- Ethical Mindset
    - ✖ Critical thinking
    - ✖ Research
    - ✖ Evaluation
- Code of Practice
- Technologies
- Professional practice
    - ✖ Communication
    - ✖ Team work

An AI/DS professional

Course Learning Outcomes

- CLO1 - Standards, ethical and social considerations
- CLO2 - Analyse and discuss impact
- CLO3 - Case studies
- CLO4 - Communication
- CLO5 - Research Principles

Refer to Course Guide for the actual description of the learning outcomes.

Responsible AI/DS

- Cost and Risk
    - �֎ Privacy
    - ✖ Fairness
    - ✖ Transparency and Explainability
    - ✖ Accountability, Responsibility and Governance
    - ✖ Human Right and Social Wellbeing
- Existing AI ethics and governance framework
- What to do this week
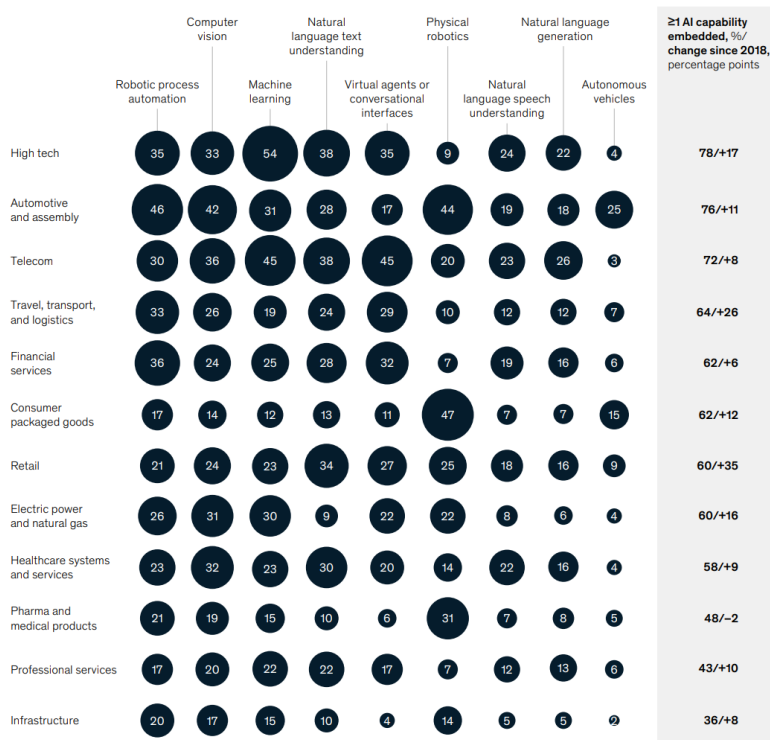- Guest talk on Research Integrity and Ethics

AI adoption in industry

AI global survey by McKinsey

https://www.mckinsey.com/featured-insights/artificial-intelligence/global-ai-survey-ai-proves-its-worth-but-few-scale-impact

High tech leads in AI adoption, and industries are generally using the AI capabilities most relevant to their value chains.

Organizations' AI capabilities,[1] % of respondents,[2] by industry

	Robotic process automation	Computer vision	Machine learning	Natural language text understanding	Virtual agents or conversational interfaces	Physical robotics	Natural language speech understanding	Natural language generation	Autonomous vehicles	≥1 AI capability embedded, %/ change since 2018, percentage points
High tech	35	33	54	38	35	9	24	22	4	78/+17
Automotive and assembly	46	42	31	28	17	44	19	18	25	76/+11
Telecom	30	36	45	38	45	20	23	26	3	72/+8
Travel, transport, and logistics	33	26	19	24	29	10	12	12	7	64/+26
Financial services	36	24	25	28	32	7	19	16	6	62/+6
Consumer packaged goods	17	14	12	13	11	47	7	7	15	62/+12
Retail	21	24	23	34	27	25	18	16	9	60/+35
Electric power and natural gas	26	31	30	9	22	22	8	6	4	60/+16
Healthcare systems and services	23	32	23	30	20	14	22	16	4	58/+9
Pharma and medical products	21	19	15	10	6	31	7	8	5	48/−2
Professional services	17	20	22	22	17	7	12	13	6	43/+10
Infrastructure	20	17	15	10	4	14	5	5	2	36/+8

[1] Embedded in ≥1 product and/or business process for ≥1 function or business unit.
[2] Respondents who said "don't know" or "none of the above" are not shown. For high tech, n = 277; for automotive and assembly, n = 128; for telecom, n = 93; for travel, transport, and logistics, n = 83; for financial services, n = 396; for consumer packaged goods, n = 72; for retail, n = 94; for electric power and natural gas, n = 82; for healthcare systems and services, n = 78; for pharma and medical products, n = 96; for professional services, n = 331; and for infrastructure, n = 91.

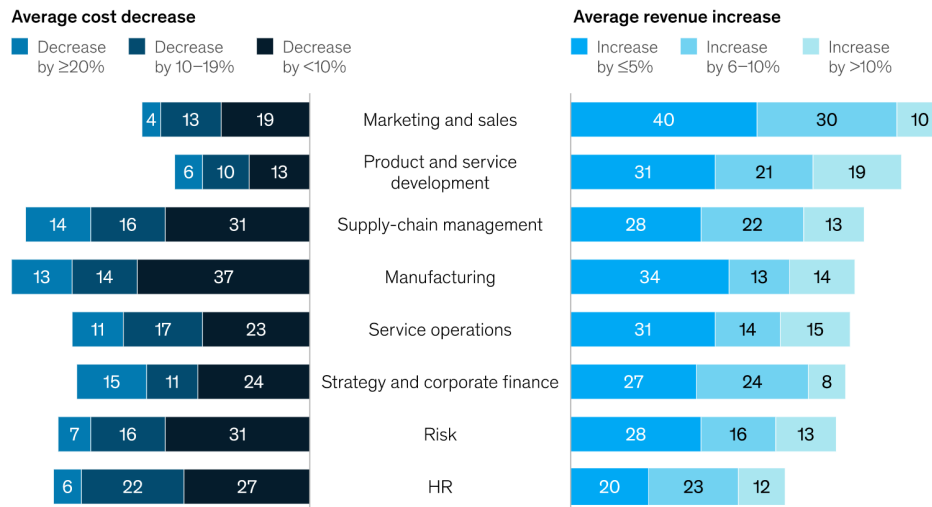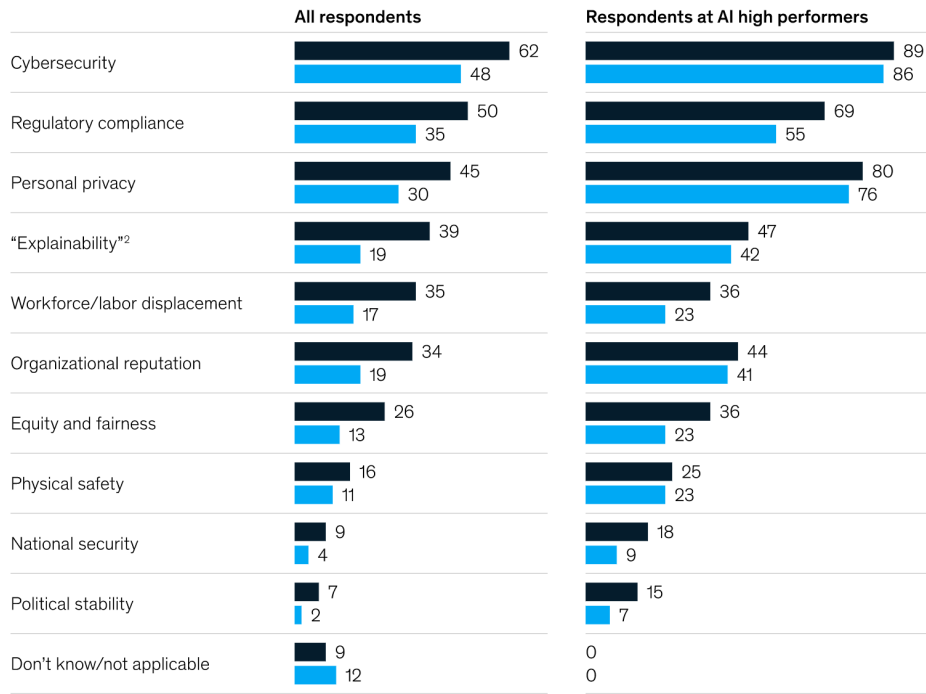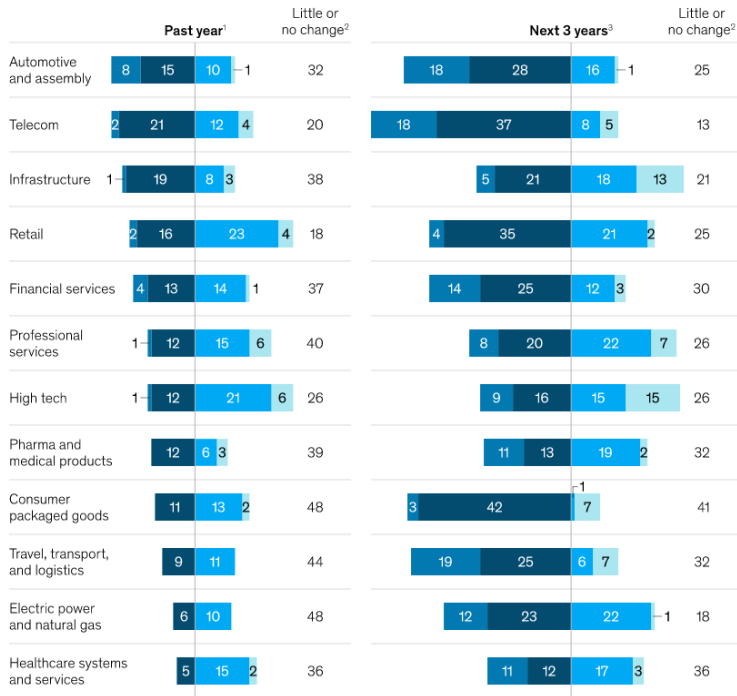# AI adoption leads to increased revenue and decreased cost

AI global survey by McKinsey

https://www.mckinsey.com/featured-insights/artificial-intelligence/global-ai-survey-ai-proves-its-worth-but-few-scale-impact

Revenue increases from adopting AI are reported most often in marketing and sales, and cost decreases most often in manufacturing.

Cost decrease and revenue increase from AI adoption, by function,[1] % of respondents[2]

	Average cost decrease				Average revenue increase		
	Decrease by ≥20%	Decrease by 10–19%	Decrease by <10%		Increase by ≤5%	Increase by 6–10%	Increase by >10%
Marketing and sales	4	13	19		40	30	10
Product and service development	6	10	13		31	21	19
Supply-chain management	14	16	31		28	22	13
Manufacturing	13	14	37		34	13	14
Service operations	11	17	23		31	14	15
Strategy and corporate finance	15	11	24		27	24	8
Risk	7	16	31		28	16	13
HR	6	22	27		20	23	12

[1]Marketing and sales includes the following use cases: customer-service analytics, customer segmentation, channel management, prediction of likelihood to buy, pricing and promotion, closed-loop marketing, marketing-budget allocation, churn reduction, and next product to buy. For product and service development: product-feature optimization, product-development-cycle optimization, creation of new AI-based enhancements, and creation of new AI-based products. For supply-chain management: logistics-network optimization, sales and parts forecasting, warehouse optimization, inventory and parts optimization, spend analytics, and sales and demand forecasting. For manufacturing: predictive maintenance and yield, energy, and throughput optimization. For service operations: service-operations optimization, contact-center automation, and predictive service and intervention. For strategy and corporate finance: capital allocation, treasury management, and M&A support. For risk: risk modeling/analytics, and fraud/debt analytics. For HR: performance management and organization-design, workforce-deployment, and talent-management optimization.
[2]Question asked only of respondents who said their companies adopted AI in given use case. Figures were calculated after removing respondents who said "don't know" or "not applicable; we are not tracking revenue related to AI"; respondents who said "no change" are not shown.

Human Rights and Social Wellbeing

Image sourced from The Intercept Source https://www.flickr.com/photos/itupictures/39621400230

[Discussion]
Are you concerned about the risk of AI making us jobless?

Image sourced from The Intercept Source from https://chatbotslife.com/will-ai-cause-mass-unemployment-f26537e5f25a

AI risks

AI global survey by McKinsey

https://www.mckinsey.com/featured-insights/artificial-intelligence/global-ai-survey-ai-proves-its-worth-but-few-scale-impact

Risks that organizations consider relevant and are working to mitigate,
% of respondents[1]

Relevant risk
Mitigated risk

	All respondents	Respondents at AI high performers
	Relevant / Mitigated	Relevant / Mitigated
Cybersecurity	62 / 48	89 / 86
Regulatory compliance	50 / 35	69 / 55
Personal privacy	45 / 30	80 / 76
"Explainability"[2]	39 / 19	47 / 42
Workforce/labor displacement	35 / 17	36 / 23
Organizational reputation	34 / 19	44 / 41
Equity and fairness	26 / 13	36 / 23
Physical safety	16 / 11	25 / 23
National security	9 / 4	18 / 9
Political stability	7 / 2	15 / 7
Don't know/not applicable	9 / 12	0 / 0

[1]Question asked only of respondents who said their companies had embedded or piloted ≥1 AI capability; n = 1,872.
[2]Ability to explain how AI models come to their decisions.

Job loss/changes due to AI

AI global survey by McKinsey

https://www.mckinsey.com/featured-insights/artificial-intelligence/global-ai-survey-ai-proves-its-worth-but-few-scale-impact
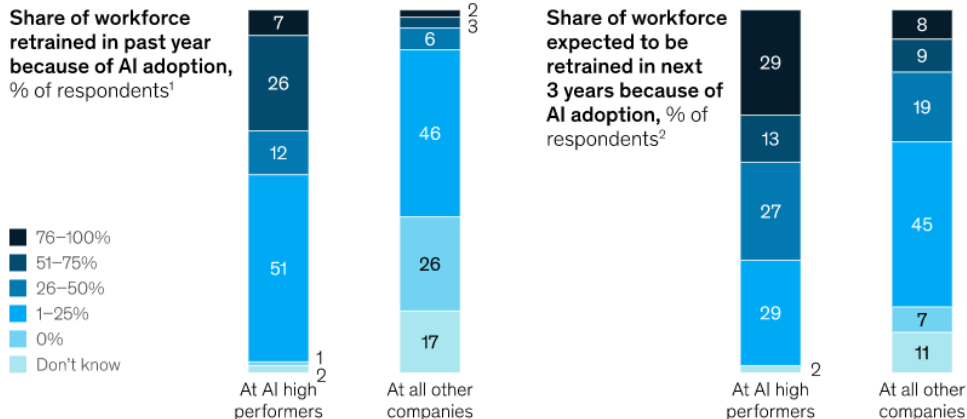
Workforce retraining due to AI

AI global survey by McKinsey

https://www.mckinsey.com/featured-insights/artificial-intelligence/global-ai-survey-ai-proves-its-worth-but-few-scale-impact

Respondents at high performers report larger retraining efforts as a result of AI than others do.

Share of workforce retrained in past year because of AI adoption, % of respondents[1]

Share of workforce expected to be retrained in next 3 years because of AI adoption, % of respondents[2]

Legend:
- 76–100%
- 51–75%
- 26–50%
- 1–25%
- 0%
- Don't know

At AI high performers / At all other companies

Note: Figures may not sum to 100%, because of rounding.
[1]Question was asked only of respondents who say their companies have piloted or embedded one or more AI capabilities. For respondents at high performers, n = 54; for all others, n = 1,818.
[2]Question was asked only of respondents who say their companies have piloted or embedded one or more AI capabilities, or plan to do so in the next three years. For respondents at high performers, n = 54; for all others, n = 1,892.

McKinsey & Company

A case study: virtual assistant

What are the benefits and the risks?

Google Duplex

https://youtu.be/D5VN56jQMWM

Jarvis (by Mark Zuckerberg)

https://youtu.be/vvimBPJ3XGQ

AI Ethics Principles

Human, social and environmental wellbeing	Human-centred values	Fairness	Privacy protection and security
Reliability and safety	Transparency and explainability	Contestability	Accountability

Source from https://www.industry.gov.au/data-and-publications/building-australias-artificial-intelligence-capability/ai-ethics-framework/ai-ethics-principles

Fairness

"AI has the potential to help humans make fairer decisions—but only if we carefully work toward fairness in AI systems as well." -- by Jake Silberg and James Manyika
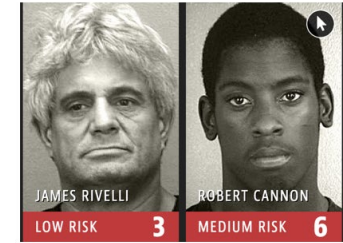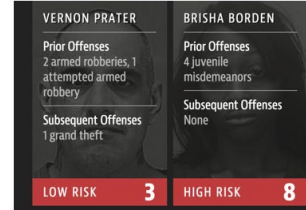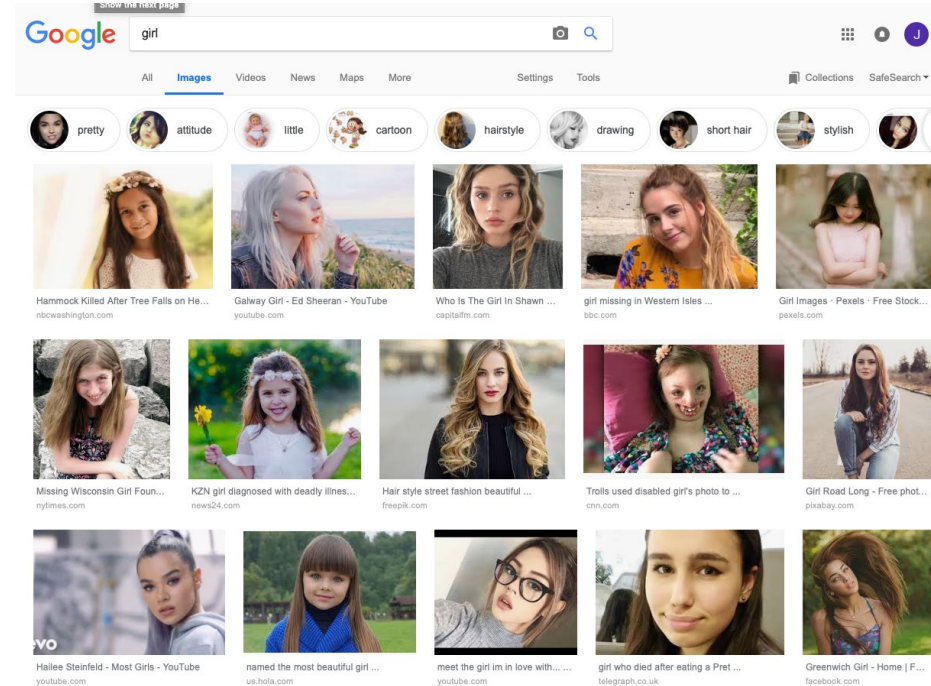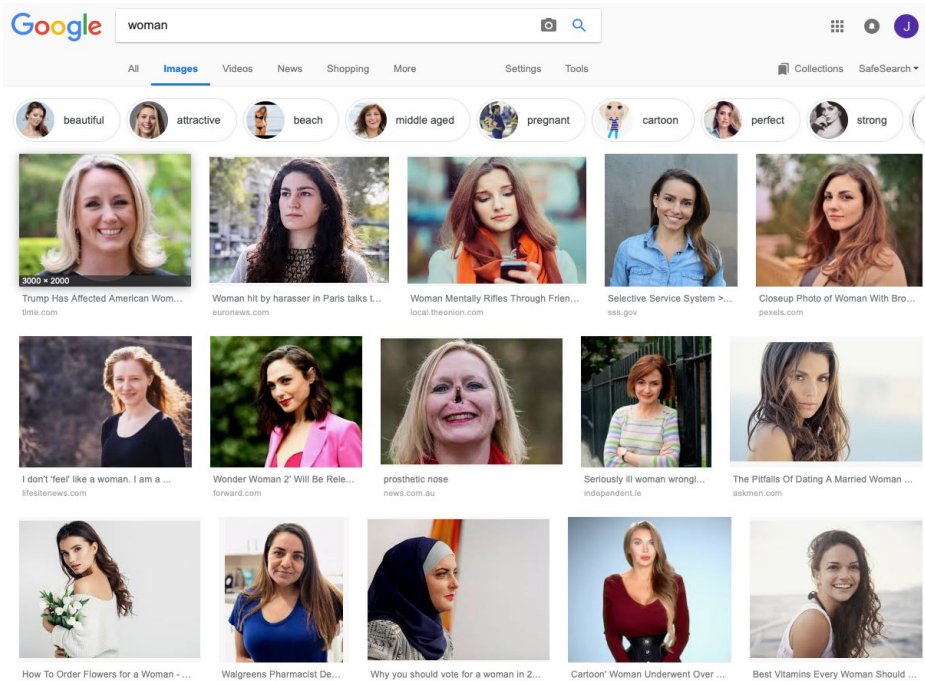
[Discussion]
Is there any AI/DS application that concern you with regards to Fairness?

Search Engine Bias

Source from https://scroll.in/article/921305/google-search-results-reflect-its-algorithms-bias-against-women-and-people-of-colour

Suggestive autocompletion
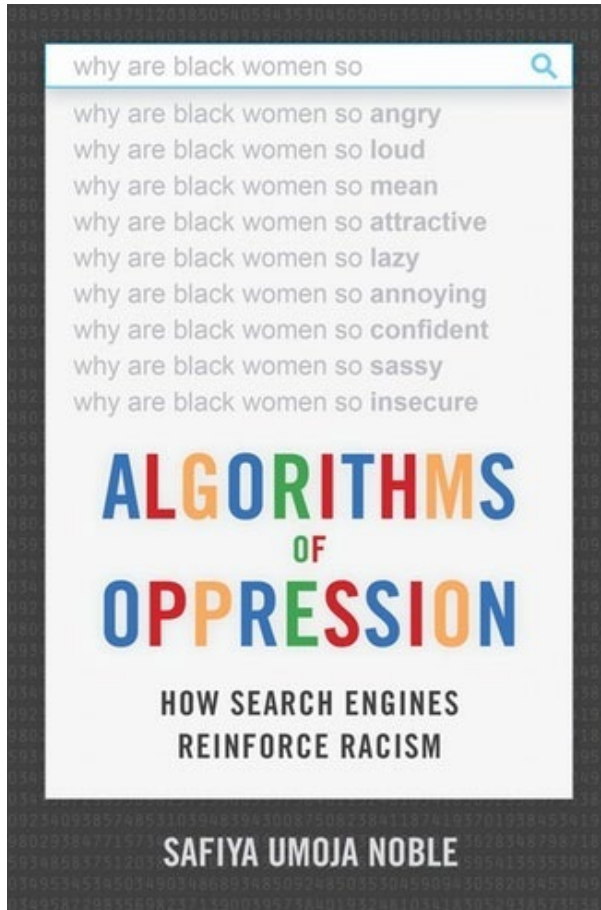
How biased are our algorithms? |
Safiya Umoja Noble | TEDxUIUC

Source from https://scroll.in/article/921305/google-search-results-reflect-its-algorithms-bias-against-women-and-people-of-colour

Other examples

- PredPol

- Facial recognition

- Google images search for "CEO"

- Facebook falsely feeds the intifada

PredPol
Predict Crime in **Real Time**™

PredPol provides targeted, real-time crime prediction designed for and successfully tested by officers in the field.

How are you?

Τι κάνετε;
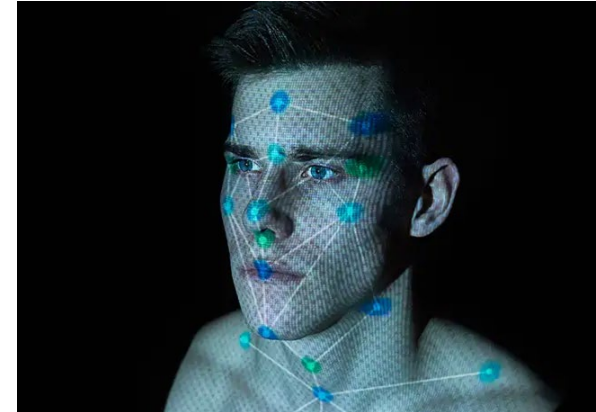
Transparency and Explainability

Image sourced from
https://www.nvidia.com/

[Discussion]
When do you think self-driving cars will widespread in Australia? And why?

[Discussion]
Are there any other AI applications that concern you in regards to Transparency and Explainability?

Privacy Concern

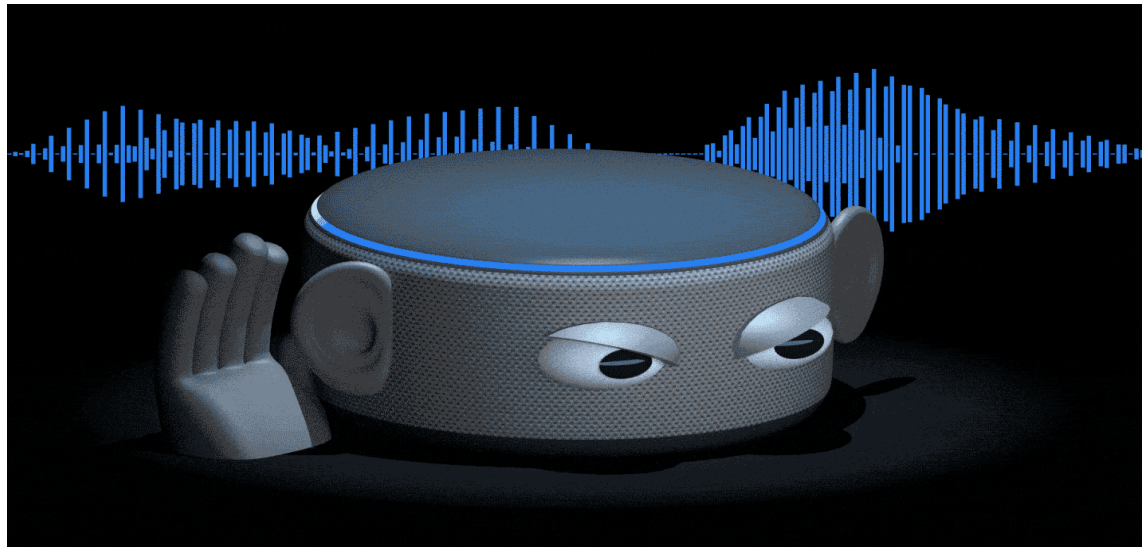Source from https://theintercept.imgix.net/wp-uploads/sites/1/2018/11/Intercept_Echo_v2-3.5MB-2-1542062294.gif

Case study: AMAZON'S ACCENT RECOGNITION TECHNOLOGY

AT THE BEGINNING of October 2018, Amazon was quietly issued a patent that would allow its virtual assistant Alexa to decipher a user's physical characteristics and emotional state based on their voice. Characteristics, or "voice features," like language accent, ethnic origin, emotion, gender, age, and background noise would be immediately extracted and tagged to the user's data file to help deliver more targeted advertising.

The algorithm would also consider a customer's physical location — based on their IP address, primary shipping address, and browser settings — to help determine their accent.
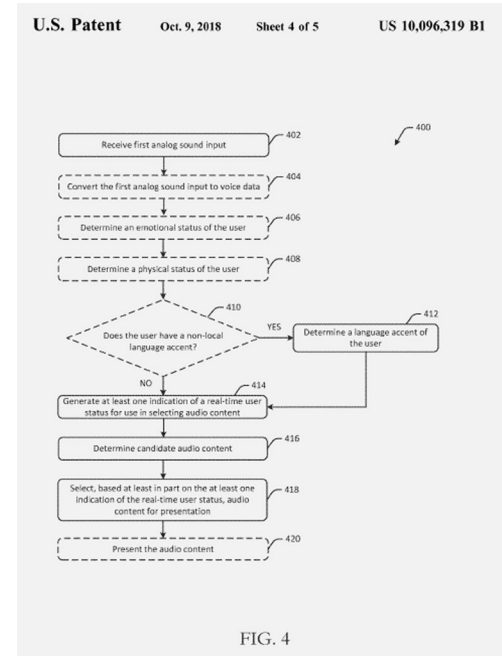
The illustration from Document: United States Patent and Trademark Office

[Discussion]
Is there any other AI/DS application that concern you with regards to Privacy?

Accountability, Responsibility and Governance
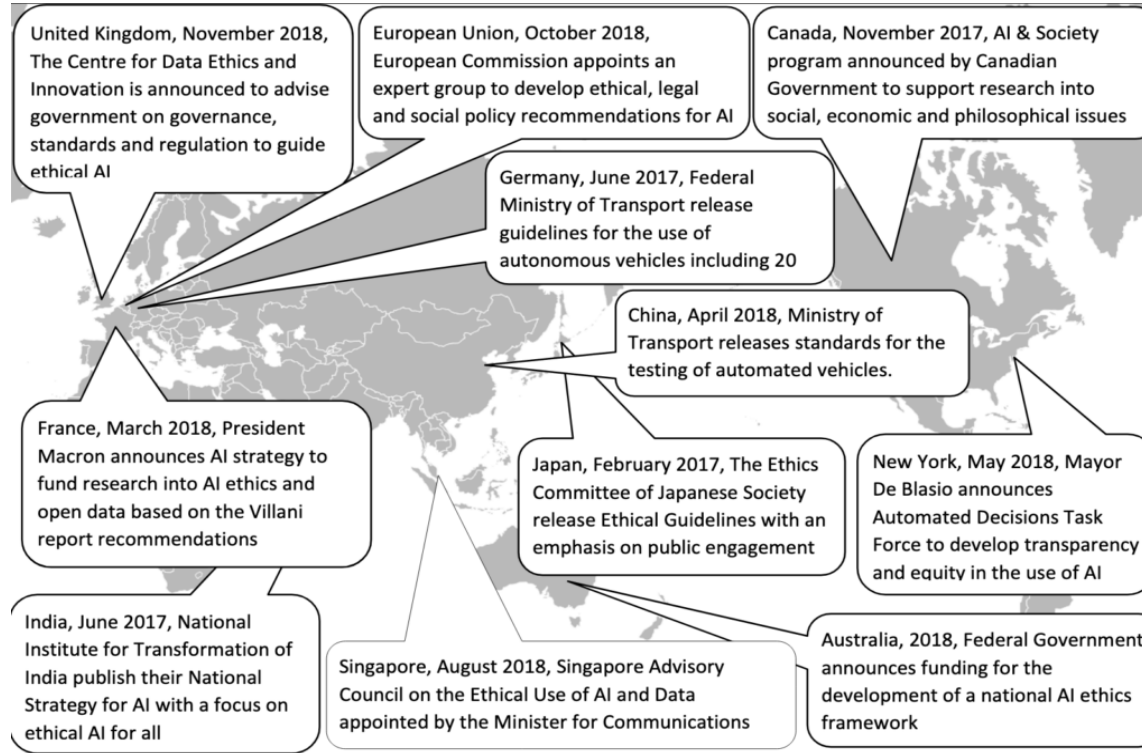

https://medium.com/@virginiadignum/the-art-of-ai-accountability-responsibility-transparency-48666ec92ea5

Existing AI ethics and governance framework

United Kingdom, November 2018, The Centre for Data Ethics and Innovation is announced to advise government on governance, standards and regulation to guide ethical AI

European Union, October 2018, European Commission appoints an expert group to develop ethical, legal and social policy recommendations for AI

Canada, November 2017, AI & Society program announced by Canadian Government to support research into social, economic and philosophical issues

Germany, June 2017, Federal Ministry of Transport release guidelines for the use of autonomous vehicles including 20

China, April 2018, Ministry of Transport releases standards for the testing of automated vehicles.

France, March 2018, President Macron announces AI strategy to fund research into AI ethics and open data based on the Villani report recommendations

Japan, February 2017, The Ethics Committee of Japanese Society release Ethical Guidelines with an emphasis on public engagement

New York, May 2018, Mayor De Blasio announces Automated Decisions Task Force to develop transparency and equity in the use of AI

India, June 2017, National Institute for Transformation of India publish their National Strategy for AI with a focus on ethical AI for all

Singapore, August 2018, Singapore Advisory Council on the Ethical Use of AI and Data appointed by the Minister for Communications

Australia, 2018, Federal Government announces funding for the development of a national AI ethics framework

Australian Framework

- Government and automated decisions

- Australia's international human rights obligations and anti-discrimination legislation

- Data-sharing legislation in Australia

- Privacy Act

https://consult.industry.gov.au/strategic-policy/artificial-intelligence-ethics-framework/supporting_documents/ArtificialIntelligenceethicsframeworkdiscussionpaper.pdf

Further discussion

- Week 3: working with data & users + *Guest talk* by Amazon

- Week 4: FATE in AI (continued)

- Week 5: Privacy & security

- Week 6: AI Bias and Fairness

- Week 7: Transparency and Explainability + *Guest talk* on Communicating DS Insights with Stakeholders by A/Prof Richard Xu, UTS

- Week 9: Human rights and wellbeing

- Week 10: AI/DS for Social Good (*Guest talk* by Prof Milind Tambe, Harvard University & Director, Google AI for Social Good)

- Week 11: Developing and Productising AI/DS project

Assignment Overview

What's next...

Recap: Assessment Structure

Task	Percentage	Due Date
Assessment Task 1 (Ethical AI/DS Case Study – Group) **Case study Report** **Case study presentation**	20%	Sundays 23:59 Week 7
Assessment Task 2 (AI/DS for Good Project – Group) **Milestone** 1 (M1 10%) – Initial Submission (Report) **Milestone** 2 (M2 40%) – Project Presentation and Report	50%	Sundays 23:59 M1: Week 5 M3: Week 11
Assessment Task 3 (Interviews & Reflections - Individual) Case study (Task 1) Reflection (5%) Project (Task 2) Reflection (5%) AI/DSP Interview (15%)	25%	Tuesdays 23:59 T1 reflection: Week 8 T2 reflection: Week 12 Interview: Week 14
Assessment Task 4 (Micro-credentials- Individual) Academic Integrity Awareness (1%) Presenting Using Story (3%) Ethical Cities (1%)	5%	Week 2 Week 2 (with badge submission due week 4) Week 9

FAQ

- Micro-cred
    - Just submit badge link – if done prior, resubmit the link to the badge
    - Content/assessment related Q: use the "ask your instructor a question" feature on Canvas
    - Technical problem: ask ITS

- Group formation
    - How many in a group? 3-4, not more than 4
    - Can group be formed across workshop? No

- How is the project (Task 2) different from the Case Studies course?
    - It's different because AI/DSP is research focused, no coding required
    - A strong focus on ethics and responsible AI/DS
    - Assessment around: Critical thinking, innovation, ethics, research components and methods, writing and presentation
    - An output of Task 2 is a solid research report and a project plan

- Challenge Topics (broad) vs Case Study/Problem (specific)

- Research materials (references, data, evidence) must be easily accessible in the public domain

Example from Past Assignment

Case study (Task 1)

- HD example

- D example

No example for Task 2

Assessment Timeline

- Week 2: Research Integrity; Presenting using Story micro-cred due

- Week 2-3: Group formation, topic discussion. By end of week 3: group to be registered on Canvas, and a channel to be set up on Teams

- Week 4: *Workshop* – discuss and finalise topic (Task 1 & Task 2) with tutor

- Week 5: Task 2 initial report due (after Easter break)

- Week 7: Task 1 (case study) due: report & recorded presentation

- Week 8: *Workshop* – Task 1 (case study) presentations (playback) & peer review; Task 1 individual reflection due

- Week 9: Ethical cities micro-cred due

- Week 11: Task 2 final report and presentation due

- Week 12: *Workshop* – Task 2 (project) presentations (playback) & peer review, Task 2 individual reflection due

- Week 14: AI/DSP individual interview

What to do this week?

- Micro-cred deadline reminder:
  - ✤ Micro-cred: Academic Integrity, Week 2
  - ✤ Micro-cred: Presenting using story, Week 2, link submission by Week 4
- Read materials
- Getting to know each other and form groups, next week is the deadline for notifying groups
- Think & discuss the topic for the project & case study (week 2 & 3)

Next Week

Guest lecture by Amazon AWS on Culture of Innovation – Working Backwards

Working with Data & Users

A brief run through of Research Methodology & Problem Formulation

Grand challenge presentation by Cybersecurity CIC & Communiteer

Week 3 Guest talk: Culture of Innovation
(The talk won't be recorded)

"Amazon's approach to innovation has remained consistent since the company first launched - start with the customer and work backwards. Culture of Innovation provides valuable insights, lessons learned, and best practices from Amazon's cultural mechanisms by helping identify the next steps with the right AWS teams to solve problems and realize the opportunity for the end customers."

Catherine Eibner, Head of Cloud Innovation Programs from AWS will be giving the talk on Monday 15th.

Catherine's Bio: Catherine is a serial entrepreneur and a leader of innovation and growth in Australia's start-up economy. She built scalable programs and communities to support Entrepreneurs and Women in Technology while at Microsoft, then going on to become the GM of Start-ups at Incubator Blue Chilli. She joined AWS in late 2017, initially working in a start-up team in supporting and developing early stages of start-up companies and then recently joining the ANZ Public Sector business where she is the Digital Innovation lead, focused on improving the digital experiences for students and citizens by enabling digital transformation in the government, non-profit and education sectors through the innovative application of cloud technologies.
Catherine has been named one of the Top Australian Women in Tech and is recognised as one of the country's leading Start-up Mentors, actively advising start-ups, while also supporting entrepreneurs through her volunteer work and angel investments across the ecosystem.

Research integrity and ethics

A guest talk by Research Integrity and Governance team

Guest Talk: 'Research integrity and ethics'

This interactive session will introduce key concepts of responsible and ethical research. It includes discussion and activities focused on the principles and dilemmas of research integrity and human research ethics.

Dr Daniel Barr BSc(Hons) PhD *Melb.* Dan is the Principal Research Integrity Advisor at RMIT University. Dan is an experienced research manager specialising in how institutions enable research integrity. Dan also conducts research into research integrity. Dan is a Training Fellow of the Australasian Research Management Society and an Honorary Senior Fellow at the University of Melbourne.

Mary Duffy BEd(Sec)Arts, PGDipArts, MA *Melb.* Mary is one of the Coordinators, Research Governance and Ethics in the Research Ethics Integrity and Governance team, and is the coordinator for the STEM College Human Advisory Network (CHEAN). Mary has worked in research governance and ethics since 2009 and joined RMIT in 2016 from the University of Melbourne.

Dr David Blades BA(Hons) *UQ* PhD *RMIT.* David is Senior Coordinator, Research Integrity and Governance and Dr Daniel Barr's minion. He has a background in human research ethics and humanities, arts, and social sciences.

Dr Fotini Toso BA, PGDipArts, MA (Research), PhD. Fotini is one of the Coordinators, Research Governance and Ethics and is the coordinator for the DSC College Human Advisory Network (CHEAN). Fotini has worked in research governance and ethics since 2015 and joined RMIT in 2020 from the University of Divinity, where she managed the Human Research Ethics Committee.