The (many) ethics of AI: A workshop for designers who want to look at AI through different analytical lenses

Jennifer Cheung @ linkedin.com/in/jenniferkhc
Dejana Draganic @ linkedin.com/in/dejana-draganic

#SDinGov

Hello! 👋

Dejana Draganic
Service Designer
HMRC's Policy Lab

Jennifer Cheung
Responsible AI Enablement
Pandora

#SDinGov

Housekeeping

Time for questions and discussion at the end.

We'll also be around on Slack to answer any questions after the session on #track4-session-chat

Aims for today

- A better understanding of what AI is and how subjectivity runs through the lifecycle of AI
- Practical tools to interrogate potential AI solutions

Session flow

Part I

- What is AI (not)?
- AI in the public sector
- AI principles and lenses

Session flow

Part I

- What is AI (not)?
- AI in the public sector
- AI principles and lenses

Part II

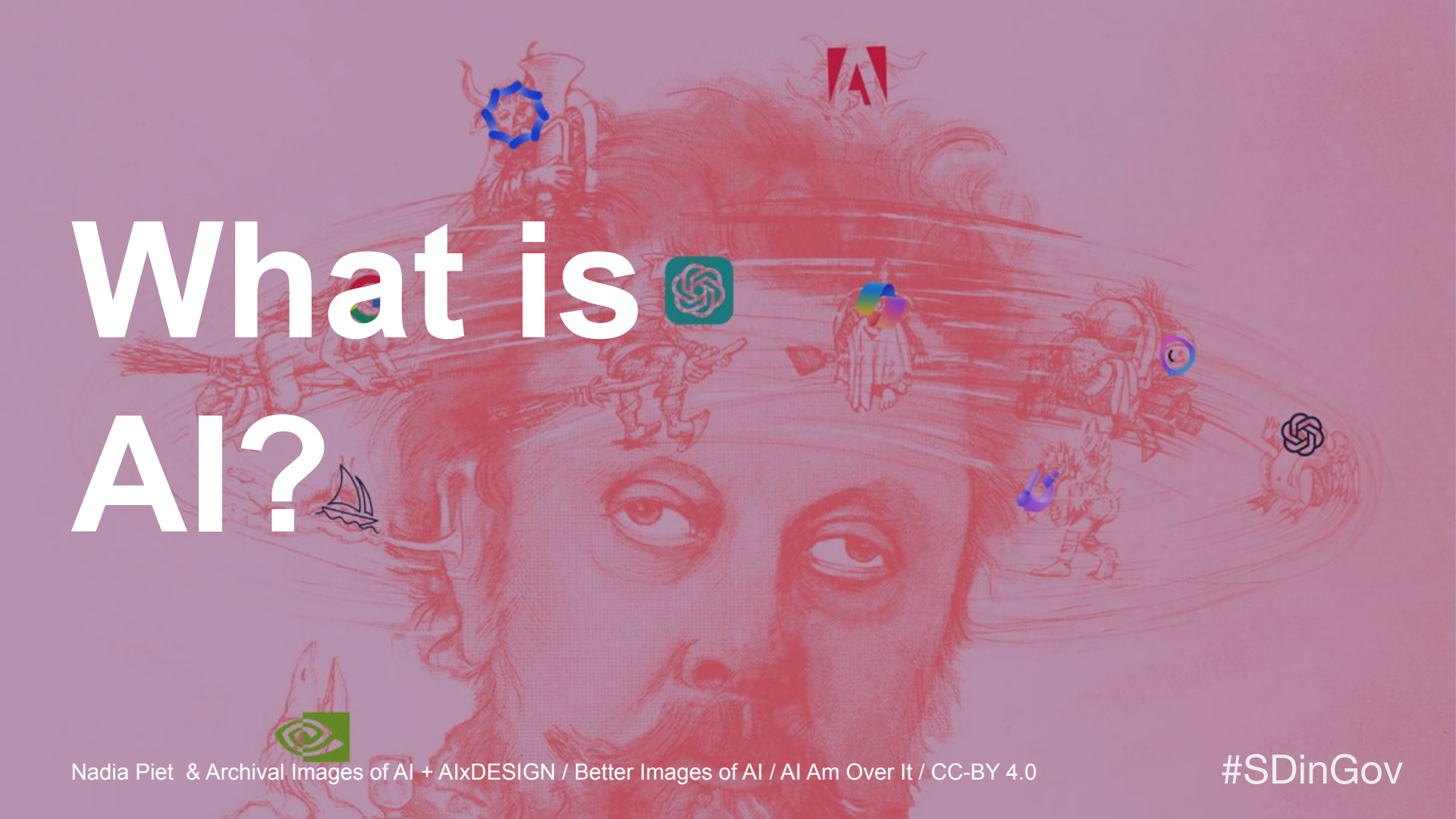
- Activity: Analysing case studies through different lenses
- Group discussion
- Feedback / wrap up

Write in the chat…

- Your name
- Role/Organisation
- Do you engage with AI in your work? How do you feel about AI? 🥶🏂📈???

#SDinGov

What is AI?



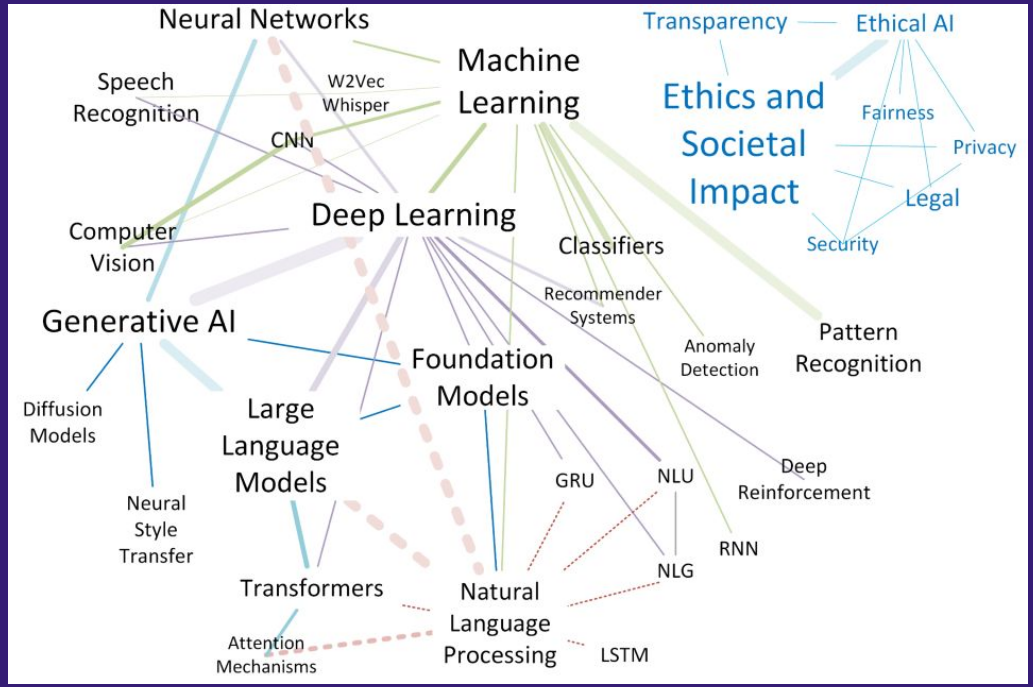
#SDinGov

AI technologies and terms

“An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.

Different AI systems vary in their levels of autonomy and adaptiveness after deployment.”

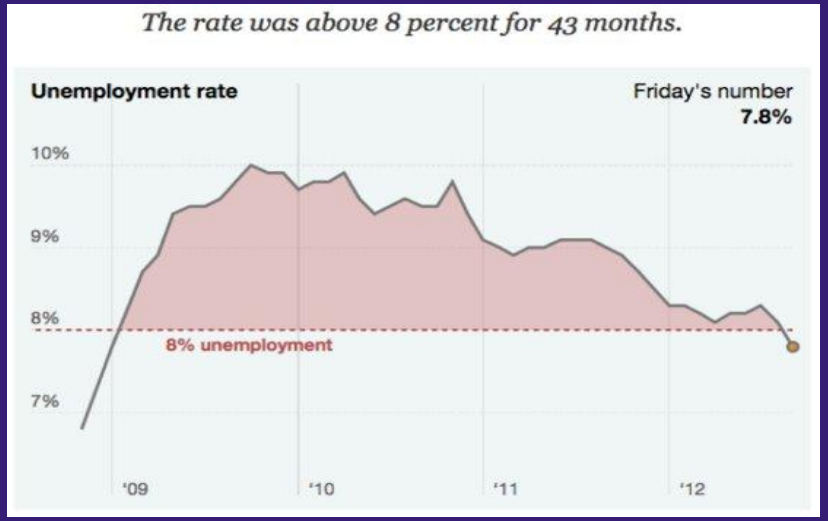
~ AI Playbook (2025), adapted from the OECD

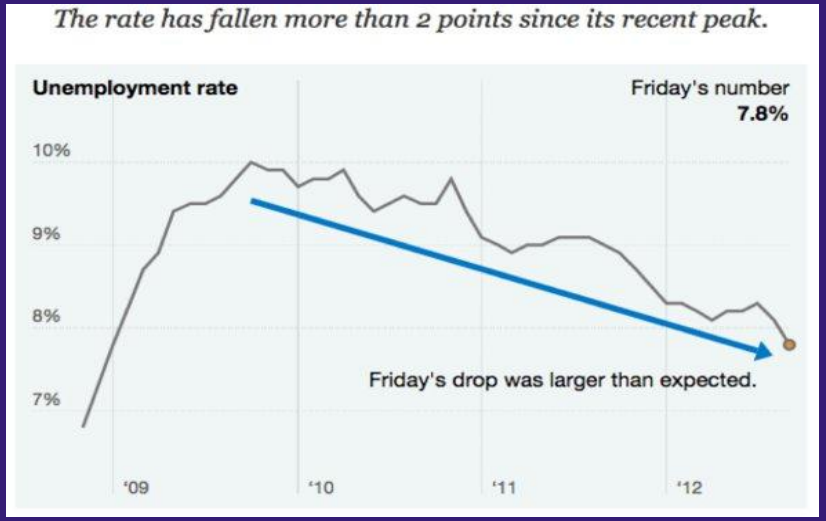


#SDinGov

AI & Data Myths

The View from Nowhere



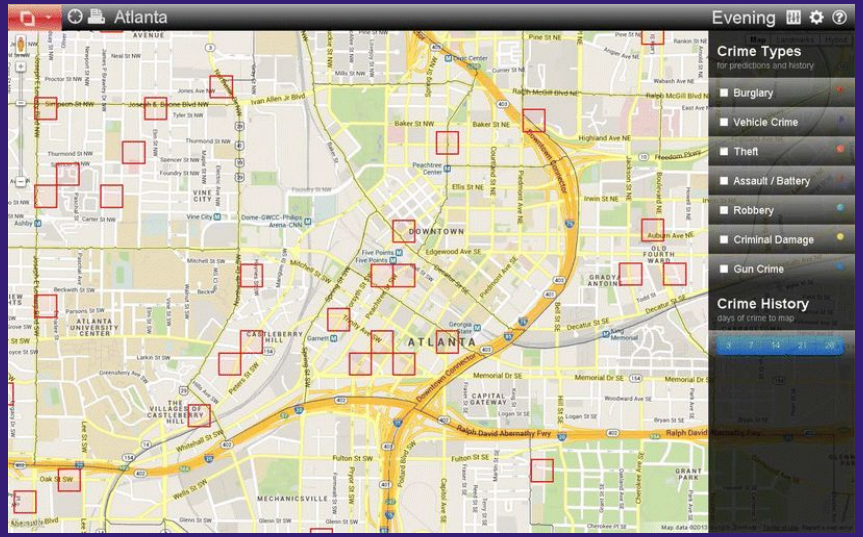


Stray, Jonathan. 2016. The Curious Journalist's Guide to Data.

#SDinGov

AI & Data Myths

More Data = Better AI



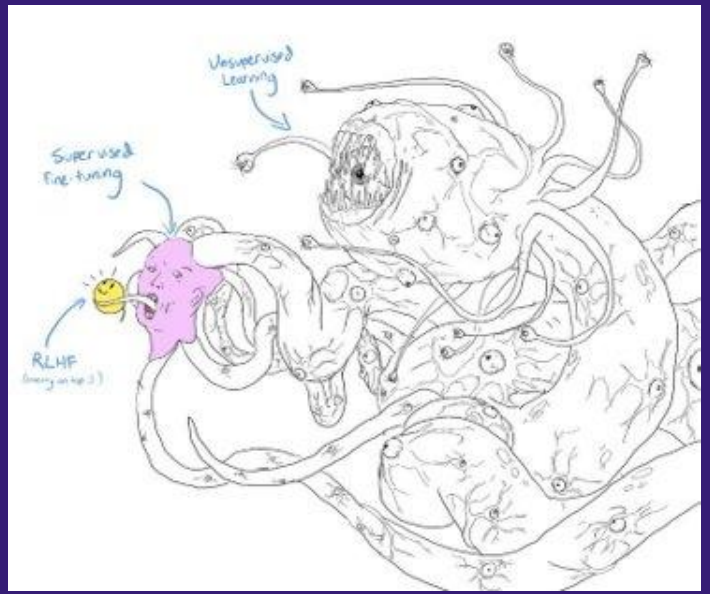
"Given the structural racism and brutality in US policing, we do not believe that mathematicians should be collaborating with police departments in this manner. It is simply too easy to create a 'scientific' veneer for racism."

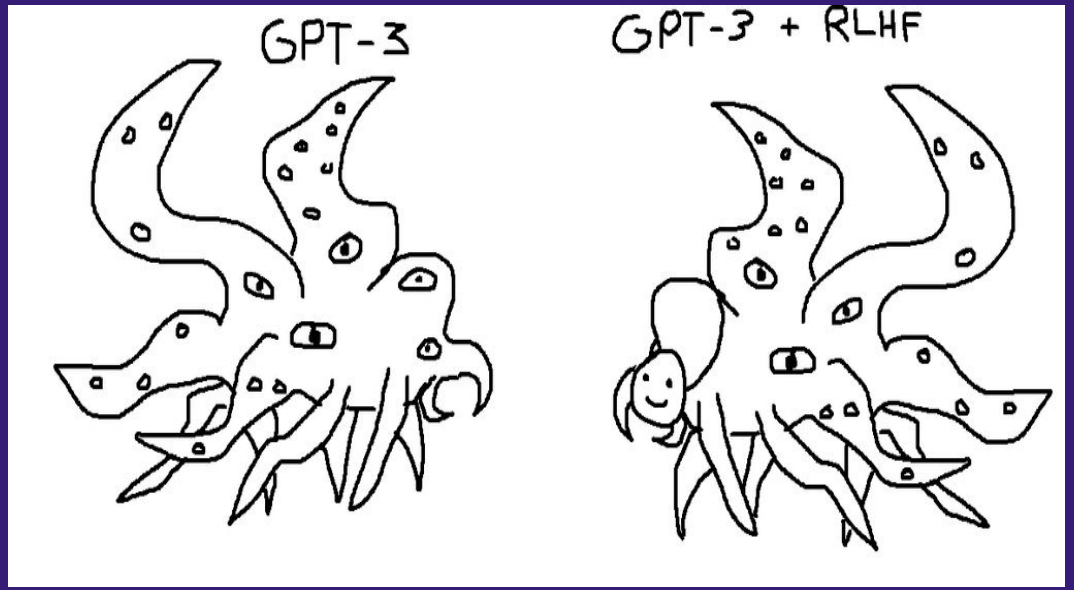
~ Notices of the American Mathematical Society

Richardson, R., Schultz, J. and Crawford, K. 2019. Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice. *New York University Law Review Online.*

AI & Data Myths

Just Follow the Data

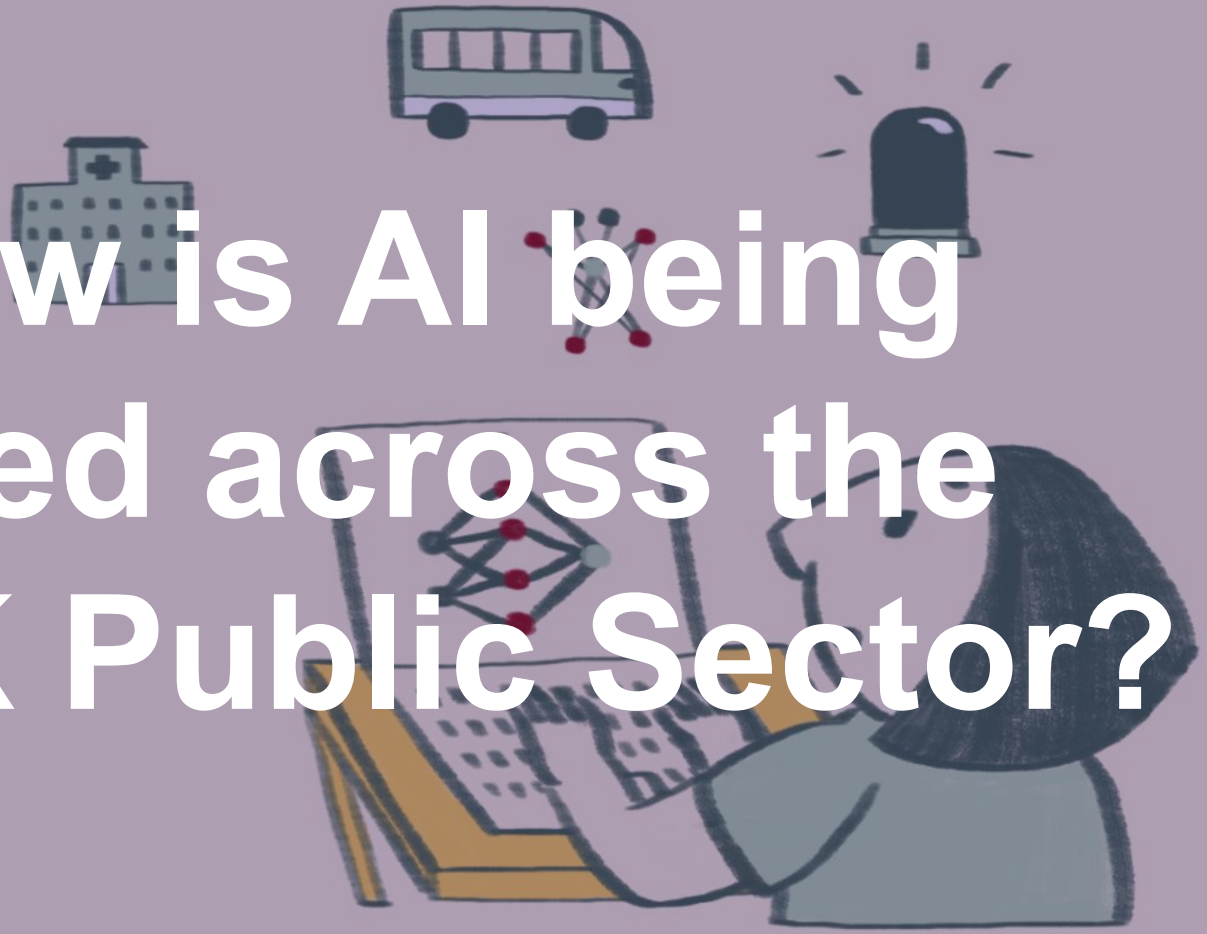




Roose, Kevin. 2023. Why an Octopus-like Creature Has Come to Symbolize the State of A.I. *The New York Times.*

#SDinGov

How is AI being used across the UK Public Sector?



#SDinGov

How (we think) AI is being used across the UK Public Sector

Healthcare
- In diagnosing eye conditions
- In analysing CT scans

Local authorities
- For prioritising social housing waiting lists

Welfare
- To detect fraud in advances claims already in payment

Policing
- To scan the faces of people in public spaces, compare them to police watchlists, and alert police officers of possible matches

Education
- To detect and provide information about the classroom environment.

Immigration
- To decide whether a couple should be investigated for sham marriage activity

Public Law Project, Tracking Automated Government 'TAG' Register (9 February 2023) http://trackautomatedgovernment.org.uk/

#SDinGov

A snapshot of AI Governance…

2021 National AI Strategy

A snapshot of AI Governance…

2021 National AI Strategy

2023 A pro-innovation approach to AI regulation

A snapshot of AI Governance…

2021 National AI Strategy

2023 A pro-innovation approach to AI regulation

2024 Generative AI Framework for HMG

A snapshot of AI Governance…

2021 National AI Strategy

2023 A pro-innovation approach to AI regulation

2024 ~~Generative AI Framework for HMG~~

2025 AI Playbook for the UK Government

#SDinGov

A snapshot of AI Governance…

2021 National AI Strategy

2023 A pro-innovation approach to AI regulation

2024 ~~Generative AI Framework for HMG~~

2025 AI Playbook for the UK Government

Honourable mention…

2025 EU AI Act

#SDinGov

A snapshot of AI Governance…

2021 National AI Strategy

2023 A pro-innovation approach to AI regulation

2024 ~~Generative AI Framework for HMG~~

2025 **AI Playbook for the UK Government**

Honourable mention…

2025 EU AI Act

#SDinGov

AI Playbook for Government | February 2025

Principle 1	You know what AI is and what its limitations are
Principle 2	You use AI lawfully, ethically and responsibly
Principle 3	You know how to use AI securely
Principle 4	You have meaningful human control at the right stage
Principle 5	You understand how to manage the AI life cycle
Principle 6	You use the right tool for the job
Principle 7	You are open and collaborative
Principle 8	You work with commercial colleagues from the start
Principle 9	You have the skills and expertise needed to implement and use AI
Principle 10	You use these principles alongside your organisation’s policies and have the right assurance in place

#SDinGov

AI Playbook for Government | February 2025

Principle 1	**You know what AI is and what its limitations are**
Principle 2	**You use AI lawfully, ethically and responsibly**
Principle 3	You know how to use AI securely
Principle 4	You have meaningful human control at the right stage
Principle 5	You understand how to manage the AI life cycle
Principle 6	**You use the right tool for the job**
Principle 7	You are open and collaborative
Principle 8	You work with commercial colleagues from the start
Principle 9	You have the skills and expertise needed to implement and use AI
Principle 10	You use these principles alongside your organisation’s policies and have the right assurance in place

#SDinGov

#SDinGov

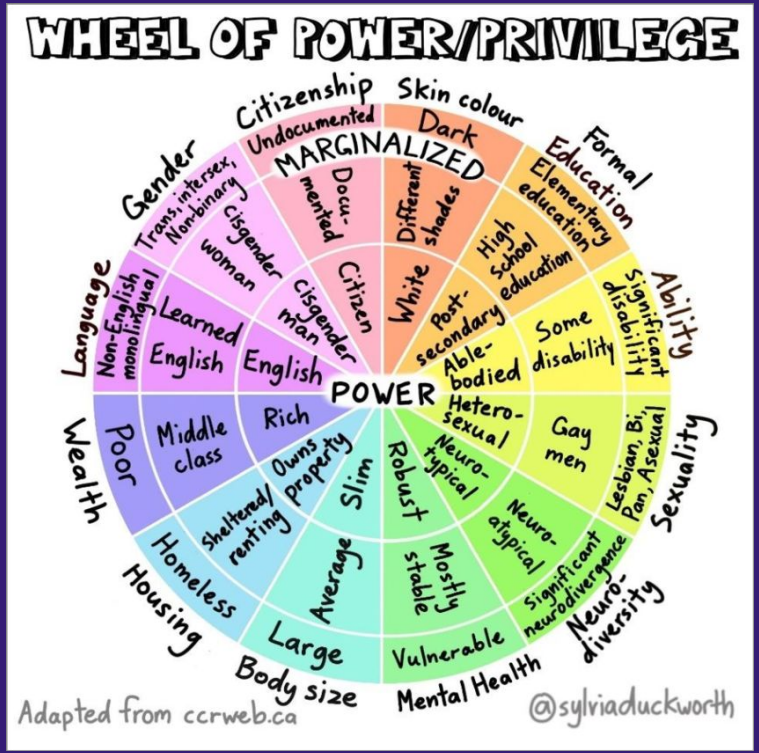
Fairness, Bias and Discrimination

“AI systems are designed, developed and deployed by humans bound by the limitation of their contexts and biases.” ~ AI Playbook (2025)

Bias can occur at different stages in the AI pipeline.

Discrimination is a direct consequence of biases and refers to the the adverse and harmful outcomes against particular groups / individuals.

Fairness means ensuring that an AI system does not amplify historical or current inequities and prejudices.



#SDinGov

Korean Chatbot Luda Made Offensive Remarks towards Minority Groups

theguardian.com · 2021

Algorithmic Bias in French Welfare System Allegedly Discriminates Against Marginalized Groups

amnesty.org · 2024

LinkedIn Search Prefers Male Names

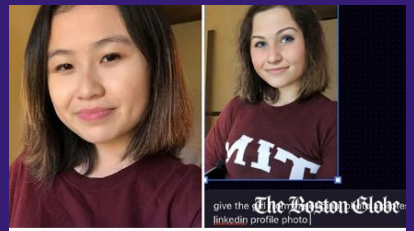
qz.com · 2016

Female Applicants Down-Ranked by Amazon Recruiting Tool

wellesley.edu · 2018

Harmful Stereotyping of Non-Cisgendered People via Text-to-Image Systems

AI-Generated Fake News Targets Black Celebrities on YouTube



AI Photo Filter Lightens Skin, Changes Eye Color in Student's 'Professional' Image

boston.com · 2023

Detroit Police Wrongfully Arrested Black Man Due To Faulty FRT

wired.com · 2022

AI Incident Database / https://incidentdatabase.ai/

#SDinGov

Stages where bias can occur

Data collection

Historical bias occurs when historical data reflects inequities that existed at that time.

Selection bias occurs when training data is not representative of the real-world population.

Discrimination by proxy Occurs when a model includes features that are correlated with protected characteristics.

#SDinGov

Stages where bias can occur

Data collection	Data labelling

Historical bias occurs when historical data reflects inequities that existed at that time.

Selection bias occurs when training data is not representative of the real-world population.

Discrimination by proxy Occurs when a model includes features that are correlated with protected characteristics.

Labelling bias occurs when the annotator's subject perceptions influence the accuracy of data labels.

#SDinGov

Stages where bias can occur

Data collection	Data labelling	Model design & training

Historical bias occurs when historical data reflects inequities that existed at that time.

Selection bias occurs when training data is not representative of the real-world population.

Discrimination by proxy
Occurs when a model includes features that are correlated with protected characteristics.

Labelling bias occurs when the annotator's subject perceptions influence the accuracy of data labels.

Algorithmic bias
Occurs when decision-making factors are unfairly weighted based on (un)conscious biases.

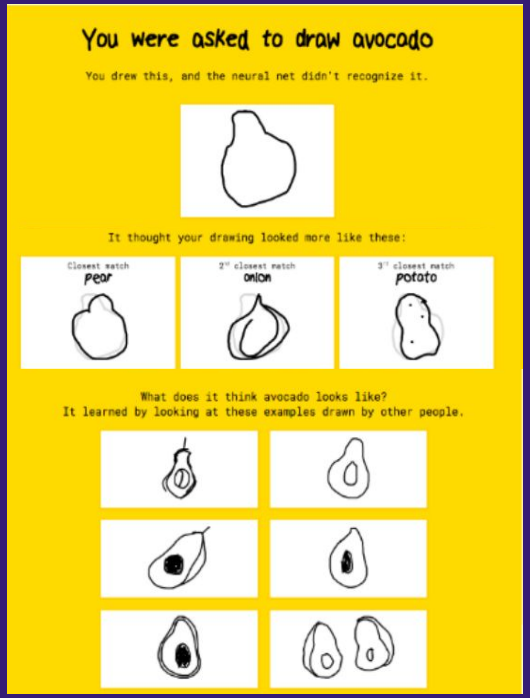
#SDinGov

Stages where bias can occur

Data collection	Data labelling	Model design & training	Deployment
Historical bias occurs when historical data reflects inequities that existed at that time.	**Labelling bias** occurs when the annotator's subject perceptions influence the accuracy of data labels.	**Algorithmic bias** Occurs when decision-making factors are unfairly weighted based on (un)conscious biases.	**Feedback loop bias** occurs when the AI model receives biased feedback, reinforcing bias across outputs.
Selection bias occurs when training data is not representative of the real-world population.			
Discrimination by proxy Occurs when a model includes features that are correlated with protected characteristics.			

#SDinGov

Transparency, Explainability and Contestability



Who are we explaining to? Explaining to everyone is explaining to no one.

A good explanation is

- **Contrastive**: why was this prediction made instead of another one?
- **Selective**: don't over-explain. Choose the most salient factors/reasons.
- **Consistent**: explanations should fit with mental models and expectations.
- **Social**: explanations should be relevant to the person. How should they use this explanation?

Carrie J. Cai, Jonas Jongejan, and Jess Holbrook. 2019. The effects of example-based explanations in a machine learning interface. In Proceedings of the 24th International Conference on Intelligent User Interfaces (IUI '19); ICO and Alan Turing Institute. 2019. Explaining decisions made with AI.

Transparency, Explainability and Contestability

Who are we explaining to?

Users
End users
Employees
Developers
Partners
Funders

What are we explaining?

Purpose/intended use/capabilities
Limitations/risks/uncertainties
Unfavourable outputs
Training data
Algorithms used
Contestability

When are we explaining?

Data collection
Model training/testing
Onboarding
Use
Monitoring/auditing

How are we explaining?

Metaphors
Source data
Visualisation
Personalisation
General vs case-based
On-demand

Who is doing the explaining?

User interface
Training
Supplementary documents
Deployer vs provider

#SDinGov

Privacy and Information Rights

The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work

Millions of articles from The New York Times were used to train chatbots that now compete with it, the lawsuit said.

Share full article 1.3K

A lawsuit by The New York Times could test the emerging legal contours of generative A.I. technologies. Sasha Maslov for The New York Times

"AI systems are so data-hungry and intransparent that we have even less control over what information about us is collected, what it is used for, and how we might correct or remove such personal information."

~ Jennifer King. 2024. Privacy in an AI Era: How do we protect our personal information? *Stanford University Human-Centered Artificial Intelligence.*

Grynbaum, M. and Mac, R. 2023. The Times Sues OpenAI and Microsoft over A.I. Use of Copyrighted Work. *The New York Times.*

We've looked at these 3 lenses…

- **Fairness, bias and discrimination**
- **Transparency, Explainability and Contestability**
- **Privacy and Information Rights**

Any questions?

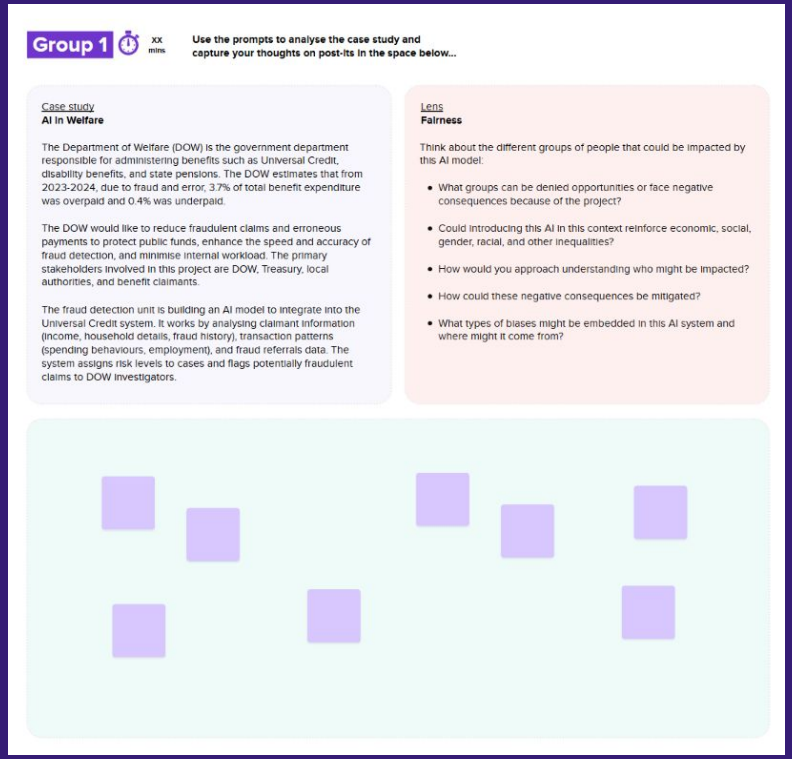
In breakout rooms…

You have been given some prompts to analyse a case study through one lens.

Capture your thoughts and discussion on post-its in the space provided.

Each group will have 2 minutes to share their key takeaways to the main room!

Any questions?



#SDinGov

Has your perspective on your case study changed after hearing how it has been analysed through a different lens? If so, how?

How might we prioritise lenses?

Are there other lenses that weren't part of this exercise but was found to be very important in the case study?

Any questions?
Pop them in the chat and we'll pick them up on Slack!

#SDinGov

Feedback wall

Thanks!

Jennifer Cheung @ linkedin.com/in/jenniferkhc
Dejana Draganic @ linkedin.com/in/dejana-draganic

Drop us any questions, comments or feedback on **#track4-session-chat** on Slack!