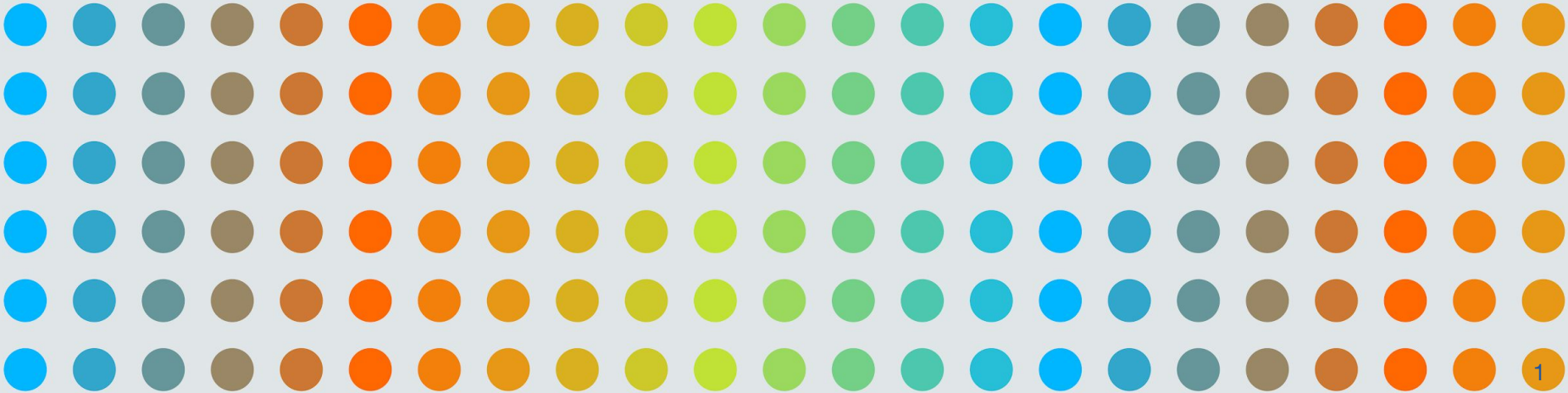
Navigating innovation responsibly with data and AI ethics

Dr Kay Achenbach
kay.achenbach@theodi.org



Aims

- **Explore** the concept of data ethics, what it is and what it isn't
- **Examine** the relationship between ethics and law
- **Review** several real-world examples, and evaluate how we identify consequences and mitigate data ethics concerns
- **Apply** Consequence Scanning to support discussions and decisions about data projects
- **Evaluate** stakeholder needs to devise a transparency approach

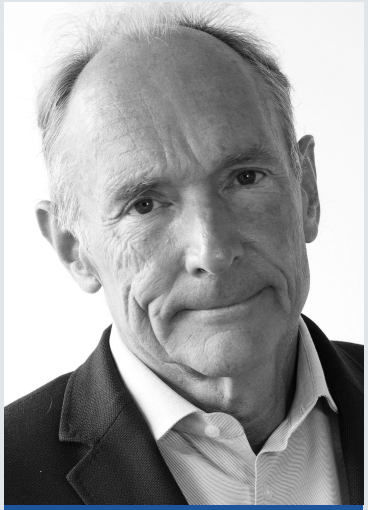
Agenda

- 9:30 - Intros and background
- 9:45 - What do we mean by data & AI ethics?
- 10:00 - Ethical frameworks
- 10:30 - Tools to facilitate ethical data & AI projects
- 10:45 - Consequence scanning activity
- 11:00 - BREAK
- 11:15 - Data Ethics Canvas activity
- 11:45 - Transparency approaches
- 12:00 - Devising a transparency strategy
- 12:30 - wrap up

Founded in 2012, the Open Data Institute (ODI) is an international, independent and not-for-profit organisation based in London, UK.

Sir Nigel Shadbolt

Executive Chair and Co-founder of the ODI



Sir Tim Berners-Lee

President and Co-founder of the ODI

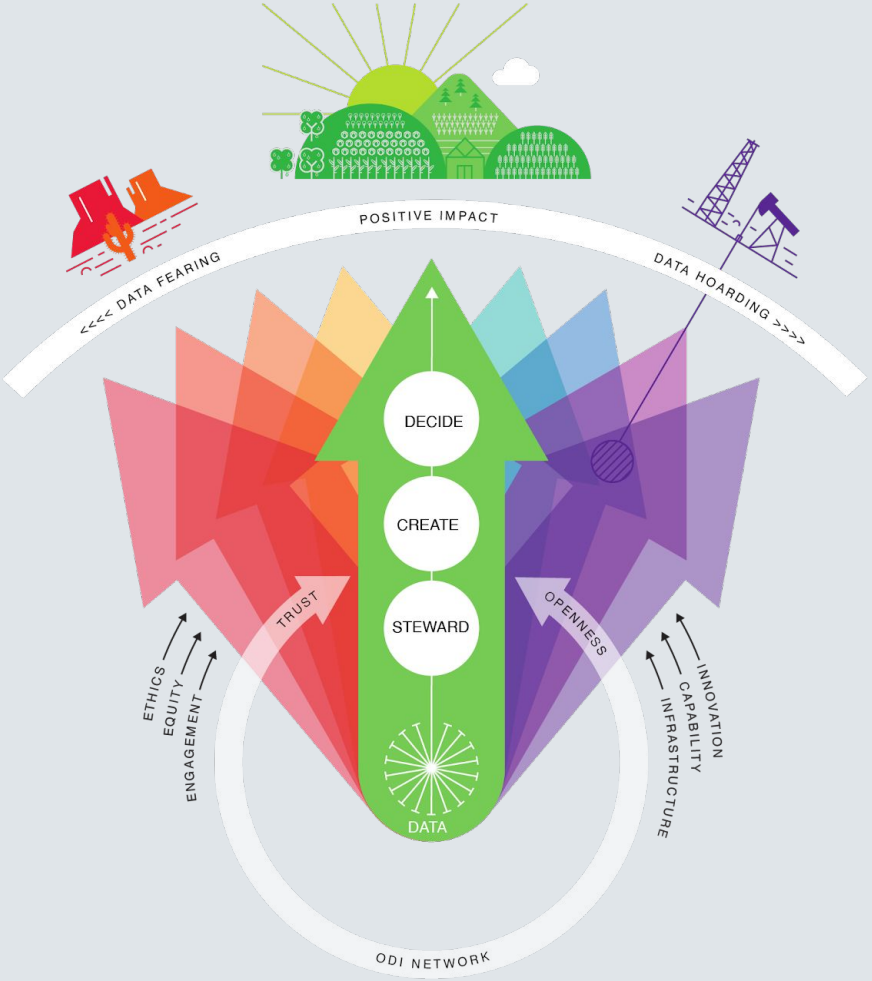
Vision

A world where data works for everyone

Mission

An open, trustworthy data ecosystem

Theory of change



Intros:

Who are you?

What brings you to the course?

Breakout Groups - 5 minutes

- What is ethics?
- What is data ethics?
- What is the difference between ethics and law?

Data Ethics

A branch of **ethics** that evaluates **data practices** with the potential to adversely impact people and society - in **data collection, sharing** and **use**.

Defined by the Open Data Institute

Childcare benefit report slams failings which ruined lives

February 26, 2024

Parents listen as the report is handed over to parliament. Photo: Robin Utrecht ANP

An Algorithm Told Police She Was Safe. Then Her Husband Killed Her.

Spain has become reliant on an algorithm to score how likely a domestic violence victim is to be abused again and what protection to provide — sometimes leading to fatal consequences.

Black Uber Eats Driver Wins Payout Over Biased Facial Recognition Checks

Pa Edriss Screenshot a Black Uber Eats driver in Oxfordshire, UK, received a payout

URBAN PLANNING

Toronto's Scrapped Smart City Reflects Distrust in Tech

An ill-fated proposal by Google-affiliate Sidewalk Labs to build a data-driven smart city in Toronto raised privacy concerns among locals. Disavowing the concept entirely, the city now plans to reimagine the site into a foliage-filled urban oasis that's more in tune with actual human needs.

BY RYAN WADDOUPS

July 11, 2022

Ethical frameworks

We need to get the best overall outcome even if it favours some people over others.

We need to ensure we are looking after the most vulnerable

Rules are rules. We need to follow them even if we don't like the outcome all the time.

We need to ensure we are treating everyone fairly.

We need to do the right thing because we have to stand up for the values we believe in.

We must make sure we uphold people's fundamental rights to [X].

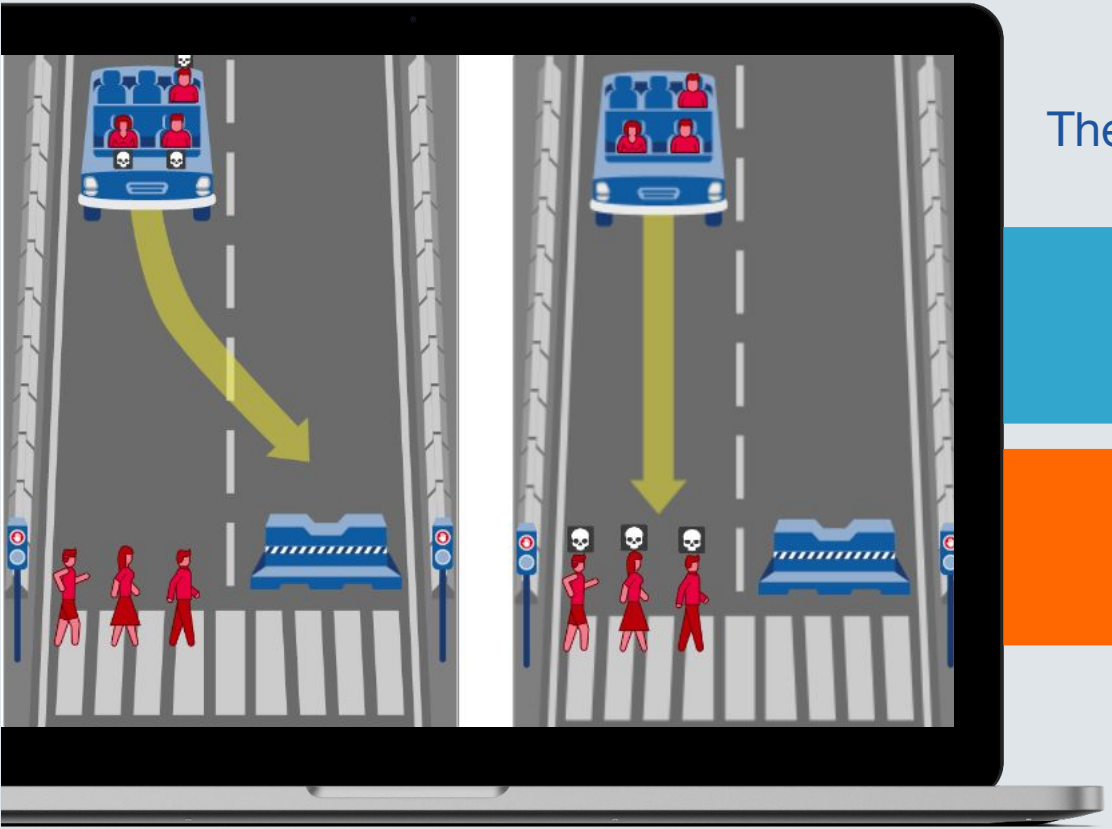
Self-driving car dilemmas

You work for the company that writes the software that controls self-driving cars. In the following dilemmas:

- The cars' brakes fail
- The steering still works, but there is no facility for human intervention
- The road has barriers on both sides, blocking access onto the pavement

Your software can only make **one** choice.

Dilemma 1: What should the self-driving car do?

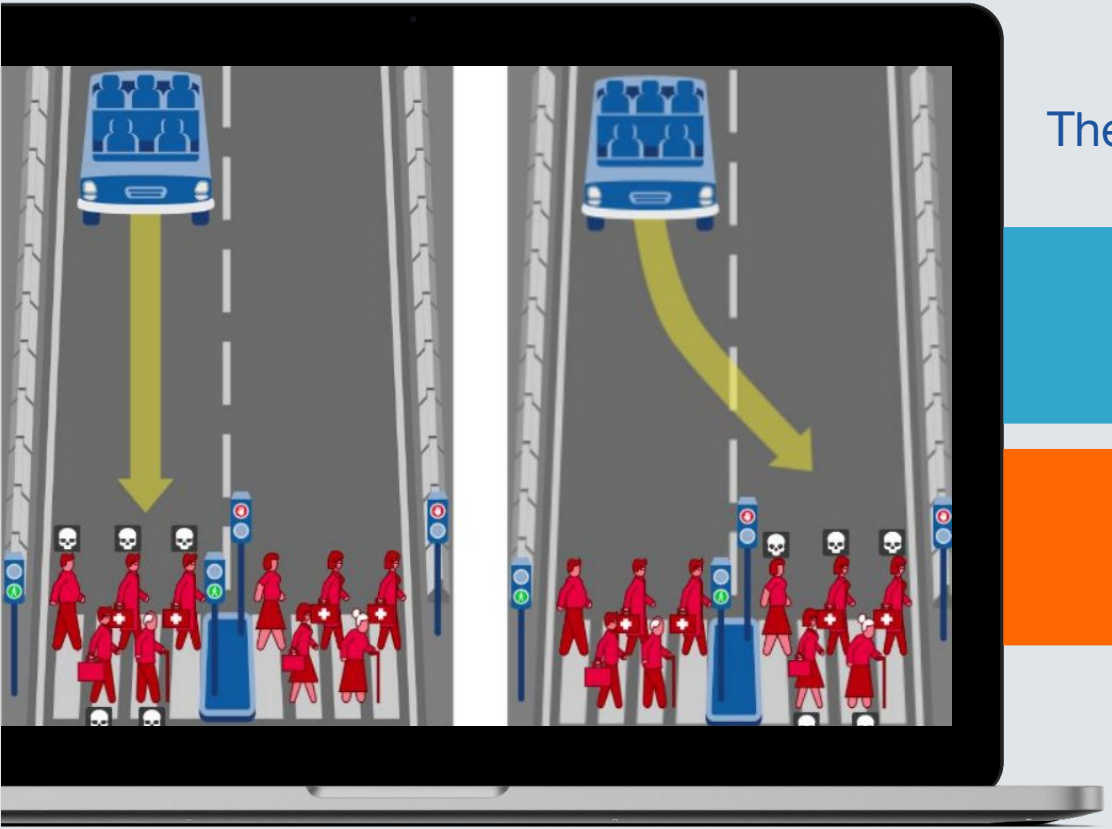


The self-driving car should:

A Swerve

B Go straight

Dilemma 2: Can you reach a shared consensus?

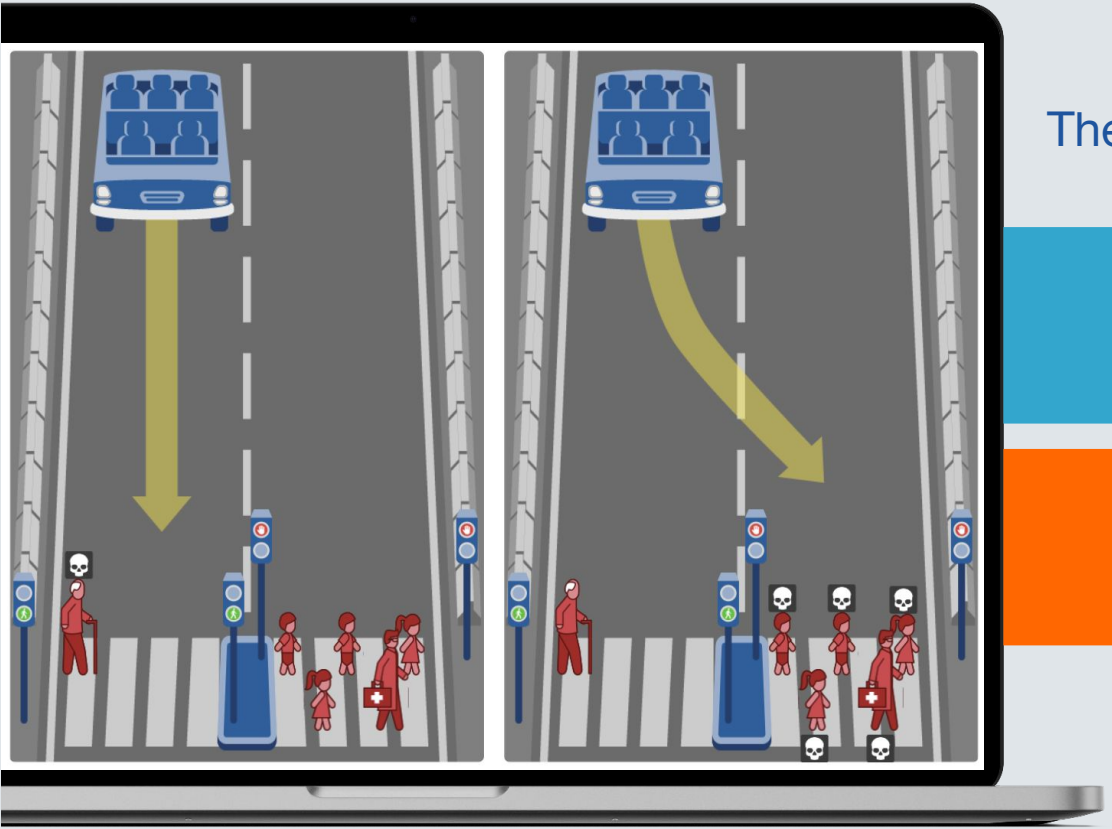


The self-driving car should:

A **Swerve**

B **Go straight**

Dilemma 3: What about now?



The self-driving car should:

A **Swerve**

B **Go straight**

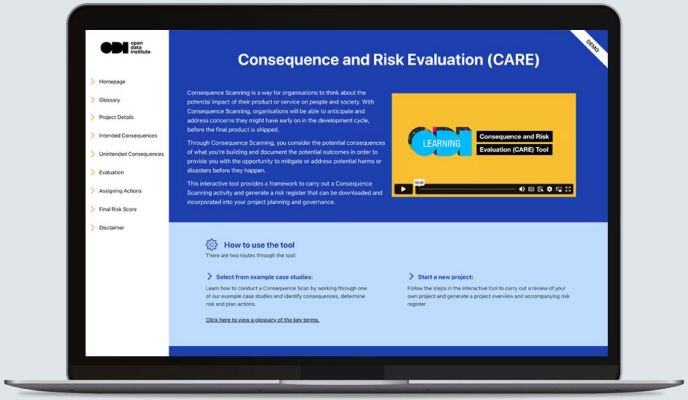
Ethics and culture

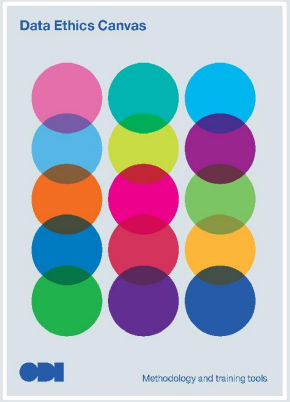
- What informs a person's perspective of ethics?
- Which set of morals/ethics are the most important?
- Whose perspective matters?
- Where do the grey areas exist?
- Is it possible to make a decision everyone agrees with?

How can we facilitate conversations about data ethics?

Data Ethics Tools

Project Level



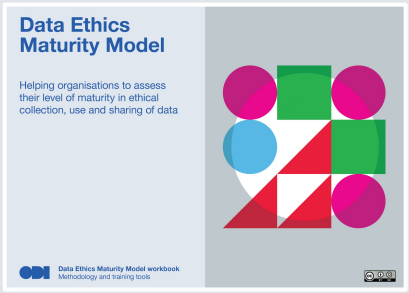


Consequence scanning
Quick assessment

Consequence and Risk Evaluation (CARE)
Interactive framework and risk register

Data Ethics Canvas
Deep consideration

Organisational Level



Data Ethics Maturity Model
Benchmarking assessment

doteveryone

Consequence scanning

Asking three questions

1

What are the intended and unintended consequences?

2

What are the positive consequences to focus on?

3

What are the consequences we want to mitigate?

What are the unintended consequences?

Imbalance in the benefits of technology

Unforeseen uses

Erosion of trust

Impact on the environment

Changes in norms and behaviours

Displacement and societal shifts

Sort potential consequences

ACT
What is your immediate responsibility?

INFLUENCE
Where do you take influencing responsibility?

MONITOR
Where do you not take any responsibility yet?

Consequence scanning

A tech startup is developing an online mental health platform that provides mental health services to clients. These services are provided via web-based interactions, as video chats telephone conversations or via text-based messaging with a qualified therapist. Group sessions and worksheets are also available. In order to provide these services, the company collects data about individuals which is categorised as the following:

- Visitor data
- Onboarding data
- Account data
- User ID
- Transaction data
- Member engagement data
- Therapy data
- Therapy quality data
- Therapist data
- Therapist engagement data
- Clinical health record

Consequence Scanning

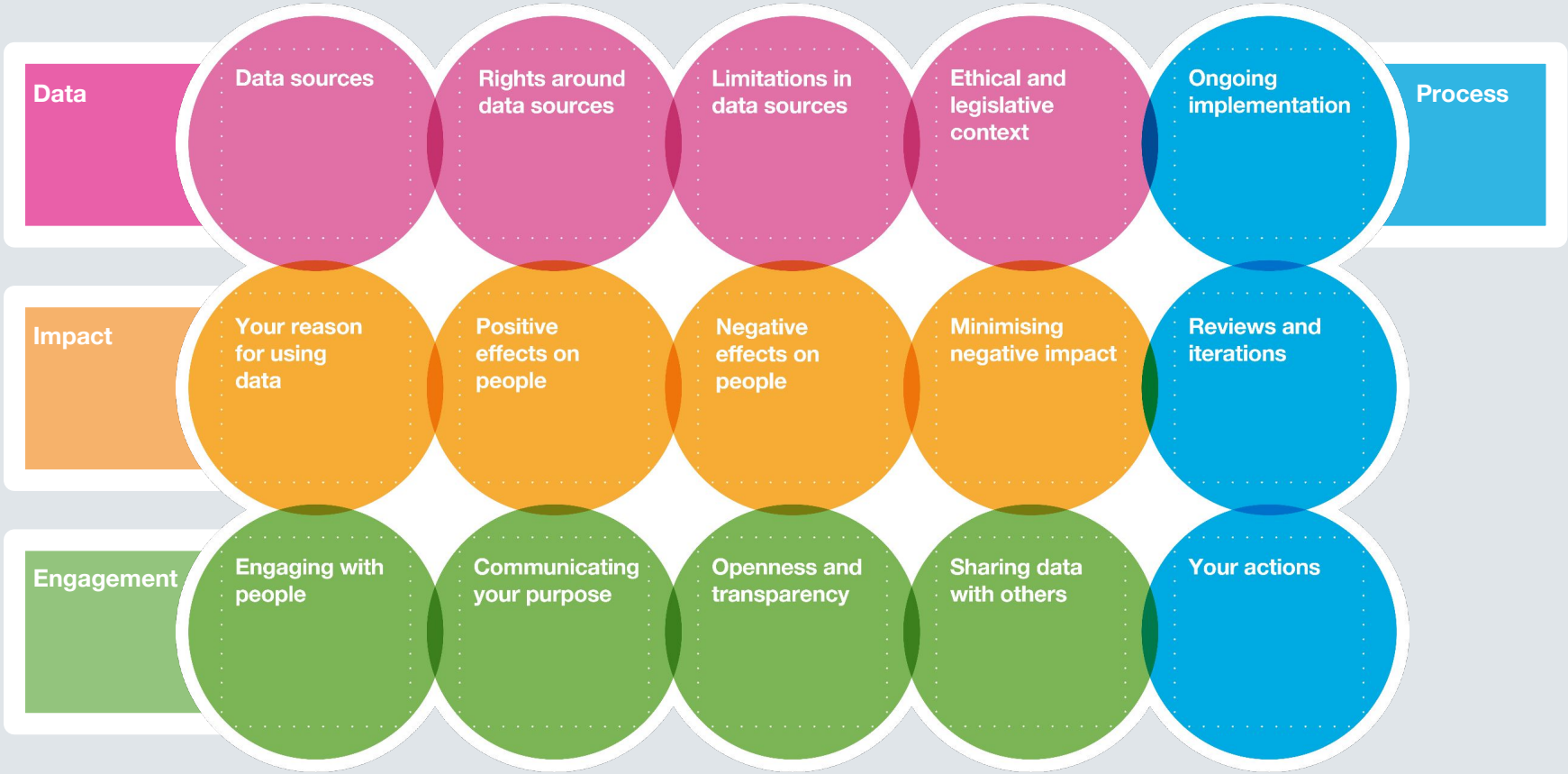
Intended consequences	Unintended consequences

What are the intended and unintended consequences?

What are the positive consequences to focus on?

What are the consequences we want to mitigate?

Data Ethics Canvas



LEARNING

This is an adaptation of the Data Ethics Canvas produced by the ODI Learning team

Data Ethics Canvas

2020-06

Data

- What do we know about the data?
- Rights/permissions?
- Limitations / bias?
- Ethical / legal / regulatory

guidance

Data Ethics Canvas

2020-06

Impact

- What is our purpose?
- What positive impacts could there be?

- What negative impacts might there be?
- How do we minimise negative impact?

Engagement

Data Ethics Canvas
2020-06

- How do we engage with people?
- Is our purpose clearly communicated?
- Are we as open as we can be about the project?
- Should we share our data and findings?

Data Ethics Canvas

2020-06

Process

- What systems/training is required?
- How often and how do we review?
- What are our actions?

Data ethics scandal: what went wrong? Whose fault was it?

‘I cry a lot, every day’: victims of the Dutch child benefits scandal fight for compensation

by Fréderike Geerdink

The government has recognized its error and resigned, but the women’s lives are still in tatters.

The Guardian

UK

Sport | Culture | Lifestyle

Australia | Middle East | Africa | Inequality | Global development

This article is more than **3 years old**

Dutch government faces collapse over child benefits scandal

Advertisement

on live TV or

Dutch childcare benefit scandal an urgent wake-up call to ban racist algorithms

The Dutch government risks exacerbating racial disc
algorithms in the public sector, Amnesty Internatior
childcare benefit scandal.

The report ***Xenophobic Machines*** exposes how racia
system used to determine whether claims for childc
fraudulent. Tens of thousands of parents and caregi
accused of fraud by the Dutch tax authorities as a r
disproportionately impacted. While the scandal brou
lessons have not been learnt despite multiple invest

Parliamentary question - O-000028/2022

European Parliament

The Dutch childcare benefit scandal, institutional racism and algorithms

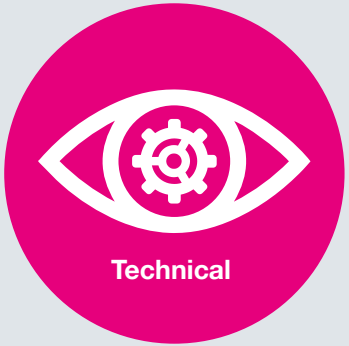
28.6.2022

Question for oral answer O-000028/2022
to the Commission

Data Ethics Canvas

The Dutch tax authority used a risk-scoring algorithm to identify people/families who were committing fraud when claiming child benefit payments. The algorithm was in use from 2013 - 2019 and now serves as an example of what can go wrong when an organisation does not consider data ethics. The algorithm was found to discriminate against people with dual nationality and against people with surnames that were not of Dutch origin. Over 26,000 people were accused of fraud, with many forced to repay tens of thousands of euros that they were actually entitled to receive. There was little transparency to the system and case workers rarely, if ever, were empowered to overrule the results of the algorithm. The appeal process was not effective and many victims were unable to receive answers as to why they were accused of fraud. The resulting scandal ultimately resulted in the collapse of the Dutch government in 2021.

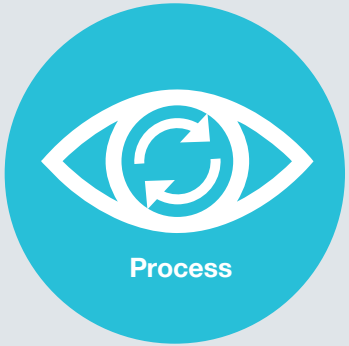
Evaluate the Dutch Child Benefit Scandal using the Data Ethics Canvas



Group 1

Group 2

Group 3



Group 4

Transparency and fairness

Terminology

Transparency:
providing information about how the system works; strategy for documentation and reporting.
What data?
What factors are considered?
What are the limitations / biases?

Explainability: providing technical reasons why the system makes certain predictions / decisions

Interpretability: allows stakeholders to predict for themselves how the system would operate in a specific instance

Transparency approaches:

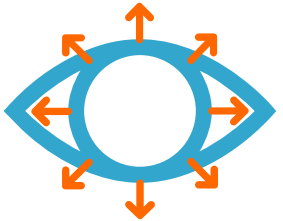
which are most important to various stakeholders? In which sectors?

Technical transparency

Training data
Model architecture
Algorithms

Process transparency

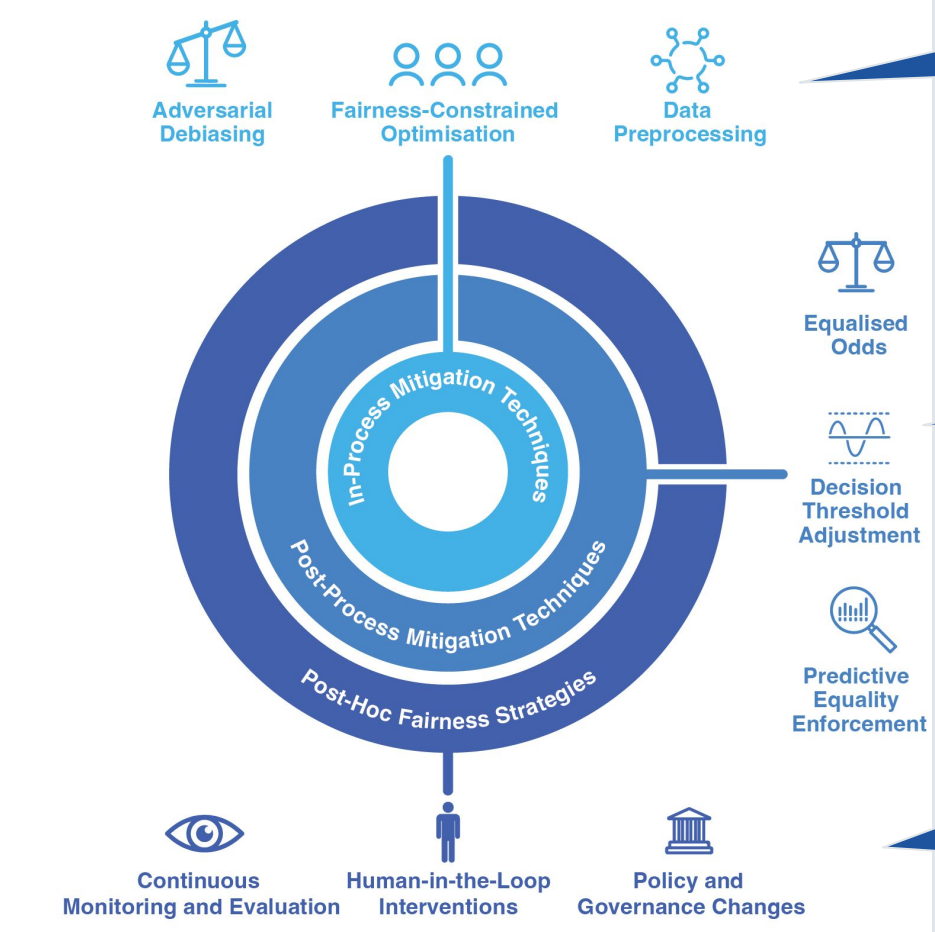
Documentation of development
Info about decision-making
How will system be used?
Audits



Outcome transparency

Info about performance for different demographic groups
Explaining specific decisions
Allowing challenges/disputes

When can bias be addressed?



In-process mitigation techniques
applied during training phase

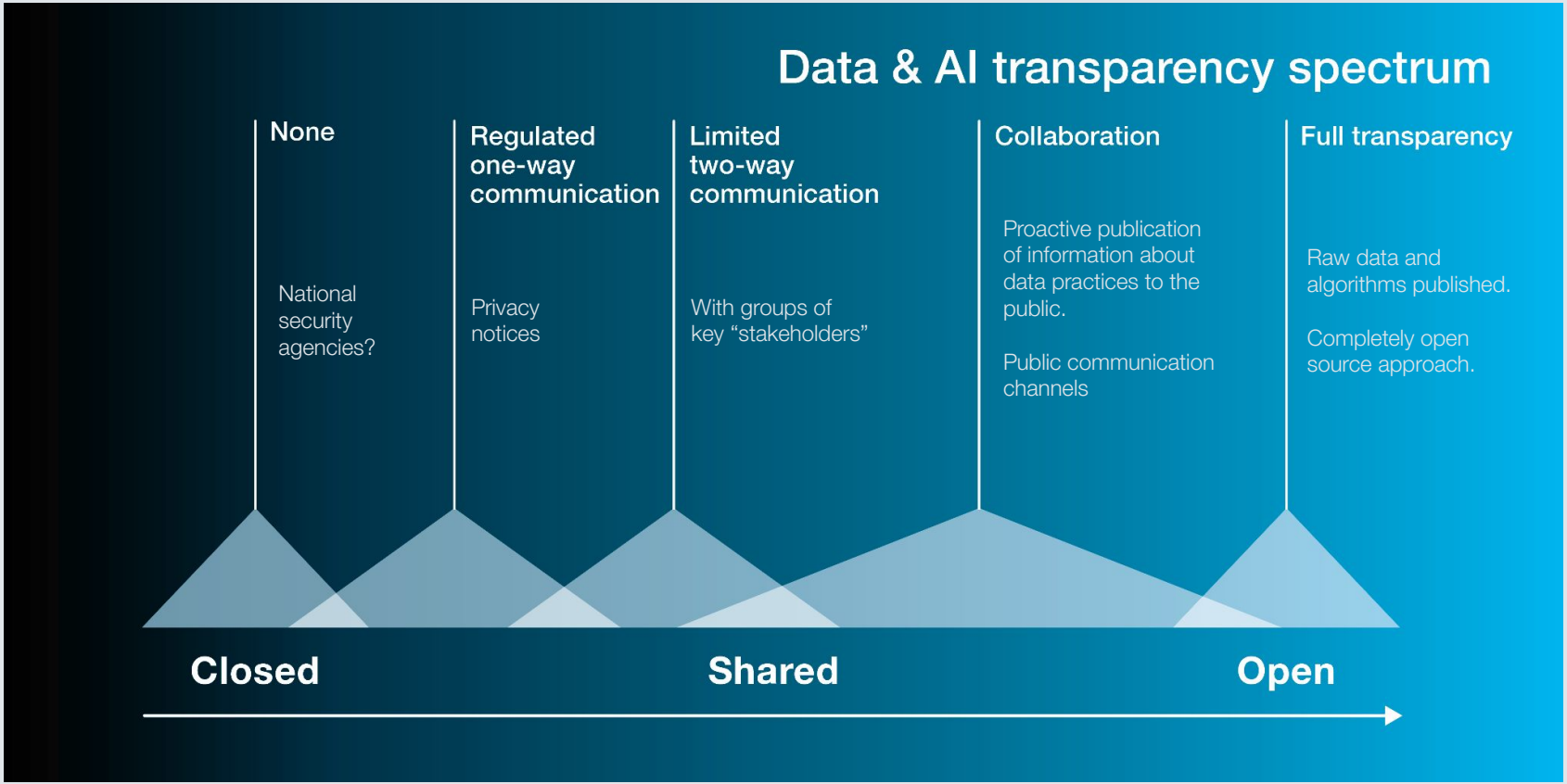
Post-process mitigation techniques
adjustments to the outputs

Post-hoc fairness strategies
do not need to be technical; don't change model outputs, but may change how they are applied

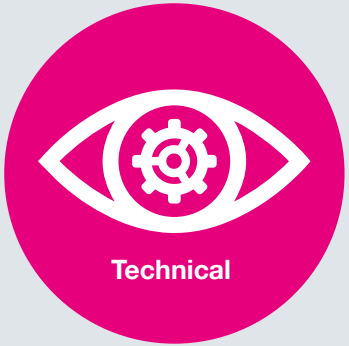
ODI research into transparency has shown:

- Transparency needs to be done ***meaningfully***
  - Address your stakeholders' real concerns
  - Think about ways to accomplish third-party auditing – and who pays for it?
  - Focus on how to reduce the asymmetry of information
- People base trust on beliefs / emotions – not necessarily just facts
  - Earn trust via two-way dialogue with community
  - Plan resourcing / time / budget to address this – knowing organisations plan for this and want to do it really does matter to stakeholders
- Technical transparency only addresses some concerns - not all
  - People don't necessarily need or want to understand the technical details
  - Do outreach, co-design your transparency strategies – make it meaningful and purposeful

Data and AI Transparency Spectrum



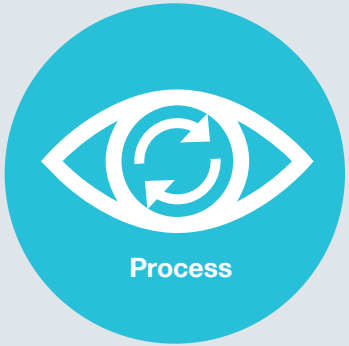
Devise the transparency strategy for the Dutch government and its fraud risk scoring for child benefits.



Group 1

Group 2

Group 3



Group 4

Session summary

Thank you!

Navigating innovation responsibly with data and AI ethics

training@theodi.org
theODI.org