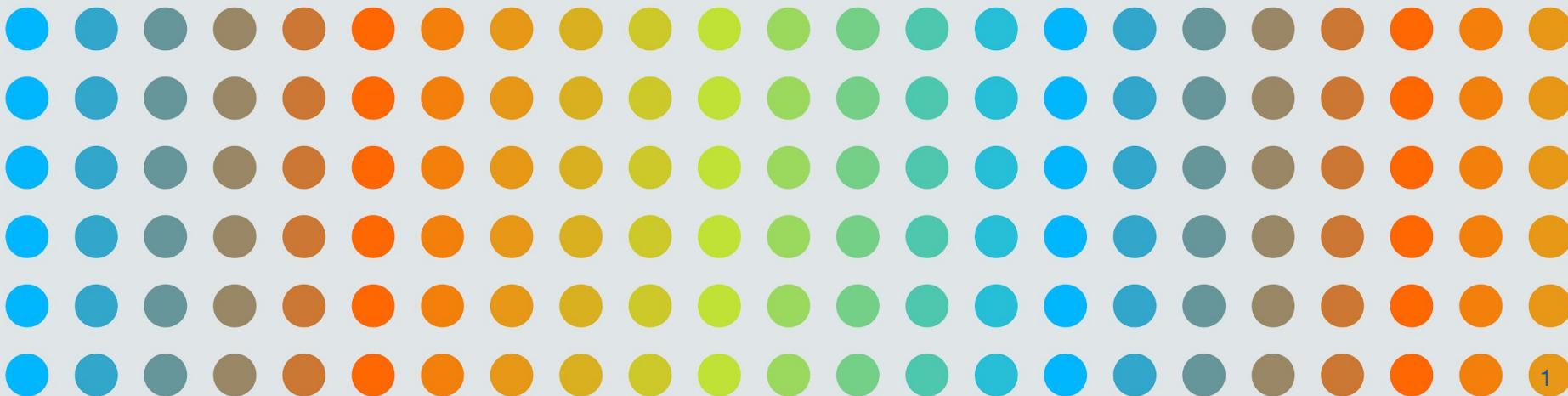


# Navigating innovation responsibly with data and AI ethics

Dr Kay Achenbach

[kay.achenbach@theodi.org](mailto: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

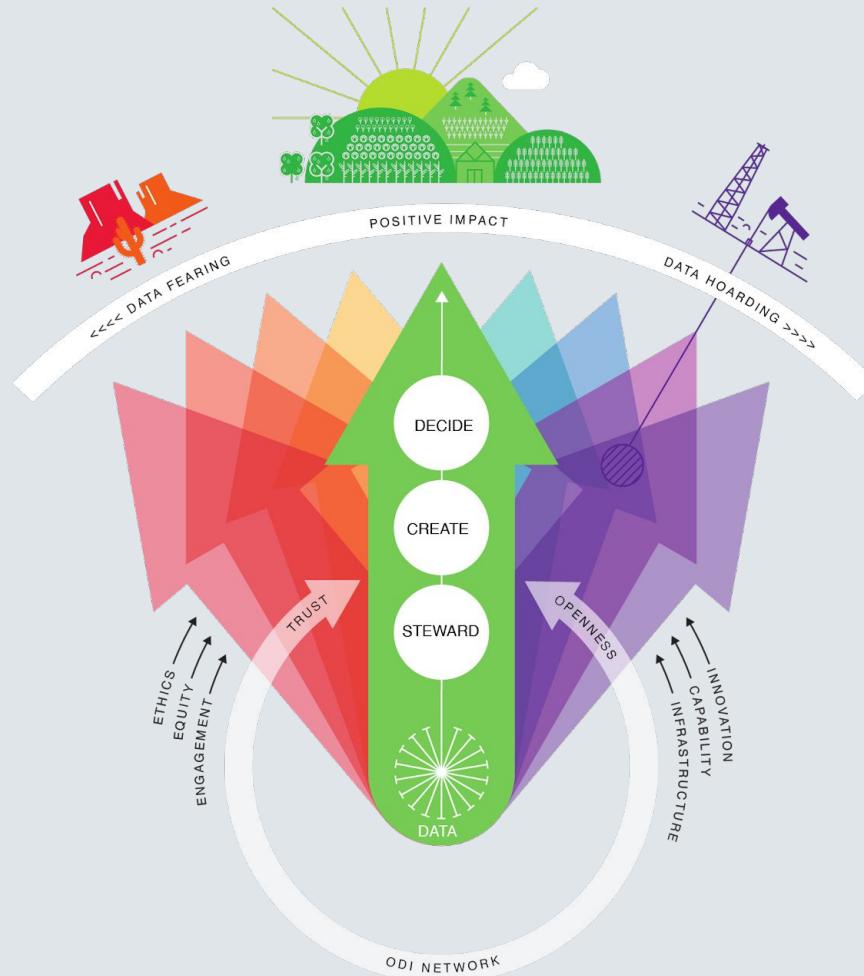
## Vision

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

## 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

## 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

## Ethical frameworks



## Self-driving car dilemmas

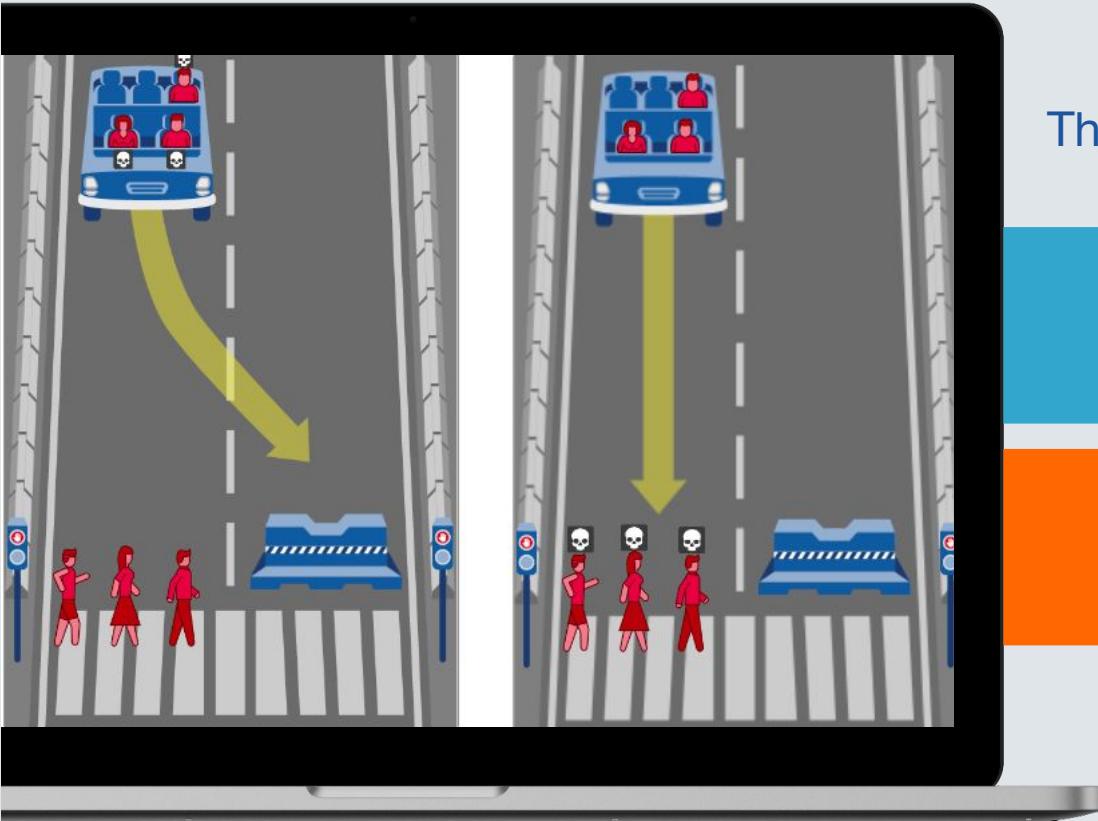
You work for the company that writes the software that controls self-driving cars. In the following dilemmas:

- 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?

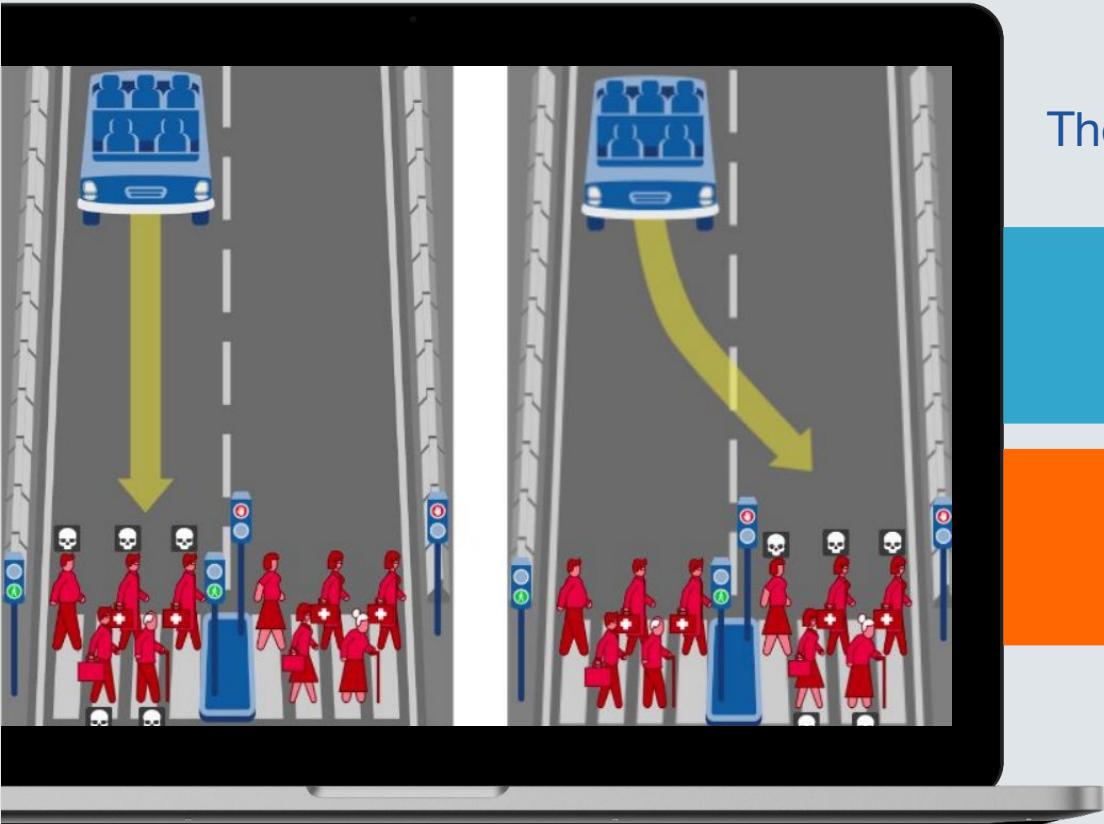


The self-driving car should:

A Swerve

B Go straight

## Dilemma 2: Can you reach a shared consensus?

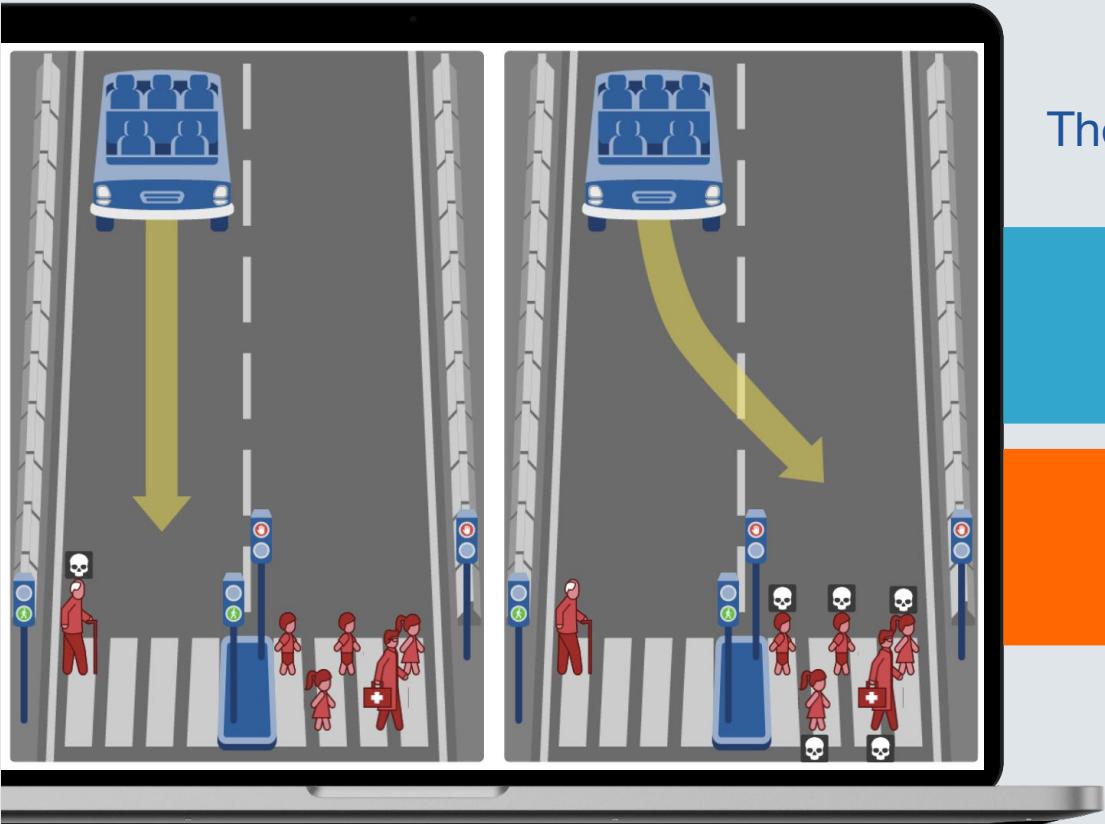


The self-driving car should:

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?



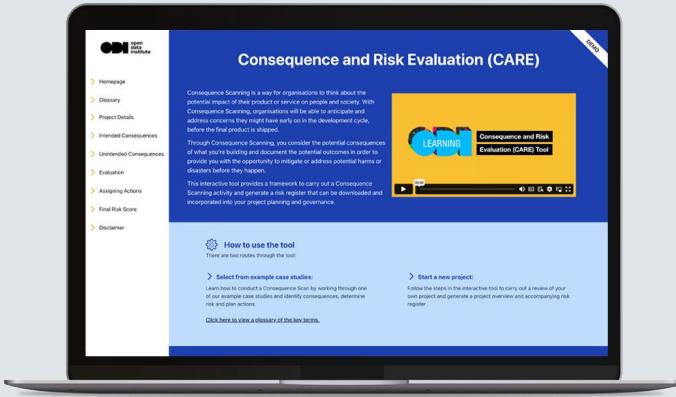
How can we  
**facilitate**  
conversations about  
**data ethics?**

# Data Ethics Tools

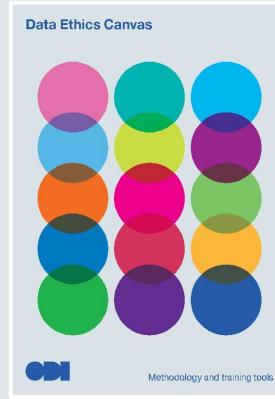
## Project Level



**Consequence scanning**  
Quick assessment

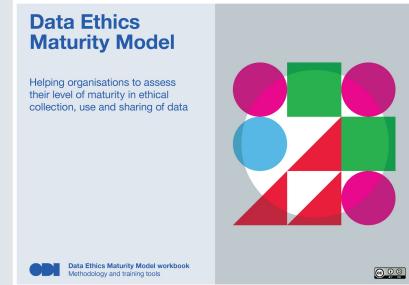


**Consequence and Risk Evaluation (CARE)**  
Interactive framework and risk register



**Data Ethics Canvas**  
Deep consideration

## Organisational Level



**Data Ethics Maturity Model**  
Benchmarking assessment

# Consequence scanning



## 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?

## Consequence Scanning

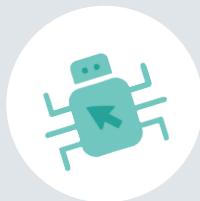
# What are the unintended consequences?



Imbalance in the benefits  
of technology



Unforeseen uses



Erosion of trust



Impact on the environment



Changes in norms and  
behaviours



Displacement and societal  
shifts

## Consequence Scanning

### 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

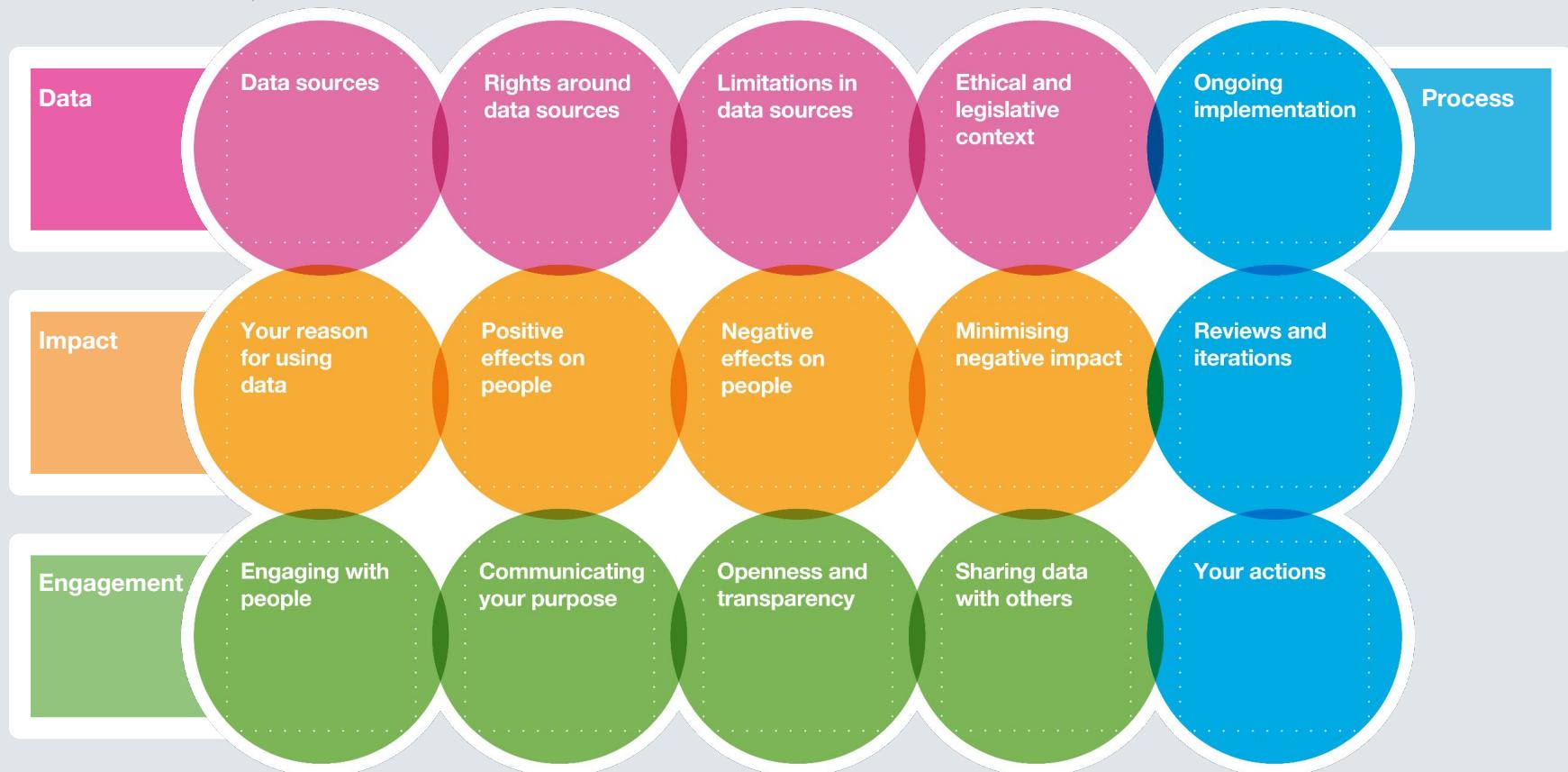
Unintended consequences

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



## Data sources

How to describe your project's key data sources, whether you're collecting data yourself or accessing via third parties.  
Is any personal data involved, or data that is otherwise sensitive?

## Rights around data sources

Where did you get the data from? Is it produced by an organisation or collected directly from individuals?  
Was the data collected for this project or for another purpose? If it was collected for another purpose, can you still use this data, or another basis on which you're allowed to use it? What ongoing rights will the data source have?

## Limitations in data sources

Are there limitations that could influence your project's outcomes?  
Consider:

- > bias in data collection, analysis, selection, analysis, algorithms
- > gaps or inconsistencies in data
- > provenance and data quality
- > other issues affecting decisions, such as team composition

## Ethical and legislative context

What existing ethical codes apply to your sector or project? What legislation, policies, or other regulation shape how you use data? What requirements do they introduce?  
Consider the rule of law, human rights, data protection, privacy, equality, non-discrimination laws and duty sharing, policies, regulation and service codes/frameworks specific to sectors (eg health, employment, taxation).

## Data

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

## Impact

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

### Your reason for using data

- What is your primary purpose for collecting and using data in this project?
- What are your main use cases? What is your business model?
- Are you making things better for society? How and for whom?
- Are you replacing another product or service as a result of this project?

### Positive effects on people

- Which individuals, groups, demographics or organisations will be positively affected by this project? How?
- How are you measuring and communicating positive impact? How could you increase it?

### Negative effects on people

- Who could be negatively affected by this project? Could the way the data is collected, used or shared cause harm or negative impacts to certain individuals, groups, demographics or organisations? Could it put people at risk, unfairly restrict access or exclude people from participating?
- How are individuals and those communities to be protected? Consider: people whom the data is about; people impacted by the use and application of the data.

### Minimising negative impact

- What steps can you take to reverse or limit harm?
- How could you reduce any偏見in your data sources? How are you keeping personal and other sensitive information secure?
- How are you measuring, reporting and acting on potential negative impacts of your project?
- What specific set of actions complete your project?

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

## Engagement

- 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?

### Engaging with people

How can people engage with you about the project?  
How can people connect information, appeal or request changes to the products/services? To what extent?  
Are appeal mechanisms reasonable and well understood?

### Communicating your purpose

Do people understand your purpose – especially people whom the data is about or who are impacted by its use?  
How have you been communicating your purpose? Has this communication been clear?  
How are you ensuring more vulnerable individuals or groups understand?

### Openness and transparency

How open can you be about this project? Could you publish your methods, key metadata, datasets, code or impact measurements?  
Can you be open for feedback on the project? How will you communicate it internally?  
Will you publish your actions and answers to this openly?

### Sharing data with others

Are you going to be sharing data with other organisations? If so, how?  
Are you planning to publish any of the data? Under what conditions?

# Process

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

# Data Ethics Canvas

2020-06

## Ongoing implementation

Are you routinely building in thoughts, ideas and considerations of people affected by your project? How?  
What information or training might be needed to help people understand data issues?  
Are systems, processes and resources available for responding to data issues that arise in the long-term?

## Reviews and iterations

How will ongoing data ethics issues be measured, monitored, discussed and actioned?  
How often will your responses to this canvas be reviewed or updated? When?

## Your actions

What actions will you take before moving forward with this project? Which show true identity?  
Who will be responsible for these actions, and who must be involved?  
Will you openly publish your actions and answers to this canvas?

# 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 Frédérique Geerdink

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

Sport

Culture

Lifestyle



Australia Middle East Africa Inequality Global development

This article is more than 3 years old

The  
Guardian

UK ▾

## 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 discrimination through the use of racist algorithms in the public sector, Amnesty International says. The report highlights the scale of the childcare benefit scandal.

The report *Xenophobic Machines* exposes how racial discrimination is built into the public sector system used to determine whether claims for childcare benefits are valid or not. The system is prone to fraud. Tens of thousands of parents and caregivers have been accused of fraud by the Dutch tax authorities as a result of the use of racist algorithms. The scandal has disproportionately impacted Black and ethnic minority families. Lessons have not been learnt despite multiple investigations.



Parliamentary question - O-000028/2022

European Parliament

## The Dutch childcare benefit scandal, institutional racism and algorithms

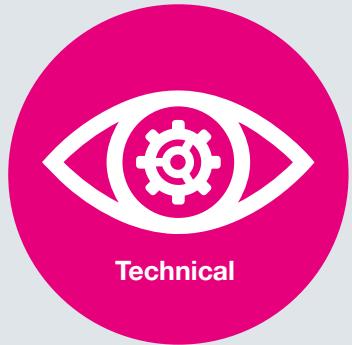
28.6.2022

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



Technical

**Group 1**



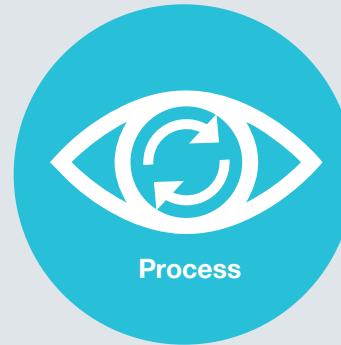
Reasons/impacts

**Group 2**



Communicating

**Group 3**



Process

**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?



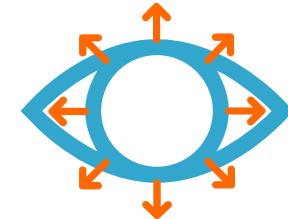
### Technical transparency

Training data  
Model architecture  
Algorithms



### Process transparency

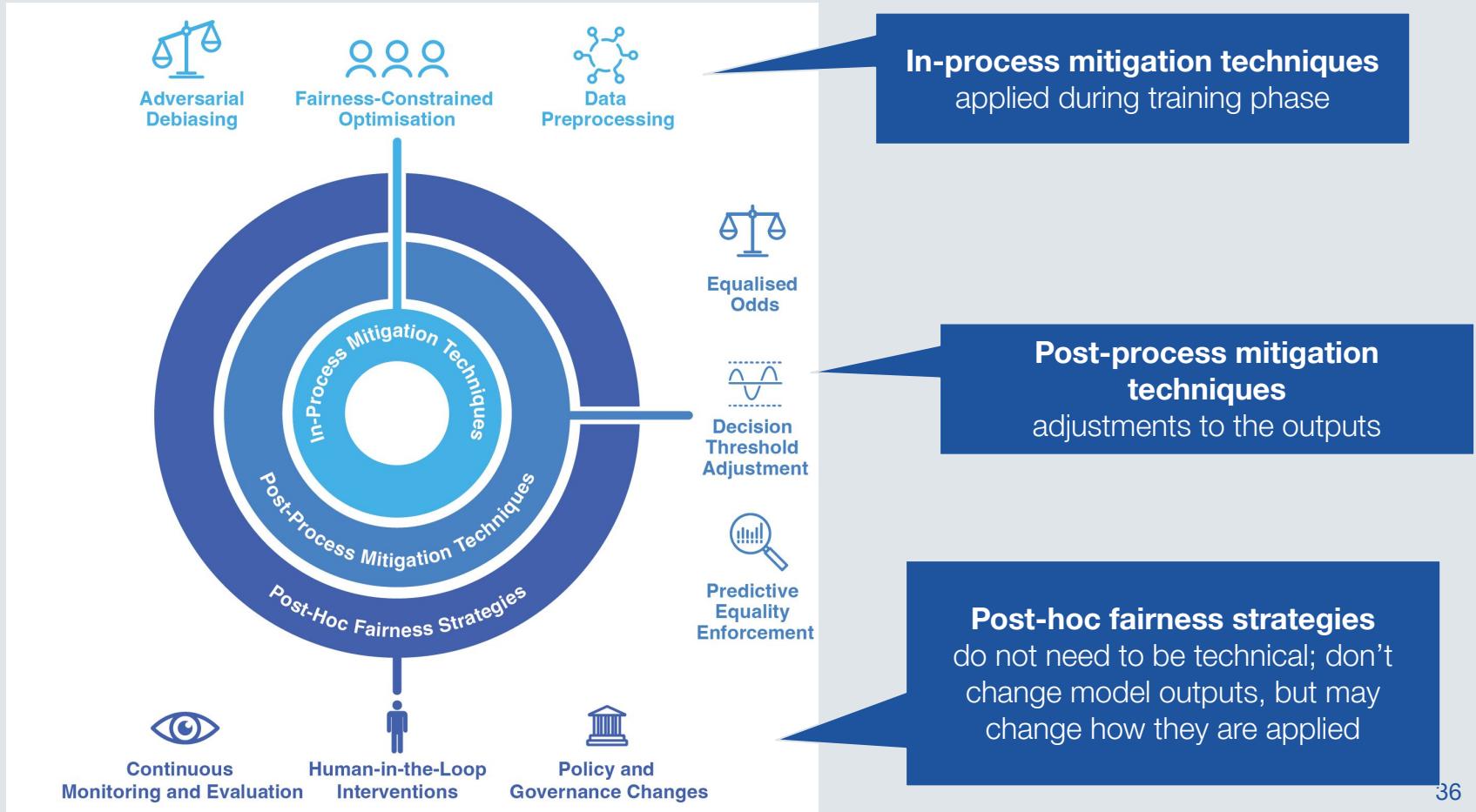
Documentation of development  
Info about decision-making  
How will system be used?  
Audits



### Outcome transparency

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

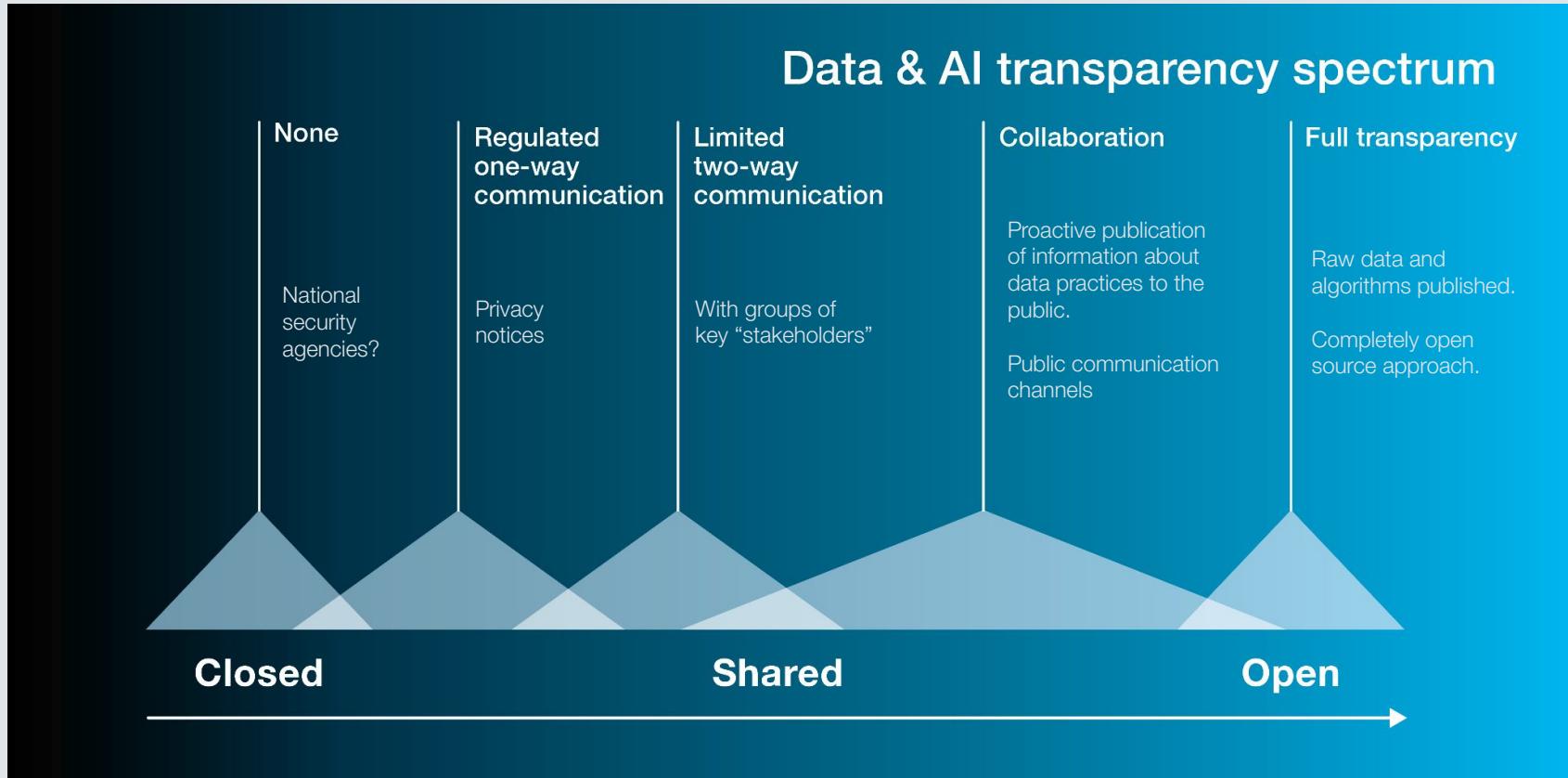
# When can bias be addressed?



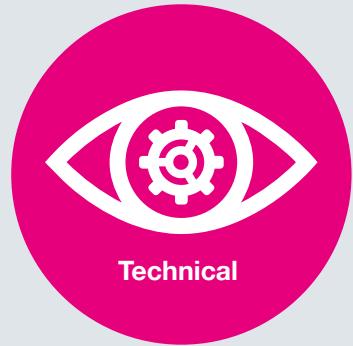
## ODI research into transparency has shown:

- 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.



Technical

**Group 1**



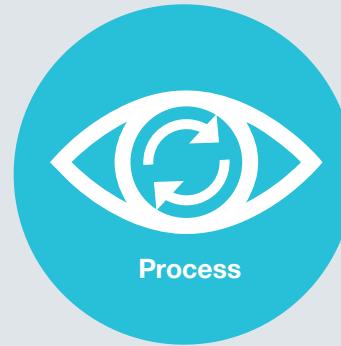
Reasons/impacts

**Group 2**



Communicating

**Group 3**



Process

**Group 4**

# Session summary

# Thank you!

**Navigating innovation responsibly with  
data and AI ethics**

[training@theodi.org](mailto:training@theodi.org)

theODI.org

