



# GOVERNMENT'S USE OF DATA TO DECISIONS

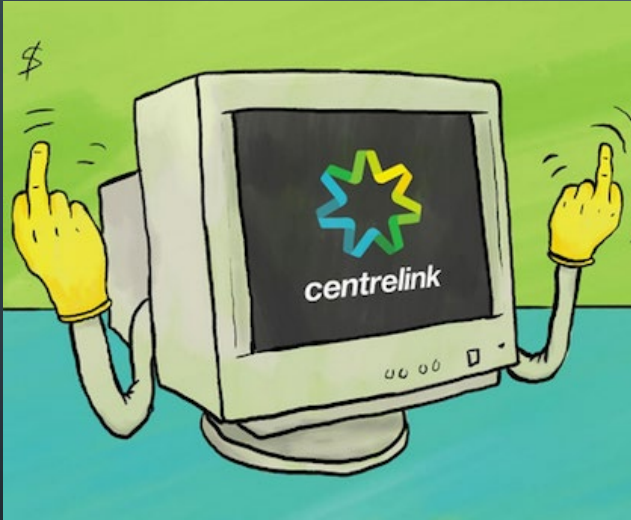
- The state vs corporations: A higher standard for the state
- Citizenship in liberal democracies. We expect:
  - Rule by law
  - Equality - differentiation must be justified (Axel Honneth)
  - Responsibility and accountability of state actors
  - Capacity for appeal, review and redress
  - Protection of rights, in recognition of balancing

# BUILDING ON THE PAST

- Public administration principles (e.g. due process; admin. justice)
- Oversight bodies and appeal mechanisms (e.g. ombudsman; audit)
- Data protection and privacy laws; EU GDPR 'right to explanation'
- Challenged by:
  - Outsourcing (distributed responsibility)
  - Use of commercial-in-confidence contracts and technology
  - Changing political values and behaviour
  - Commercial pressures
  - Ensuring compliance and appropriate restitution



# WE HAVE FAILED



**Centrelink recipient's data released by department to counter public criticism**

PAUL HENMAN @PW

# CONCEPTUAL WAYS FORWARD

- Rights based approaches
- Universalism vs differentiation: 'objective'
- Analytical framework for evaluating algorithms

# ANALYTICAL FRAMEWORK FOR ETHICAL AI (HENMAN 2019)

## DATA:

- *Does the AI incorporate social categories associated with disadvantage (e.g. sex)?*
- *Are proxies used for these (e.g. names, place of birth, address, nationality)?*
- *Is the training set data 'biased' (e.g. poor quality, reflects structurally inequalities)?*

## CODE:

- *Does the AI differentiate along social categories associated with disadvantage?*
- *If so, does it make scientific or ethical sense to treat individuals differently based on cohort characteristics?*
- *Is the algorithm based on proxy data to make decisions (e.g. neglect = abuse; student = teacher perf.)?*



# ANALYTICAL FRAMEWORK FOR ETHICAL AI (HENMAN 2019)

## USE

- *Does the AI decision involve (a redistribution of) an increase of ‘harms’ or disadvantages (particularly for disadvantaged groups)?*

## USE CONTEXT

- *Is human intervention possible in halting, reversing and correcting an AI-based decision?*
- *Do subjects have the right to an explanation for an AI based decision?*
- *Can subjects affected by an AI based decision effectively appeal and overturn that decision and seek appropriate redress?*

# ANALYTICAL FRAMEWORK FOR ETHICAL AI (HENMAN 2019)

- *Data*
- *Does the AI incorporate social categories associated with disadvantage (e.g. sex)?*
  - *Are proxies used for these (e.g. names, place of birth, address, nationality)?*
- *Is the training set data 'biased' (e.g. poor quality, reflects structural inequalities)?*
- *Does the AI differentiate along social categories associated with disadvantage?*
  - *If so, does it make scientific or ethical sense to treat individuals differently based on cohort characteristics?*
- *Is the algorithm based on proxy data to make decisions (e.g. neglect = abuse; student = teacher perf.)?*
- *Does the AI decision involve (a redistribution of) an increase of 'harms' or disadvantages (particularly for disadvantaged groups)?*



# TECHNICAL WAYS FORWARD

- Developing programming languages that:
  - encodes privacy (University of Melbourne)
  - encodes time limited data
- Developing tools to test algorithms for:
  - (level of) bias
  - Capacity to manually intervene
  - Providing an explanation for a decision

# ADMINISTRATIVE – LEGAL WAYS FORWARD

- Enhancing and extending legal provisions
  - Responsibility for algorithms used by/for government
- Appropriate funding for overview and redress systems
- Responsible Innovation Organisation
  - Preventative; front end of innovation
  - Start up culture

# CITIZENSHIP & POWER

