

# Big data, machine learning, consumer protection and data privacy in digital financial services

Rory Macmillan

ITU Webinar on Digital Credit and Consumer Protection

11 June 2020

[rory@macmillankeck.pro](mailto:rory@macmillankeck.pro)

## Financial inclusion benefits of big data and machine learning

Kenya



China



India



South Africa



ZestFinance's CEO: "All data is credit data"

## Let's start with some operational and regulatory policy dilemmas

When deciding whether or not, or how much, to lend to you, should a digital lenders' algorithm be allowed to take into account:

- a. a sudden drop in **activity in your mobile money account** as your income shrivels under lock-down?
- b. your family's recent **purchase history** suddenly buying a lot of cough medicine at the pharmacy?
- c. entries on your **social network** page about you having COVID-19?
- d. your **browsing history** on sites about dry cough, trouble breathing and losing your sense of smell?
- e. the fact that you live in a densely packed **neighbourhood** populated predominantly by an **ethnic or racial group** that is particularly hard-hit by COVID-19?

## More operational and regulatory policy dilemmas

2. Should you have **access to a human** to explain that you have had COVID-19 and so you are able to return to work and resume your normal income? Is this realistic to require?
3. Should credit reference bureaus or lenders be allowed or required to **suspend negative reporting of loan defaults** during the pandemic? Is there an equivalent measure for lenders using big data and machine learning?
4. How much of the above should be **disclosed and explained** to consumers? Really?

# FINANCIAL INCLUSION GLOBAL INITIATIVE (FIGI)

TELECOMMUNICATION  
STANDARDIZATION SECTOR  
OF ITU

(10/2019)

Security, Infrastructure and Trust Working Group

---

**Big data, machine learning, consumer protection and privacy**

Report of Trust Workstream



## Scope of discussion

- Big data and machine learning
- Consumer protection
- Data privacy

## Pre-engagement: notice and consent

- Notice and consent requirements

## Engagement: operations

- Accuracy
- Bias and discriminatory treatment
- Breach and re-identification
- Data intermediaries

## Post-engagement: accountability

- Rights of access, rectification and erasure
- Transparency and explanations
- Right to contest decisions
- Harm and liability

## Risk management, design and ethics

## Areas for further exploration

**Alternative data** collected from a wide range of non-traditional digital sources (i.e., beyond application documents and credit reference reports):

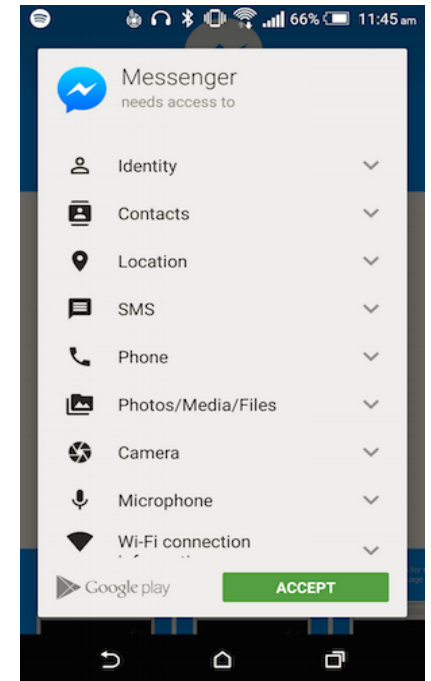
- Data from financial aggregators
- Credit card data
- Geospatial and location data
- Web scraping datasets
- App engagement data
- Shipping data from customs
- Ad spend data
- Data made available through APIs
- Location/foot traffic data from sensors and routers
- Social media data
- B2B data acquired from parties in the supply chain
- Agriculture data (e.g., feeds on corn production)
- Point of sale data
- Pharmaceutical prescription data

1993 New Yorker



*“On the Internet, nobody knows you’re a dog”*

2018 – 90% of apps on Android smartphones transfer data to Google

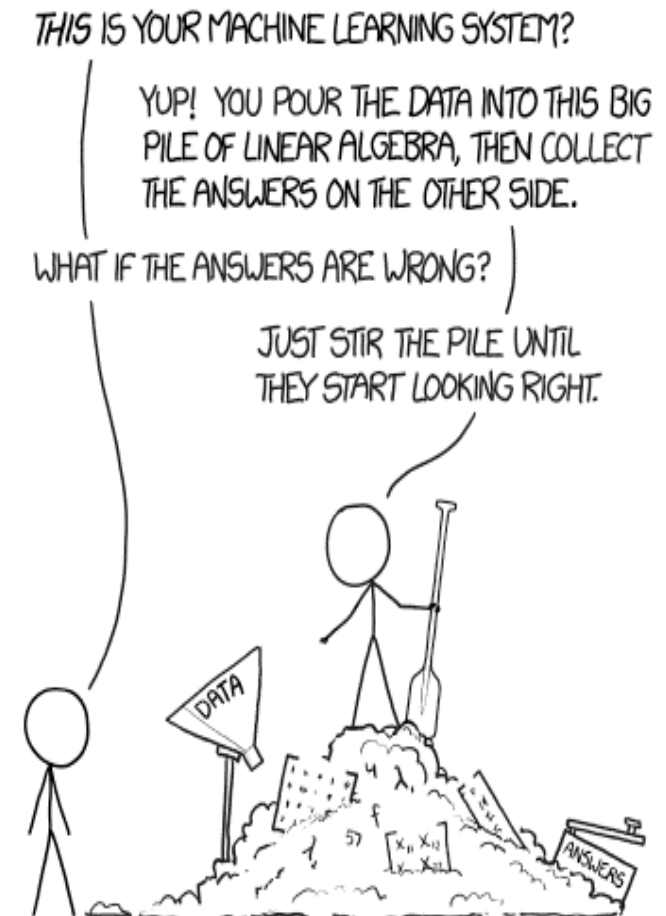


## Profiling

- Automated processing of personal data
- To evaluate, analyse or predict likely aspects of a person's interests, personal preferences, behaviour, performance at work, economic situation, health, reliability, location or movements

## Automated decisions

- Decisions made by computer processing systems
- Without any human involvement (beyond the coding)
- Typically based on inferences produced by profiling

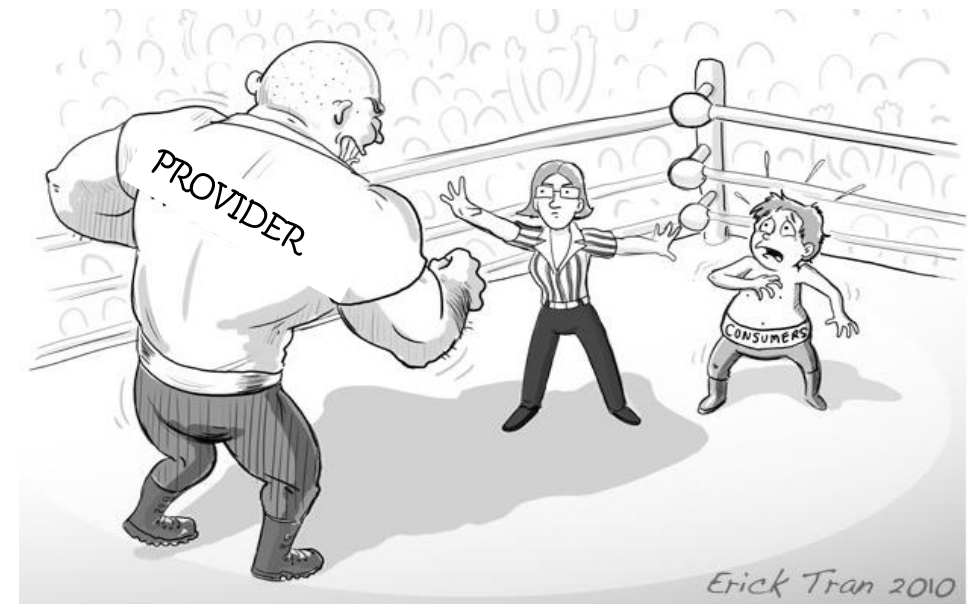


## Consumer benefits

- Suit products and services to consumer preferences, needs and risk profile
- Price at levels the consumer is willing and able to pay

## Consumer protection concerns

- **Asymmetries**
  - Information
  - Bargaining power
- **Values**
  - Fairness, accountability, transparency (FAT) and similar formulations

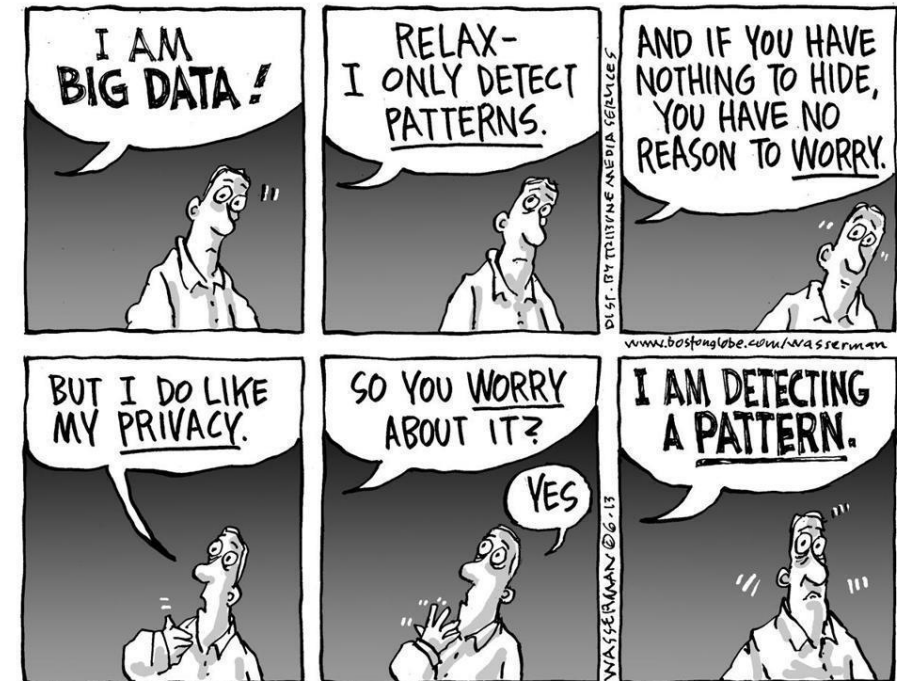


*"LET'S GET READY TO REGULATE!"*



## Privacy

- **Values, rights, protections**
  - “Individuality, autonomy, integrity and dignity”
  - Freedom in personal and family life
- **Digital context**
  - Controls on the collection, use and sharing of personal data
  - Data privacy ≠ data security
- **Risks**
  - Fraud and identity theft
  - Invasive advertising
  - Surveillance (government and commercial)



# FINANCIAL INCLUSION GLOBAL INITIATIVE (FIGI)

TELECOMMUNICATION  
STANDARDIZATION SECTOR  
OF ITU

(10/2019)

Security, Infrastructure and Trust Working Group

---

**Big data, machine learning, consumer protection and privacy**

Report of Trust Workstream



## Scope of discussion

- Big data and machine learning
- Consumer protection
- Data privacy

## Pre-engagement: notice and consent

- Notice and consent requirements

## Engagement: operations

- Accuracy
- Bias and discriminatory treatment
- Breach and re-identification
- Data intermediaries

## Post-engagement: accountability

- Rights of access, rectification and erasure
- Transparency and explanations
- Right to contest decisions
- Harm and liability

## Risk management, design and ethics

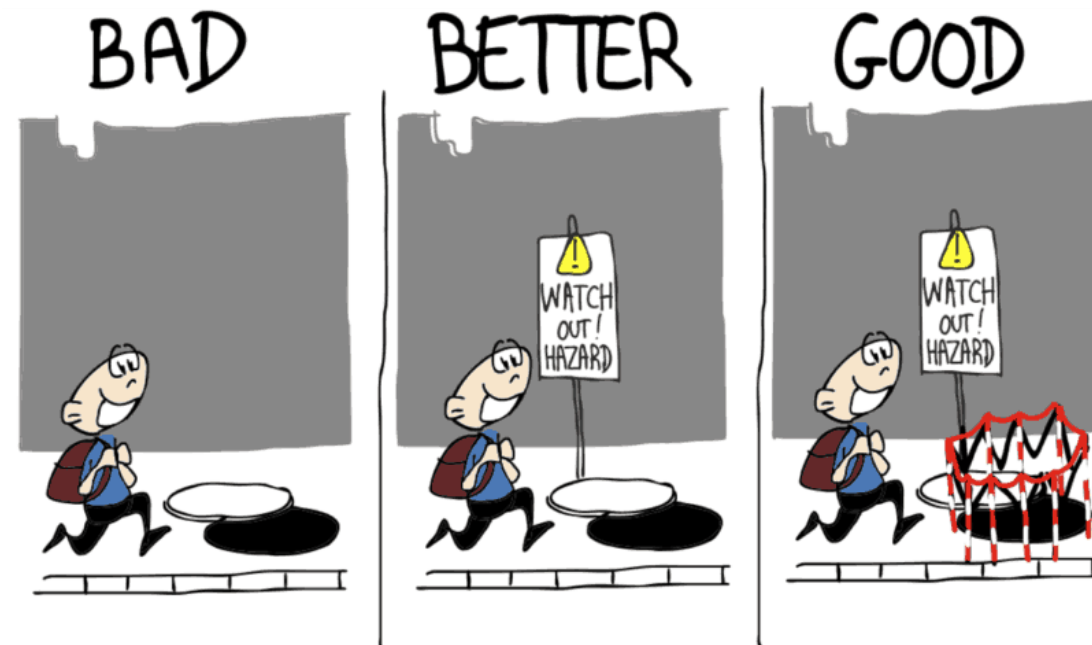
## Areas for further exploration

## Disclosure-based regime of notice and consent

- **Concern**
  - Impossibility of digesting number, scale, complexity and implications of notices
- **Data collection**
  - Purpose specification principle?
  - Data minimisation principle?
- **Automated decision-making**
  - **Notice** of “automated decision-making, [...] profiling, [...] meaningful information about the logic involved, as well as the [...] consequences”? (GDPR Arts 13-15)
  - **Opt-in/out** – e.g., “right not to be subject to a decision based solely on automated processing? (GDPR Art 22)

*The notice and consent is defeated by exactly the positive benefits that big data enables: new, non-obvious, unexpectedly powerful uses of data.*

Big Data and Privacy: A Technological Perspective, The White House, 2014



# FINANCIAL INCLUSION GLOBAL INITIATIVE (FIGI)

TELECOMMUNICATION  
STANDARDIZATION SECTOR  
OF ITU

(10/2019)

Security, Infrastructure and Trust Working Group

---

**Big data, machine learning, consumer protection and privacy**

Report of Trust Workstream



## Scope of discussion

- Big data and machine learning
- Consumer protection
- Data privacy

## Pre-engagement: notice and consent

- Notice and consent requirements

## Engagement: operations

- Accuracy
- Bias and discriminatory treatment
- Breach and re-identification
- Data intermediaries

## Post-engagement: accountability

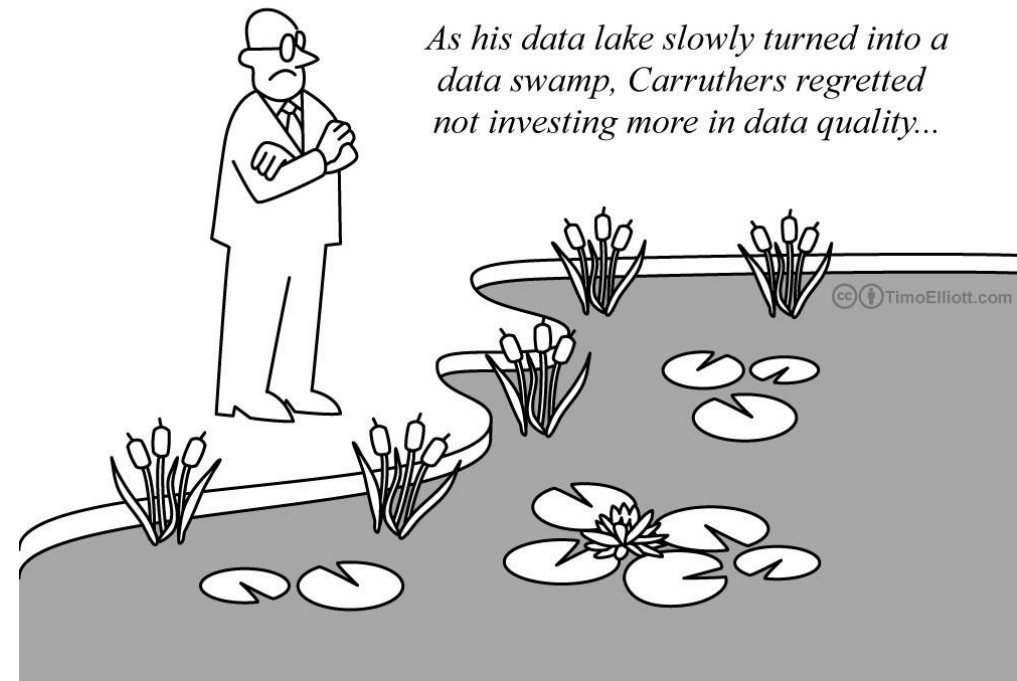
- Rights of access, rectification and erasure
- Transparency and explanations
- Right to contest decisions
- Harm and liability

## Risk management, design and ethics

## Areas for further exploration

## Responsibility for accuracy of information in the system

- **General data quality principles** – e.g., Mexico: data controllers must verify that personal data in their databases is correct and updated for the purposes for which it was gathered
- **Financial services laws** – e.g., “credit reporting systems should have relevant, accurate, timely and sufficient data – including positive – collected on a systematic basis from all reliable, appropriate and available sources, and should retain this information for a sufficient amount of time.” WB General Principles for Credit Reporting (GPCR)
- **Concern – can any of this apply to big data?**
  - Large volumes of data
  - Structured and unstructured
  - Accumulated over time
  - Multiple sources
  - Personal data and algorithm training data



## Risk of bias and discriminatory treatment

- **Concern**
  - Non-sensitive data can be used to infer or be a proxy for sensitive data, blurring the distinction

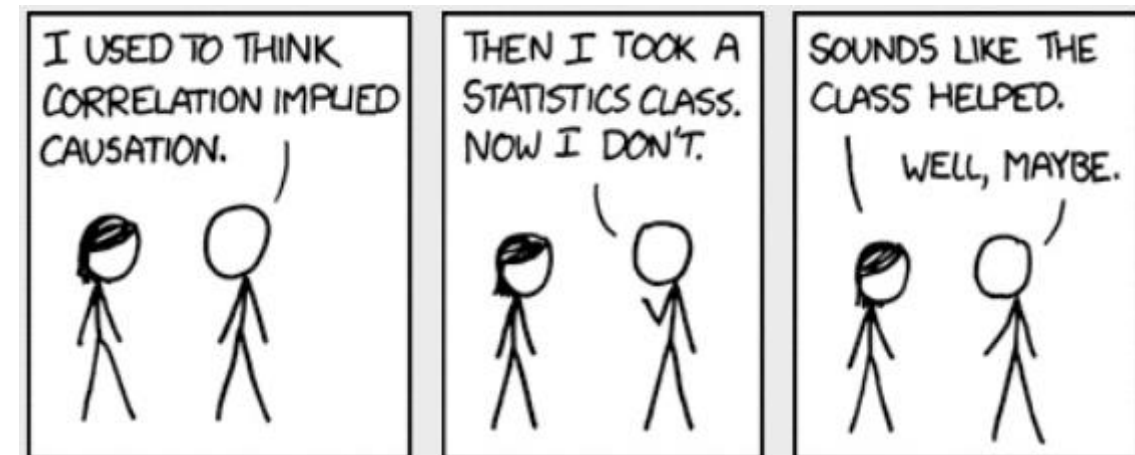
Name → religion or place of birth → race

Medical purchases → health condition → eligibility for health insurance

Postcode → ethnic or racial group

- **Correlation ≠ causation**, but a lot of machine learning is precisely about weighted clusters of proxies

*Data should “not be used in an unfair discriminatory manner in relation to digital financial services” (Principle 5 of the High Level Principles for Digital Financial Inclusion)*



2015 patent for loan application filtering based on average credit rating of an applicant's friends

# FINANCIAL INCLUSION GLOBAL INITIATIVE (FIGI)

TELECOMMUNICATION  
STANDARDIZATION SECTOR  
OF ITU

(10/2019)

Security, Infrastructure and Trust Working Group

---

**Big data, machine learning, consumer protection and privacy**

Report of Trust Workstream



## Scope of discussion

- Big data and machine learning
- Consumer protection
- Data privacy

## Pre-engagement: notice and consent

- Notice and consent requirements

## Engagement: operations

- Accuracy
- Bias and discriminatory treatment
- Breach and re-identification
- Data intermediaries

## Post-engagement: accountability

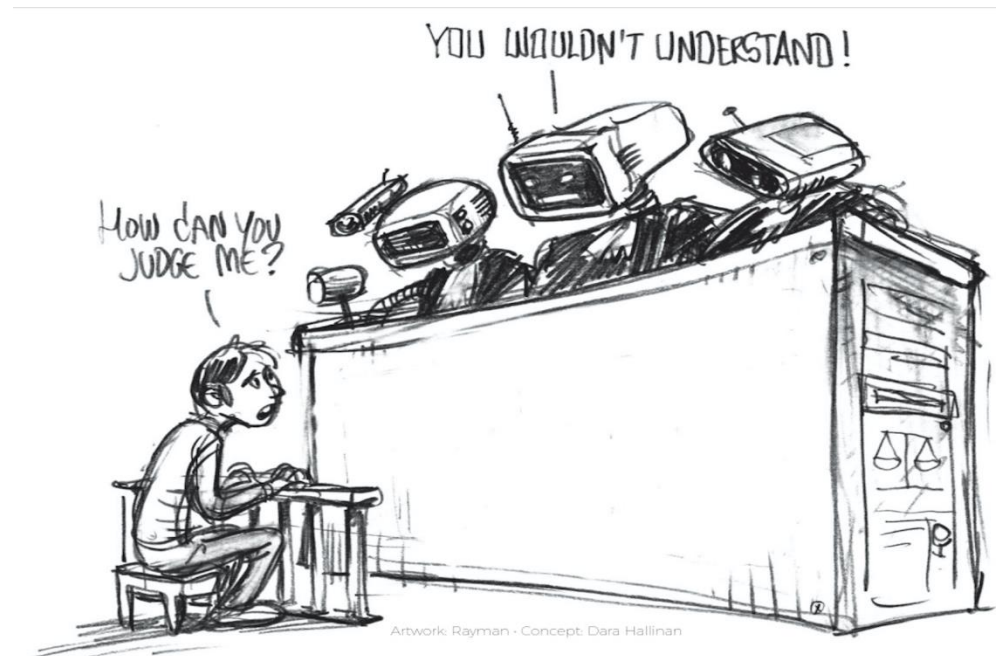
- Rights of access, rectification and erasure
- Transparency and explanations
- Right to contest decisions
- Harm and liability

## Risk management, design and ethics

## Areas for further exploration

## Difficulty of transparency and explanations

- **Concern**
  - Trade-offs between explainability and accuracy
  - Protection of trade secrets and software copyright
- **Regulating for adequate explanations**
  - Brazil's Data Protection Act 2018: right to request review of automated decisions using personal data, including profile definition, personality evaluation, and clear and relevant information on the criteria and procedures used
  - Counterfactual tools?
  - Documenting decisions made during design of model?





# FINANCIAL INCLUSION GLOBAL INITIATIVE (FIGI)

TELECOMMUNICATION  
STANDARDIZATION SECTOR  
OF ITU

(10/2019)

Security, Infrastructure and Trust Working Group

---

**Big data, machine learning, consumer protection and privacy**

Report of Trust Workstream



## Scope of discussion

- Big data and machine learning
- Consumer protection
- Data privacy

## Pre-engagement: notice and consent

- Notice and consent requirements

## Engagement: operations

- Accuracy
- Bias and discriminatory treatment
- Breach and re-identification
- Data intermediaries

## Post-engagement: accountability

- Rights of access, rectification and erasure
- Transparency and explanations
- Right to contest decisions
- Harm and liability

## Risk management, design and ethics

Areas for further exploration

## Efforts at progress



*Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems*



*Privacy Framework*



*Principles to Promote Fairness, Ethics, Accountability and Transparency (FEAT) in the Use of Artificial Intelligence and Data Analytics in Singapore's Financial Sector*



*Digital Credit Standards*

## Areas for further exploration

- **Standards for integrating privacy principles in design** (Ann Cavoukian)
- **Standards for acceptable inferential analytics**
  - assessment of output data and decisions
  - privacy acceptability of inferences
  - standards for establishing the reliability of inferences, and testing inferences before and after deployment
- **Standards for explanations** of automated decisions
  - relevance of data, relevance of inferences, accuracy and statistical reliability of the methods
  - sharing key attributes, relative weighting, documentation, audit trails
  - examining the potential for using counterfactuals to inform the consumer how, with different input attributes, they might obtain different decisions from the automated decision-making system
- **Best practices in processes** for allowing consumers to obtain human intervention
- **Principles for harmonisation of accountability**
  - procedures for contesting automated decisions
  - standards for establishing prima facie harm
  - frameworks for assessing liability