AI and Accountability in the Public Sector

Scotland's Futures Forum, Scottish Parliament, 22 June 2022

Shannon Vallor
Baillie Gifford Professor of Ethics of Data and AI
The University of Edinburgh
Director, Centre for Technomoral Futures
Turing Fellow, The Alan Turing Institute

Centre for Technomoral Futures

THE UNIVERSITY of EDINBURGH
Edinburgh Futures Institute

The Alan Turing Institute

1. Emerging Use Cases for AI in the Public Sector
2. Ethical v. Legal Accountability
3. Accountability, Answerability and Public Trust
4. Ethical Risks and Vulnerabilities
5. Barriers and Opportunities
6. Q&A

Emerging Use Cases for AI in the Public Sector

AI in the Public Sector Worldwide

Policing: face recognition, crime prediction, criminal profiling, lie detection, licence plate reading, image analysis and reconstruction

Judicial: bail/pretrial release decisions, recidivism risk assessment

Health Care: diagnostic image reading, medication monitoring and delivery, robotic surgery, triaging, risk flagging, organ and bed allocation, personalised treatment selection

Social Care: fall/activity monitoring, location tracking, medication monitoring, biometric monitoring (breathing, sleep patterns, pulse), behavioural analysis and prediction, social care support matching

Education: student risk assessment, exam proctoring/cheating detection, classroom attention/gaze tracking, behavioural monitoring, student sentiment analysis, automated marking

AI in the Public Sector Worldwide

Immigration: border security, detention surveillance, claim approval, risk assessment, identity verification, fraud detection, lie detection

Public Welfare: benefits fraud and abuse detection, automated application review, benefits determination and automatic adjustments

National Security/Defence: autonomous weapons and vehicles, cybercrime and cyberattack detection/prevention, encryption/decryption, surveillance, behavioural analytics and profiling, suspect or target identification, suspect or target tracking, risk and strategy assessment

Transportation and Infrastructure: smart city sensors, traffic management, autonomous vehicles, road/sewer/bridge defect detection, maintenance prioritisation, water and air quality monitoring, weather/flood prediction, emergency services dispatching

Questions

Which of these use cases are **scientifically legitimate**?

Which of these are Scotland's public agencies **adequately resourced** to deploy safely and reliably?

Which of these have **worked well** in other jurisdictions, and which **poorly**?

Which of these present currently **unmanageable ethical risks**?

What are those **specific** risks?

Who is **endangered** by these risks, and what is their path of **redress**?

What is required to **mitigate** the manageable risks?

The big question

Who is **responsible** for answering these questions?

That is, who is responsible for ensuring that public sector use of AI in Scotland is scientifically legitimate, safely and reliably implemented, ethically deployed, and accountable to the public and those at risk?

Ethical v. Legal Accountability

Ethical v. Legal Accountability

In the UK we have robust legal frameworks for data protection, intellectual property and copyright, and information governance

BUT:

These often get **conflated** with ethical requirements, so public agencies (and other actors) assume that if they are legally compliant with these frameworks, they are in the clear ethically.

This is false.

Ethical v. Legal Accountability

Ethics enters gaps where legal accountability is (or is seen by publics to be) porous, weakened, or incomplete.

Almost all AI use cases in the public sector have sizable gaps of this kind.

Ethical v. Legal Accountability

In these cases, ethical demands emerge from:

- Formal ethical codes of professional societies (e.g. medical, legal, and engineering professions)
- Organised advocacy and activism in civil society
- Critical media and journalistic investigation
- Spontaneous public concern (amplified in social media)
- Internal whistleblowers

Accountability, Answerability and Public Trust

Accountability, Answerability and Public Trust

Trust requires the security provided by accountability for **power,** where that power endangers **specific vulnerabilities and interests**:

- **Retrospective (backward-looking) accountability** (the person or agency deploying this power can and will *answer* for an unjust harm done by that power to me or my community)
- **Prospective (forward-looking) accountability** (the person or agency has accepted and undertaken *specific duties of care* to protect my legitimate interests in this context)
- **Character accountability** (the person or agency has thus far been trustworthy with my interests in this context)

Accountability, Answerability and Public Trust

Where accountability has not yet bridged a trust gap, there are three options for securing or restoring trust:

- **Hard Constraints** (local or global prohibition of the power, or restriction of the power from acting in certain domains/conditions)
- **Robust Duties** (creation of new, enumerated duties of care in the exercise of the power, which are allocated to specific and appropriate persons or roles, and aligned with parallel liabilities for negligence, that will be executed by reliable mechanisms)
- **Strict Liability for Harm** (irrespective of performance of duties of care, undue harms will be answered by specific and appropriate sanctions of the responsibilised party)

Ethical Risks and Vulnerabilities

Ethical Risks and Vulnerabilities for AI Use Cases in the Public Sector

1. **Unpredictability/brittleness** of performance of AI/ML systems
2. **Unjust bias** in AI/ML applications inherited from systematic bias in historical data or inappropriate design decisions
3. **Opacity** of AI/ML decisions that are proprietary, challenging to interpret/justify, or hard to audit
4. **Speed, scale, distributed, and automated** implementation of AI that blocks ‘meaningful human control’ and can deskill human supervisors who may succumb to ‘automation bias’
5. **Distinctive vulnerabilities** of groups targeted for public sector use cases, whose autonomy, dignity, rights and well-being may be disregarded in order to attain key efficiencies or satisfy political aims

Barriers and Opportunities

Barriers to Accountable AI in the Public Sector

- **Underresourced public agencies** that cannot afford the expertise or staff time to identify or manage AI/ML risks appropriately
- **Optimism bias** creates and incentivizes cultures of ‘see no evil, speak no evil’ that count solely on legal compliance + noble ambitions
- **Technosolutionist** imperatives that seek to apply AI/ML where it is not needed or fit for purpose, displacing more robust solutions
- **Lack of technical skills in AI/ML** needed to craft appropriate, robust models and safeguards, which makes public agencies vulnerable to exploitation by shoddy third-party providers.
- **Fears of over-regulation** that stifles innovation and adoption (which failure to appropriately regulate or govern *also* stifles)
- **Inadequate channels for identifying, reporting and contesting harms**

Opportunities for Accountable AI in the Public Sector

- **Growing AI ethics resources in UK to guide public sector agencies** (from Alan Turing Institute, Ada Lovelace Institute, CDEI, others)
- **New training pipelines will increase availability of AI/ML ethics expertise** that can be employed or seconded by public agencies
- **Scotland's strong commitment to responsible AI** allows it to learn lessons now from public uses of AI/ML elsewhere that were less careful
- **Scotland's advantage in public trust** makes our responsible AI work more likely to land, *if* followed by the actions needed to match the words
- **Devolved agencies** can create new cultures of accountability and care in AI/ML deployments that provide a sound model for others

Thank You!

Questions?

Shannon Vallor
Baillie Gifford Professor of Ethics of Data and AI
The University of Edinburgh
Director, Centre for Technomoral Futures

Centre for Technomoral Futures

THE UNIVERSITY of EDINBURGH
Edinburgh Futures Institute

The Alan Turing Institute