

Transparency Fallacy: Insights from Healthcare Professionals on the Safe Use of Medical AI

TATIANA ARANOVICH

RITA MATULIONYTE

FARAH MAGRABI



ARC Grant: 'No' to Black Box: Towards Transparent and Safe AI in Healthcare

Agenda

1. Medical AI Transparency and
Methodology

2. Interview Results

3. Limitations to Medical AI
Transparency

4. Possible Policy Solutions

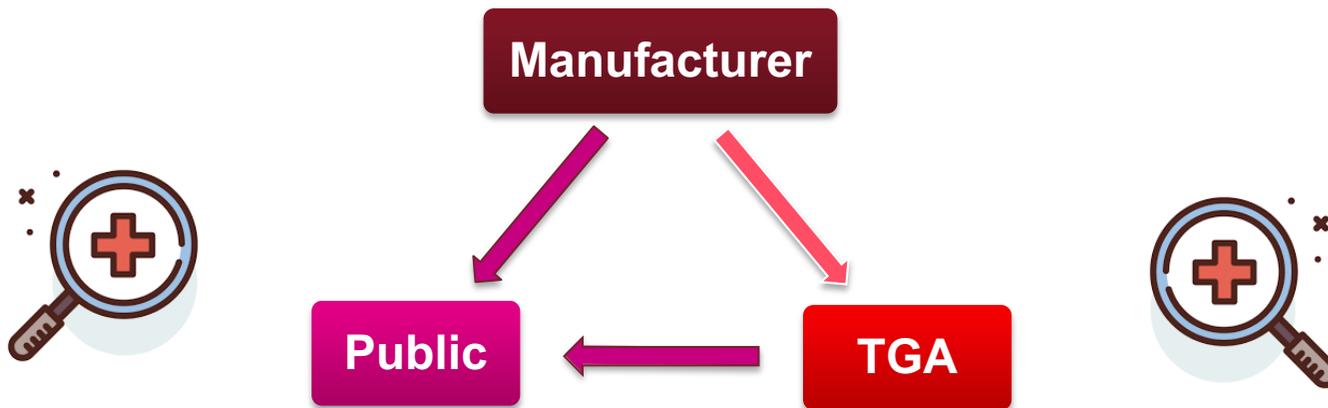


1. Medical AI Transparency

**AI
Transparency**

Information **ABOUT** the overall model

(**✗** NOT Explainability)



Medical AI

AI used for **medical purposes**
(even without approval)

1. Methodology

Semi-structured Interviews - Health Professionals

1. What information is needed **with the AI Technology (Labels / Instructions)** and in the **Pipeline** ?
2. What information should be provided to **Patients** ?
3. How to Avoid **Information Overload** ?

12 Interviewees AUS/NZ:

Anaesthetist

Cardiologist

Clinical Informatics Officer

Clinical Research Officer

Consumer Advocacy

Intensivist

Ophthalmologist

Paediatrist

Physician 1

Physician 2

Psychologist

Surgeon



2. Results: Info Relevant to Health Professional



Acknowledging AI usage

With the Technology
In the Pipeline

*' [We want to know if the clinician] was actually **aware of the result** or if it was **just spat out by a computer.**'*

Core Information

Intended Purposes
Intended Use
How to Use
Contraindications
Risk or Side Effects

*'[This first-tier of information] is what **99% of doctors need to know.**'*

Clinical Evidence and Validation and Datasets

Bias Mitigation
Performance (inc. Generalis)

*'If you are **comparing** two different **products** [the information about datasets and validation is important], but **once you have bought a product**, this information is **no longer important.**'*

Risk Management

Data (Privacy)
(Cyber)Security



2. Results: Info Patients Need to Get



Acknowledging AI usage

Risk and Side Effects
(Informed Consent)

Privacy Protection
(Terms of Use)

*'[Patients] are being **pushed to the side**: [they should] know the **top [risks and side effects]** and the **prevalence** of them.'*

*'When we think about this **interface with the patient**, especially regarding the risks and side effects, then clinicians should **get all the information** related to it and **philtre** this information to the patients.'*

3. Limitations to AI Transparency Rationales: Information to Clinicians

Safety

- **Clinicians rarely read Labels**
- Literacy and Info Overload

Accountability & Liability

- Deterrence
 - Eg. harms given automation bias
- Evidence

Trust

- Medical AI Novelty, Opacity and Uncertainty

*'Look at the **instructions** I have [show a lengthy manual/ instructions]. Do you think that [clinicians] are reading this? **No one is reading this.**'*

*'We are **drowning** - it is hard to get to the end of the day. [We do not have time to read labels].'*

3. Limitations to AI Transparency Rationales: Information to Patients

Safety

- Limited Patient's Sharing Data Mechanisms

Informed Consent

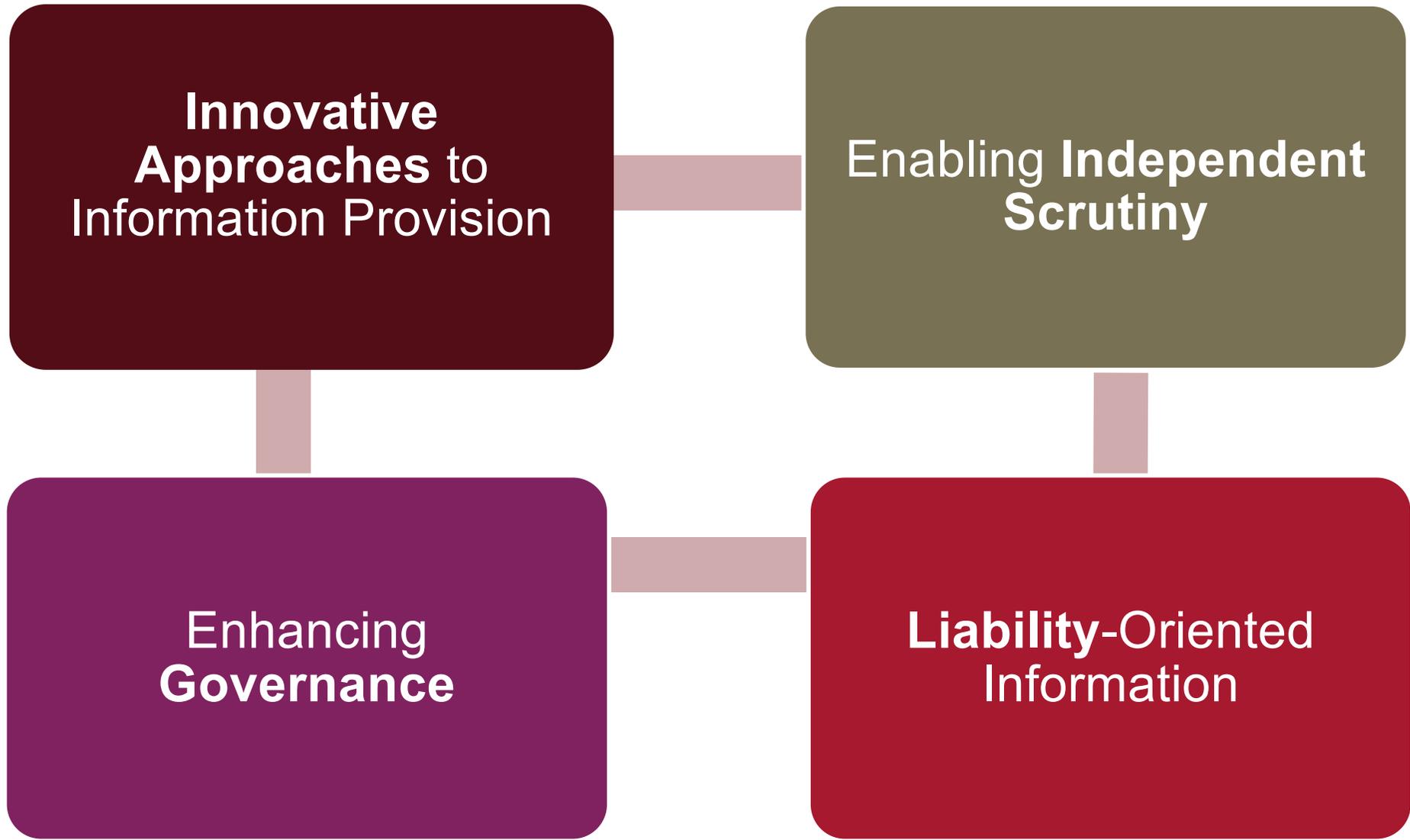
- Limited Patients and Clinicians' Literacy

Privacy

- Information Overload Risk

*'If I was **not confident** about my ability to **answer patients' questions** (...), saying 'I do not even want to let people know I am using it for **fear** that it will **overwhelm me** when they **ask a million questions**', I would just not use it.'*

4. Possible Policy Proposals



4.1 Innovative Approaches to Information Provision

- **Layered and Friendly**
 - In Different Levels
 - Language
 - Nudging to be Read + Understood
 - Eg. Model Cards

*'[Transparency] has to be also **smoothly and smoother.** Information must be much **easier to understand.**'*



4.2 Enabling Independent Scrutiny

- **External Expert Assessment as Key**
- **But requires Higher Transparency to the Public**
 - Eg. Datasets, Clinical Evidence, Validation

*'The [clinician] is **not** in a **reasonable position to understand** any of that. Then, it is really important that we have **experts to assess those things**. We will **never get rid of all risks**. It comes down to defining what is **acceptable risk**.'*



4.3 Enhancing Governance

- **Medical Associations**
 - Guidelines
 - Committees

*'Medical associations are taking **risk-adverse approach** - they **do not want to sign off any orientation or guidance** on these technologies.'*

- **Hospitals and Clinics**
 - Procurement Team
 - Guardrails

*'It is **counterintuitive** to be **spending hours combing through the fine print**. This is why it was helpful to have (...) a **procurement person** from the clinic to **check all information necessary**.'*

- **Medical Training**
 - De-skilling and Automation Bias
 - Junior Professionals

*'Clinicians will trust that the **machine is going to get it right**. (...) We have a problem especially regarding **junior clinicians** - we need to **teach them how to do medicine**.'*

4.4 Liability-Oriented Information

- **Information Available in a Hierarchical Way**
 - According to Risk
- **Assertive Language**
- **Producing a Chain of Evidence**
 - Disclosing AI Use

*'It is **hard** to **prove harm**. (...) In all the clinical incidents that we recorded in our organisation, there were multitude of causal factors. Someone has to do **something really terribly negligent** and a **departure from expected practice** to have that consequence'.*





tatiana.aranovich@hdr.mq.edu.au

rita.matulionyte@mq.edu.au

farah.magrabi@mq.edu.au



www.linkedin.com/in/tatiana-aranovich-848a16188/

www.linkedin.com/in/rita-matulionyte-29856b105/

www.linkedin.com/in/farah-magrabi/