#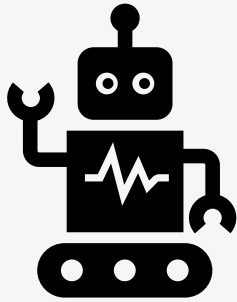 *Introducing a trauma informed assessment tool for evaluators of AI (artificial intelligence) assisted programs and services.*

Dr Lyndal Sleep

Senior Lecturer

Queensland Centre for Domestic and Family Violence research

Central Queensland University

- AI and other forms of digital technologies are being increasingly used in programs and services delivery.

- They promise increased efficiency, accuracy and objectivity, however these technologies can also cause significant harm and trauma, as seen in Robodebt.

- It is vital for evaluators to be aware of key questions to ask to prevent AI having unintended impact on program processes, outputs and outcomes, or cause harm to service users.

Core argument

(1) AI is increasingly being used in programs and services, and understanding the resulting risks is essential for evaluators to assess whether services are meeting intended outcomes.

(2) many evaluators are unaware of what types of risks to look for when assessing AI assisted services, or what questions to ask – especially when conducting trauma informed evaluations.

(3) To address this need, researchers from the University of Queensland and Central Queensland University have developed a practical trauma-informed assessment tool for AI-assisted services, which will be briefly introduced. The tool was funded by Notre Dame University IBM Ethics Lab.

Outline

Highlight the **problem** that AI is increasingly being used to assist program and services delivery, but many evaluators are unaware of the main risks to consider when evaluating these services.

Suggest the **solution** of a practical tool which considers these risks, with technological knowledge and within a trauma informed framework, that can be employed by evaluators.

Introduce a **trauma informed AI assessment tool**, the method used to develop it, as well as its intended practical use by evaluators (both internal and external to organisations).
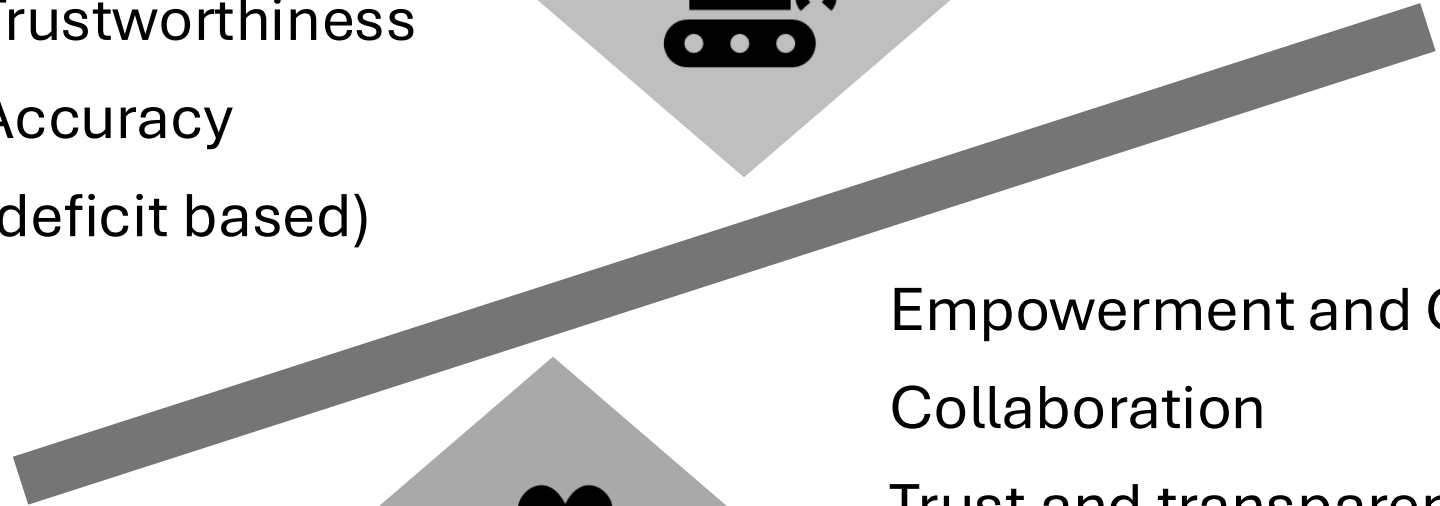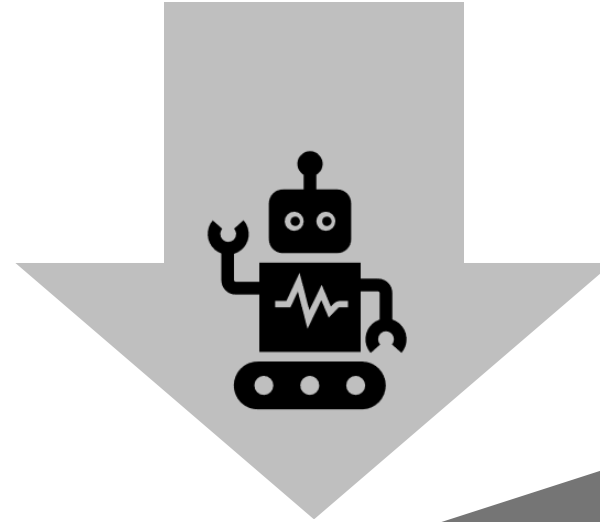
AI is increasingly being used to assist program and service delivery, but evaluators may be unaware of the main risks to consider when evaluating these services.

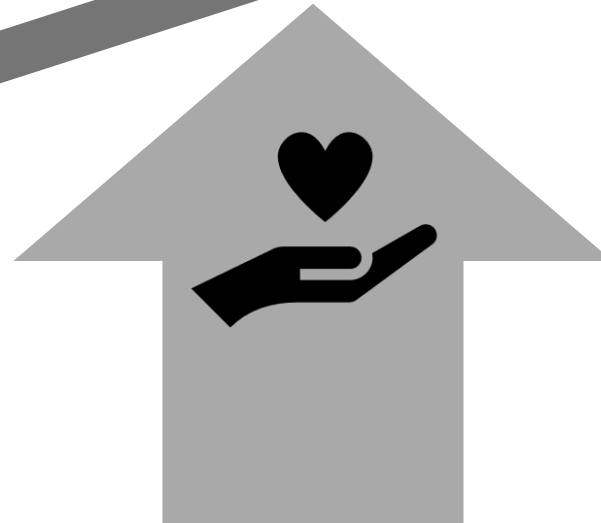Examples of **problems** and harms ....
- Robodebt
- Aleghany County
- Tessa (will come back to this example later)
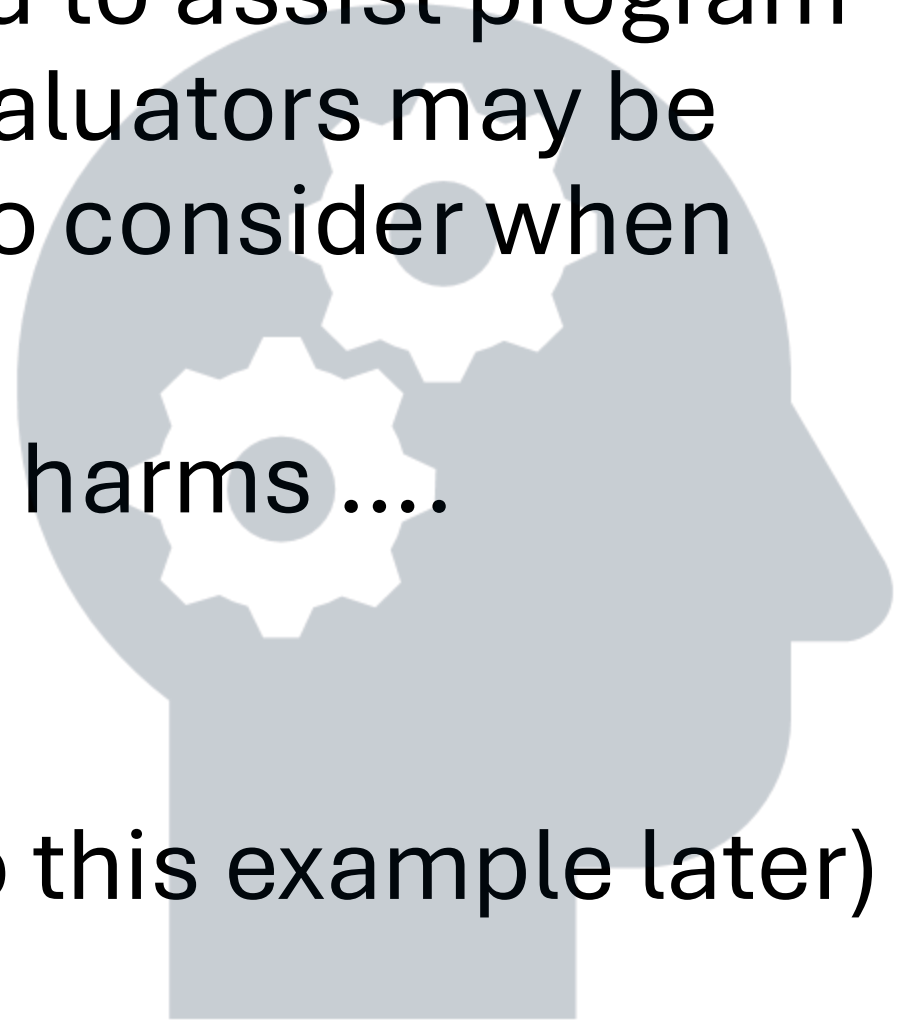
Bias

Fairness

Explainability

Trustworthiness

Accuracy

(deficit based)

Empowerment and Choice

Collaboration

Trust and transparency

Safety

Intersectionality
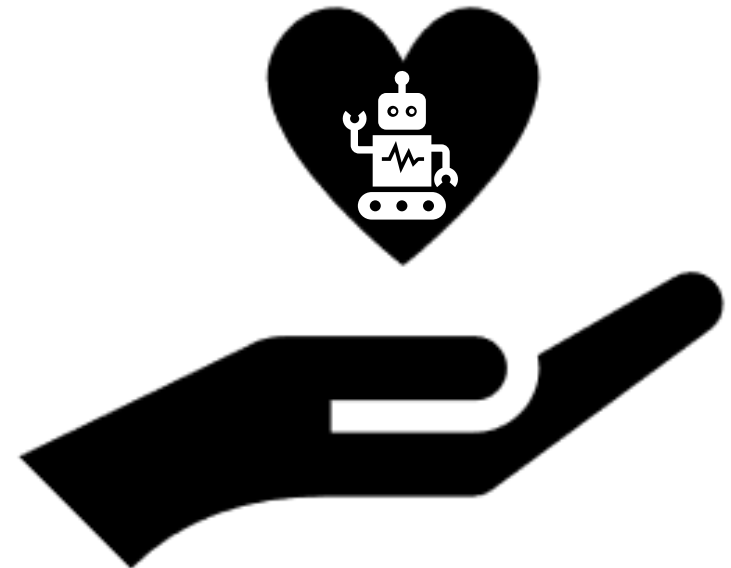
(needs and strengths based)

Suggest the **solution** of a practical tool which considers these risks, with technological knowledge and within a trauma informed framework, that can be employed by evaluators.

The tool aims to:

• Inform

• Instigate knowledgeable reflection

• Provide guidance

Introduce a **trauma informed AI assessment tool**

The tool was developed at the University of Queensland by Professor Paul Henman, Associate Professor Suzanna Fay, Dr Lyndal Sleep and Suvradip Maitre with funding from the IBM Data Ethics Lab in the University of Notre Dame, United States.
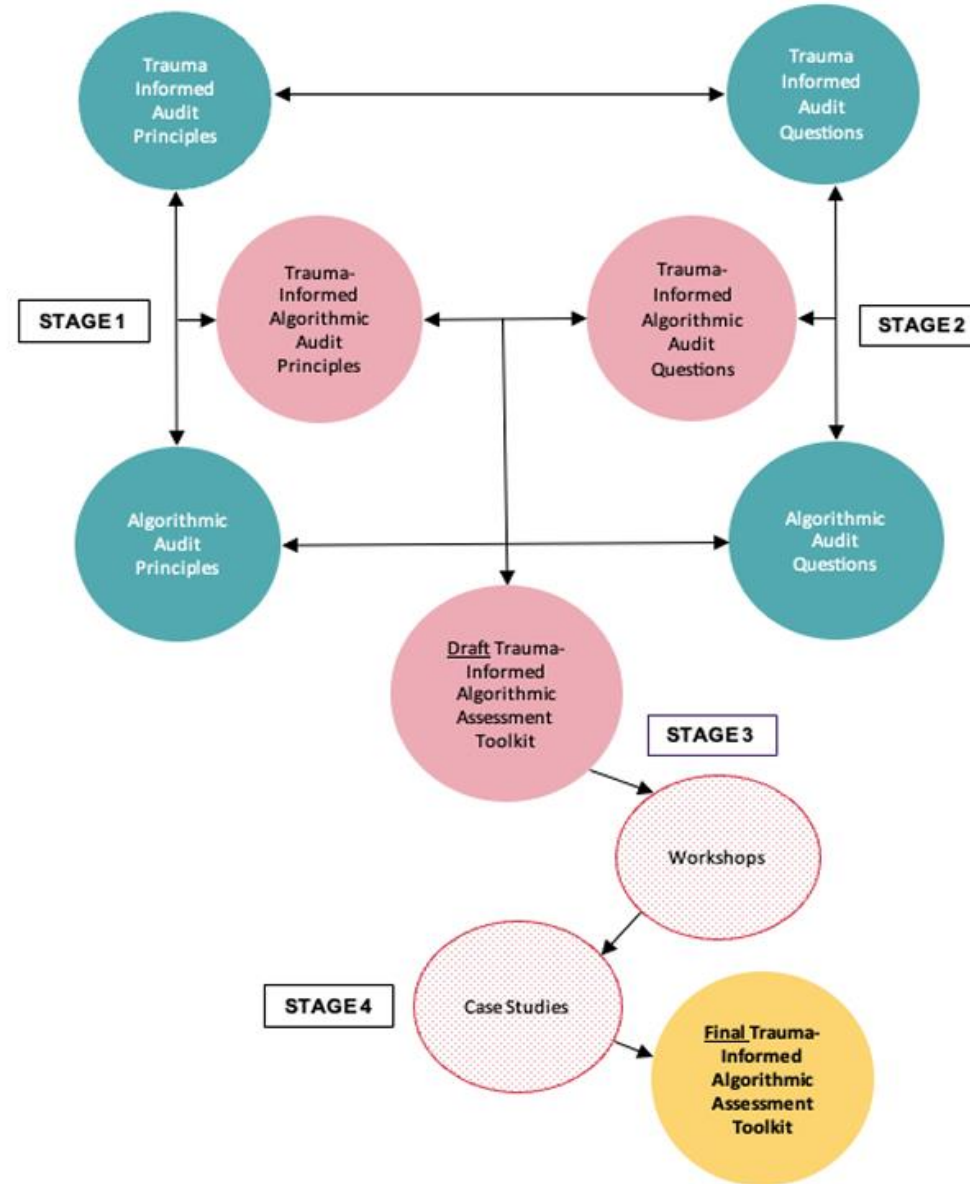
A report explaining the development of the tool, as well as a first version of the tool is available at https://www.admscentre.org.au/trauma-informed-ai/

SCAN ME

Research Design stages for developing the tool

The tool

The tool is structured into 3 main sections, allowing multiple ways to apply the tool.

(1) a short assessment of 10 pertinent questions is provided,

(2) a domain-focused assessment, and

(3) a technology-focused assessment.

1. CRITICAL CONSIDERATIONS

The following section prompts you to reflect on critical considerations that are applicable to all algorithm supported services. You should ensure that you have thoroughly reflected and engaged with the below considerations. The questions in this section represent the absolute minimum when you are implementing a trauma informed approach within your algorithm supported service.

#	For the algorithm supported service, to what extent…	Not at all	To some extent	To a great extent	Evidence	Action/Plan	Priority	Principle
1.	are service users aware that an algorithmic system is being used?	☐	☐	☐			☐	Trust and Transparency
2.	have designers considered if it could cause or trigger trauma?	☐	☐	☐			☐	Safety
3.	has the psychological, emotion and cultural safety of service users been considered?	☐	☐	☐				
4.	does the service user have an informed choice about when their personal information is accessed or shared?	☐	☐	☐				
5.	can the service user choose to interact with a human?	☐	☐	☐				
6.	can service users regularly provide feedback?	☐	☐	☐			☐	Collaboration Choice
7.	can service users choose to make a complaint, appeal or review directly to a human?	☐	☐	☐			☐	Trust and Transparency Empowerment and Choice
8.	could it unfairly disadvantage different service user groups?	☐	☐	☐			☐	Intersectionality Safety
9.	could the algorithmic system harmfully discriminate against any service user?	☐	☐	☐			☐	Intersectionality Safety
10.	are there supportive processes for responding to harm or distress caused by the algorithmic system?	☐	☐	☐			☐	Safety Trust and Transparency

Intended use by evaluators

- To assist evaluation of AI-supported services in general
- To assist with trauma informed assessment of AI supported services
- Tool will be freely available for no cost

Next steps

- Currently undertaking a follow on project with Australian Research Council Automated Decision Making and Society Centre of Excellence researchers, developing an online interactive version of the tool and evaluating it in situ.

- Please contact researchers if you would like more information on the tool, or to use it in your evaluation.

- We are interested in hearing from you if you would like to take part in an in-situ evaluation.
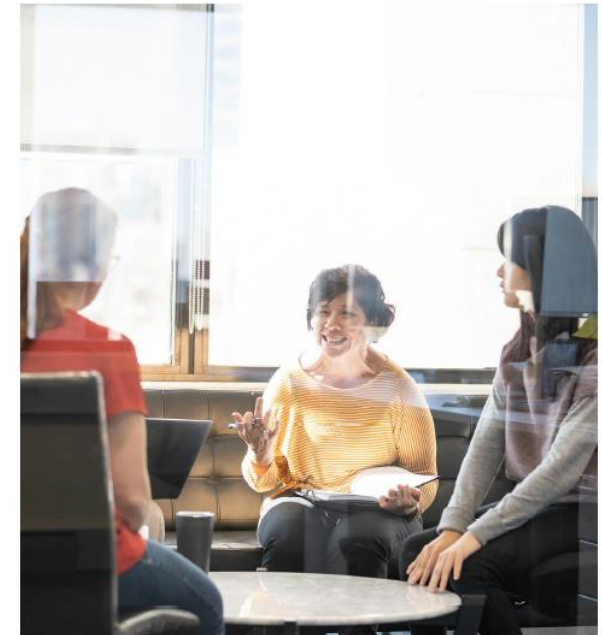
Thank you!

- L.sleep@cqu.edu.au

Access the first generation version of the tool at:

https://www.admscentre.org.au/trauma-informed-ai/

SCAN ME

Building a Trauma-Informed Algorithmic Assessment Toolkit

Suvradip Maitra, Lyndal Sleep, Suzanna Fay & Paul Henman

ADM+S
ARC Centre of Excellence for Automated Decision-Making and Society

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA
CREATE CHANGE