MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

@docsunny100

Comprehensive evaluation framework for artificial intelligence in healthcare and an example of its application

Dr Sandeep Reddy

Director, Healthcare Management

Deakin School of Medicine

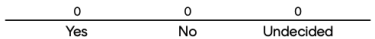
AIDH IMIA

Outline

- Why evaluation of AI in healthcare?
- Current Approaches
- TEHAI and Components
- Application and Findings
- Discussion

Go to **menti.com** and use the code **1559 9056**

Do we need evaluation of AI in Healthcare



SR
Account
Content
Design

Why Evaluate AI in Healthcare

> Progress in artificial intelligence (AI) has opened new opportunities

> However, in limited assessments that have taken place so far, it has been found AI systems have fallen short of their translational goals

> This is because many AI systems have intrinsic inadeq
that don't get assessed until after deployment

Context


HOT TOPIC
BUZZ
FOX BUSINESS
6:49A ET
DOW FUT 33,556.00 183.00 -0.54%
S&P FUT 4,069.00 30.75 -0.75%
INDEX FUTURES
S&P 500 FUTURES (ES/H3) 4,067.75 32.00

Context

- Utilising and integrating AI systems in clinical settings can be potentially expensive and disruptive

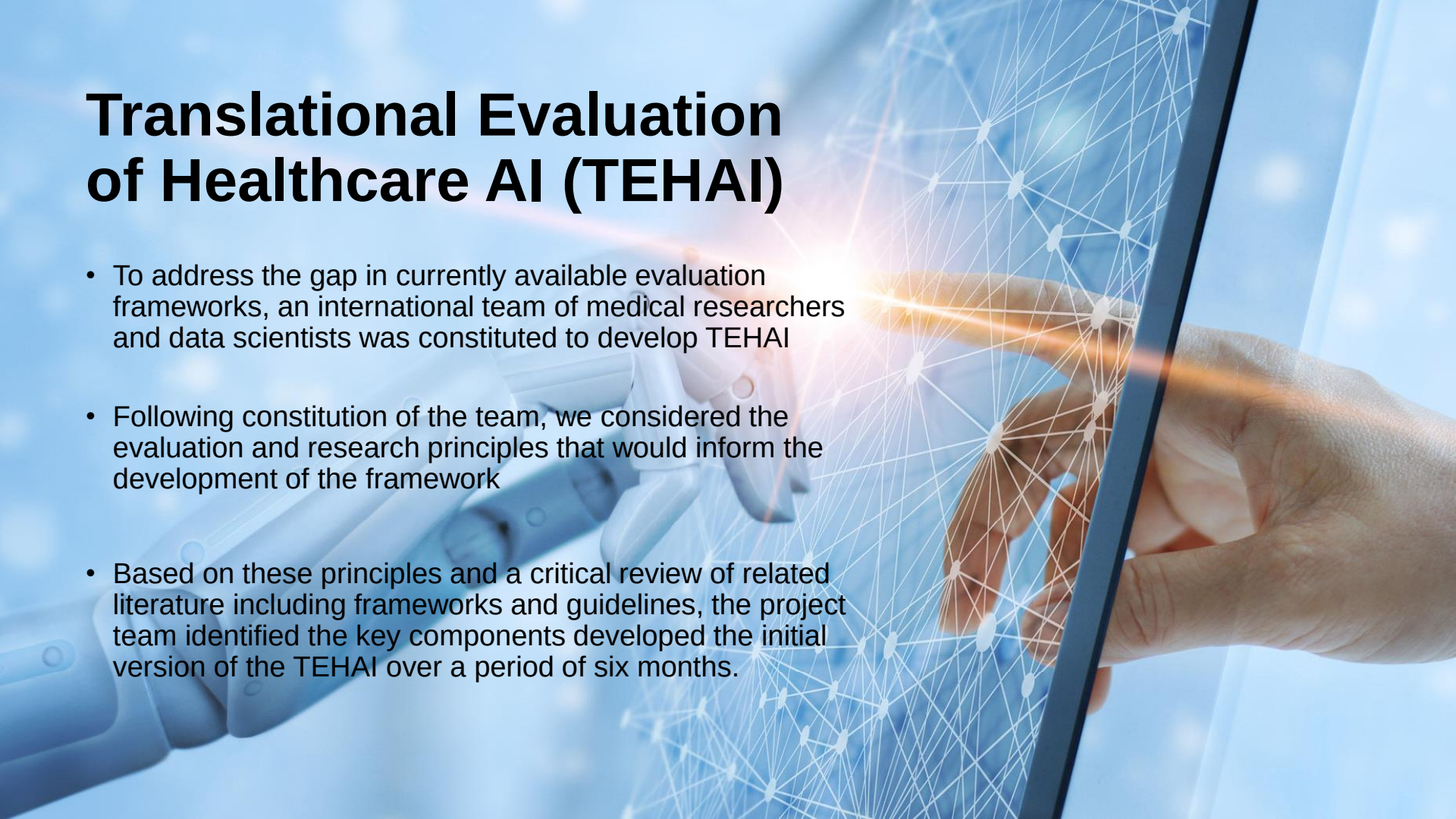
- Therefore, a rigorous evaluation that assesses AI systems early and at various stages of their deployment can support or contradict the use of a specific AI tool

Current Evaluation Approaches

- Currently available evaluation frameworks generally focus on the reporting and regulatory aspects
- It is evident there is an absence of an evaluation framework that assesses various stages of development, deployment, integration and adoption of AI systems
- Dependence on disparate evaluation frameworks to assess different aspects and phases of AI systems is unrealistic
- Also, currently available evaluation and reporting frameworks fall short in adequately assessing the functional, utility, and ethical aspects of the models

Translational Evaluation of Healthcare AI (TEHAI)

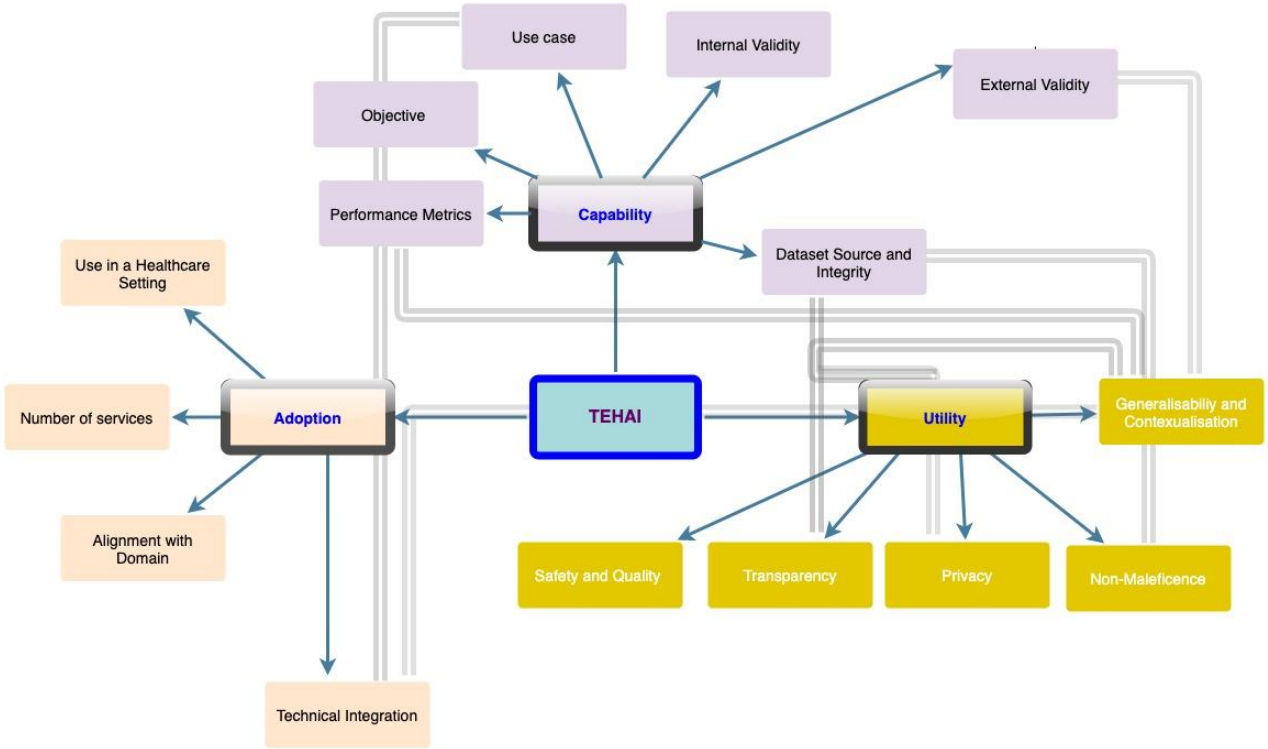
- To address the gap in currently available evaluation frameworks, an international team of medical researchers and data scientists was constituted to develop TEHAI
- Following constitution of the team, we considered the evaluation and research principles that would inform the development of the framework
- Based on these principles and a critical review of related literature including frameworks and guidelines, the project team identified the key components developed the initial version of the TEHAI over a period of six months.



Translational Evaluation of Healthcare AI (TEHAI)

- To provide a layer of independent review before finalization of TEHAI, the draft consensus framework was then reviewed by an international panel
- The eight-member international panel had expertise in medicine, data science, healthcare policy, biomedical research and healthcare commissioning, and were drawn from the United Kingdom, United States of America and New Zealand
- The panel members were provided the framework and documentation and after, meetings were convened with panel members to receive their feedback.
- Following collation of the feedback from the expert panel, TEHAI was refined to incorporate panel members feedback and was then finalised

TEHAI


Use case
Internal Validity
External Validity
Objective
Performance Metrics
Capability
Dataset Source and Integrity
Use in a Healthcare Setting
Number of services
Adoption
TEHAI
Utility
Generalisability and Contexualisation
Alignment with Domain
Safety and Quality
Transparency
Privacy
Non-Maleficence
Technical Integration

TEHAI

1. Capability

- 1.1.Objective
- 1.2. Use Case
- 1.3. Dataset Source and Integrity
- 1.4. Performance Metrics
- 1.5. Internal Validity
- 1.6. External Validity

2. Utility

- 2.1. Generalisability and Contextualisation
- 2.2. Safety and Quality
- 2.3. Transparency
- 2.4. Privacy
- 2.5. Non-Maleficence

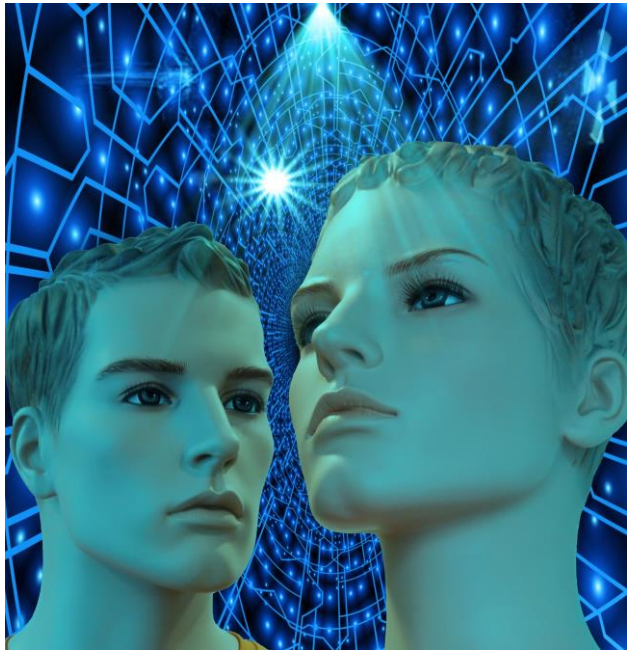
3. Adoption

- 3.1. Use in a Healthcare Setting
- 3.2. Number of Services
- 3.3. Alignment with Domain
- 3.4. Technical Integration

Capability: This component assesses the intrinsic technical capability of the AI system to perform its expected purpose, by reviewing key aspects as to how the AI system was developed

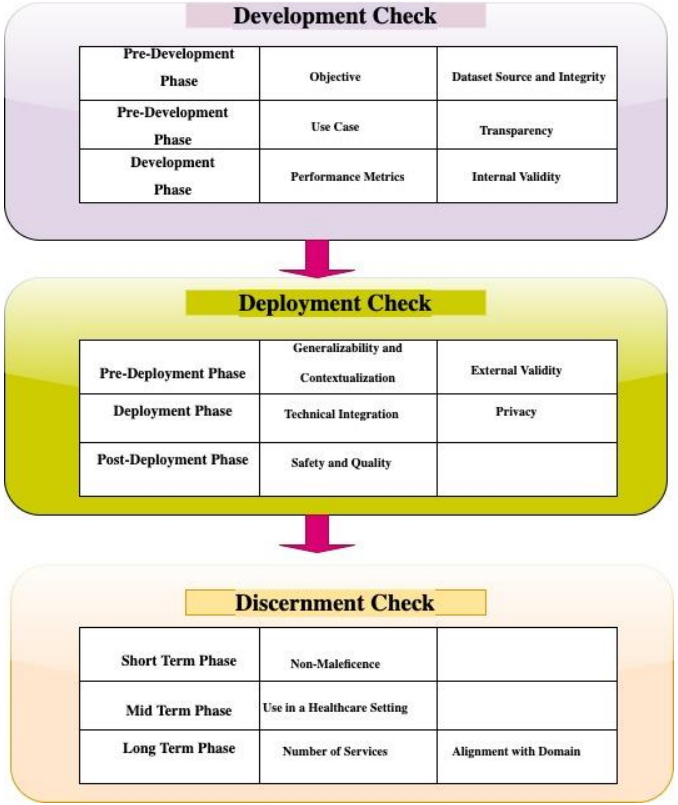
Utility: This component evaluates the usability of the AI system across different dimensions including the contextual relevance, and safety and ethical considerations. It also assesses the efficiency of the system

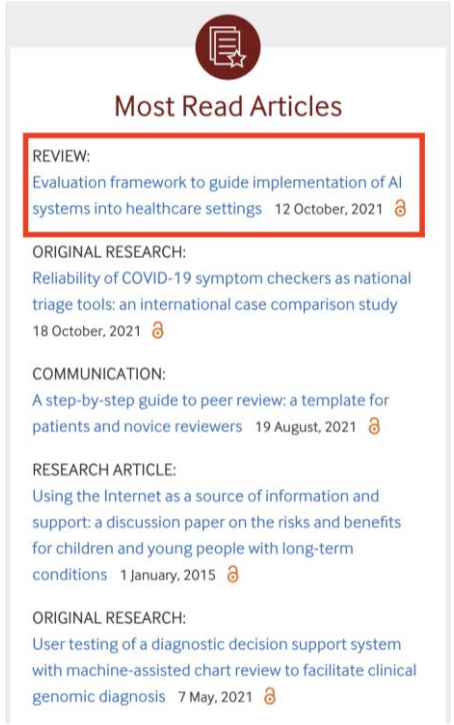
Adoption: This component appraises translational value by evaluating key elements that demonstrate the adoption of the model in real life settings



Component	Sub-component	Initial Score	Weight	Subcomponent Score= Initial Score x Weight	
Capability	*Objective of Study*	0-3	10	Weight 5	0-9 10-14 15 and above
	Dataset Source and Integrity		10		
	Internal Validity		10		
	External Validity		10		
	Performance Metrics		10		
	Use Case		5		
Utility	*Generalizability and Contextualisation*	0-3	10	Weight 10	0-19 20-29 30 and above
	Safety and Quality		10		
	Transparency		10		
	Privacy		10		
	Non-Maleficence		10		
Adoption	*Use in a Healthcare Setting*	0-3	10		
	Technical Integration		10		
	Number of Services		5		
	Alignment with Domain		5		

TEHAI usability





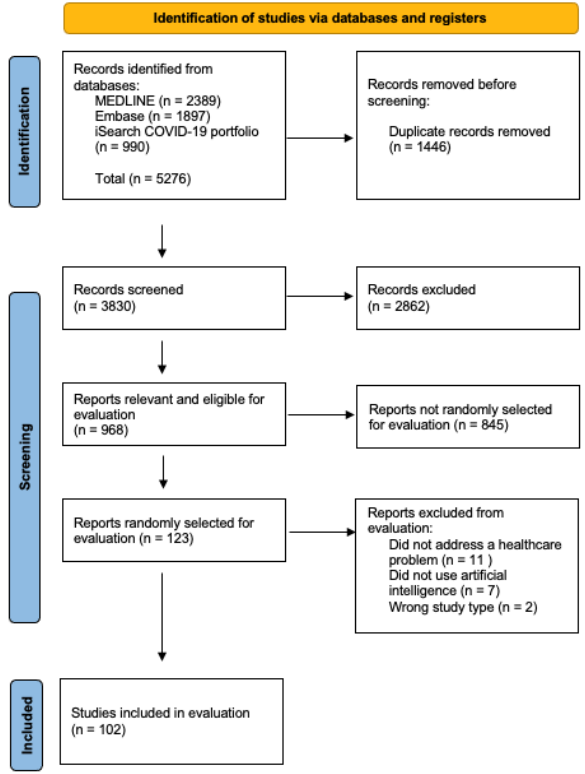
Reddy, S et al. (2021). Evaluation framework to guide implementation of AI systems into healthcare settings. *BMJ Health & Care Informatics* 2021;**28:**e100444.

Application

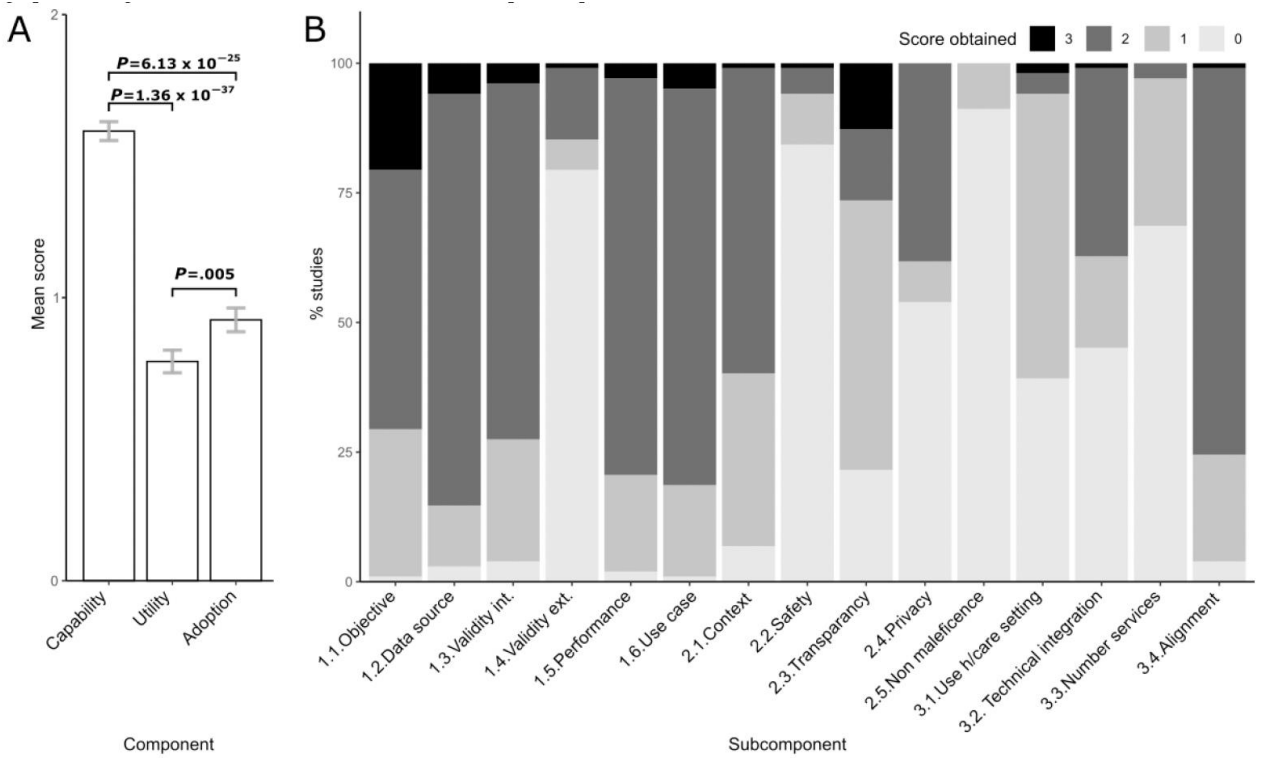
Application of a Comprehensive Evaluation Framework to COVID-19 Studies: Systematic Review of Translational Aspects of Artificial Intelligence in Health Care

Aaron Edward Casey[1,2], PhD; Saba Ansari[3], GCHE (Teaching and Learning), MSc; Bahareh Nakisa[4], BSE, MCS, PhD; Blair Kelly[5], Grad Dip (InfoLibStds), BCom; Pieta Brown[6], MPS; Paul Cooper[3], PhD; Imran Muhammad[3], MIS, MSc, PhD; Steven Livingstone[6], BSc, GradDipSci, MDataSci; Sandeep Reddy[3], MBBS, MSc, PhD; Ville-Petteri Makinen[1,2,7,8], DSc

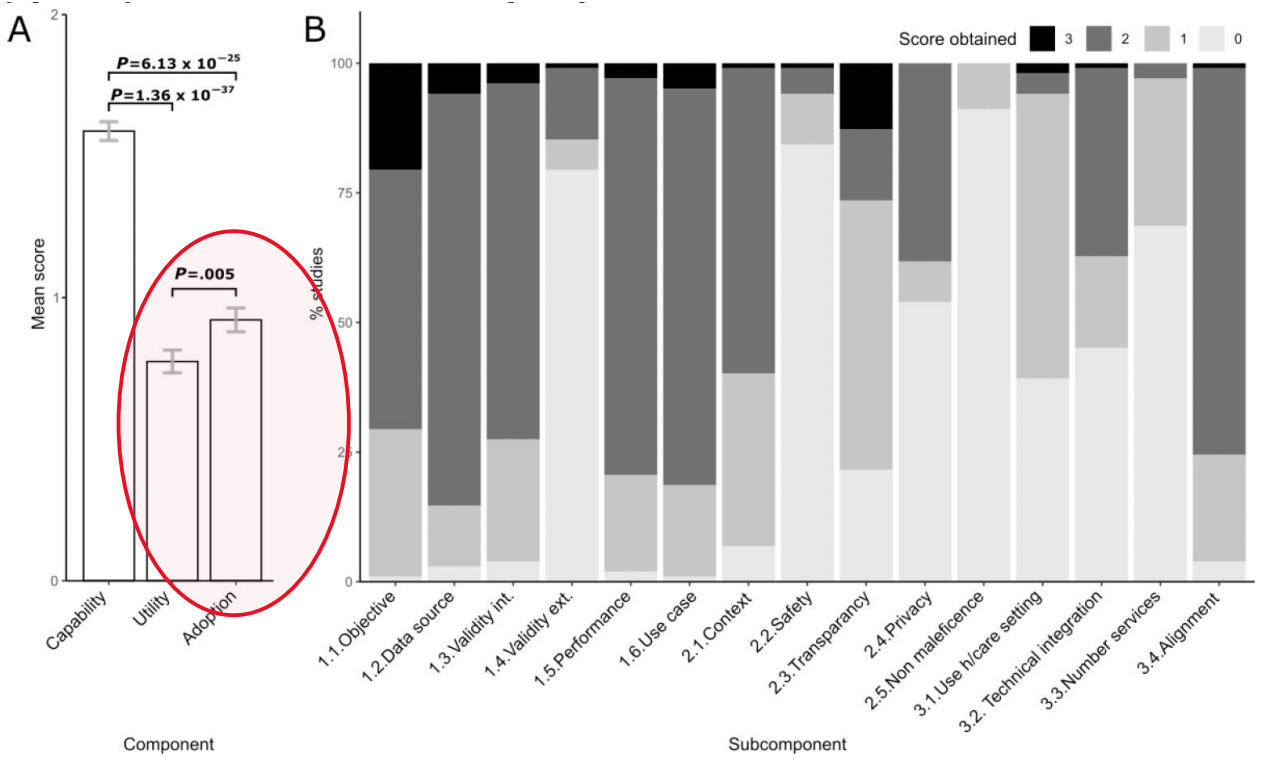
Application



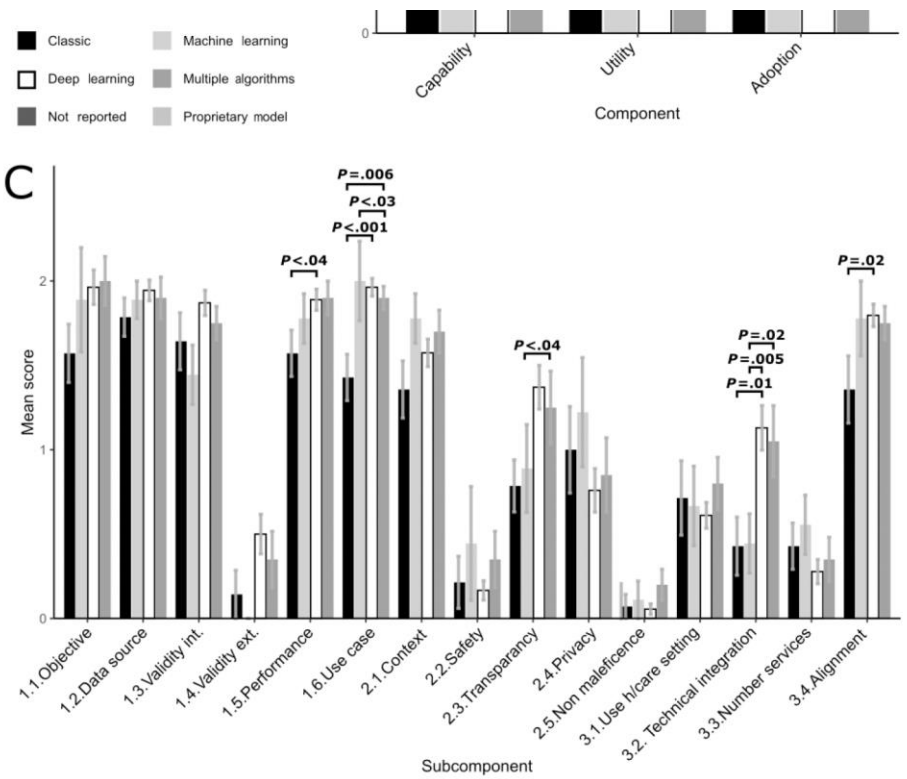
Application



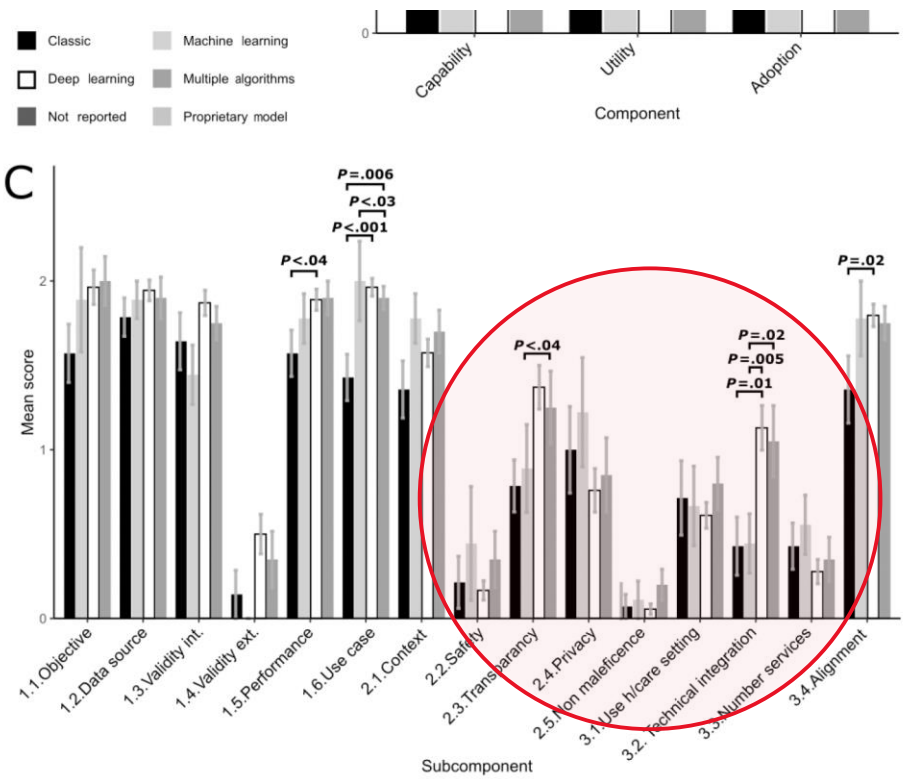
Application



Application



Application



Learning

- The ‘utility’ component assesses how safely can the AI model be used in healthcare

- Very few studies scored well across the criteria and scored poorly especially with the safety and quality and non-maleficence subcomponents

- One of the distinguishing aspects of TEHAI framework compared to other evaluation framework is its assessment of how well the AI model is adopted

- This is assessed through the actual use of the AI model in health services or healthcare delivery.

- Considering many of the COVID-AI models were experimental and the time frames we assessed were short, very few included studies did well in the component

- The top-ranking study in this component was an online COVID-19 mortality prediction model that was deployed as an open-source tool making it highly accessible and adoptable.

In conclusion

- TEHAI- A comprehensive evaluation framework
- Three main components (Capability, Utility and Adoption) and 15 subcomponents
- Can be used in development, deployment and discernment stages
- Applied to COVID-19 AI studies
- Very few studies have a translational component i.e., did poorly in utility and adoption components
- Therefore, evaluation has to in-built in product/application development cycle

Discussion/Questions