

MANCH

The University of

NING

Designing for Integrity at Scale: A Cross-Institutional AI-Driven Framework for Safe, Automated, Scalable and Student-Centred Assessment

Dr Homa Molavi, Dr Kamal Qazi

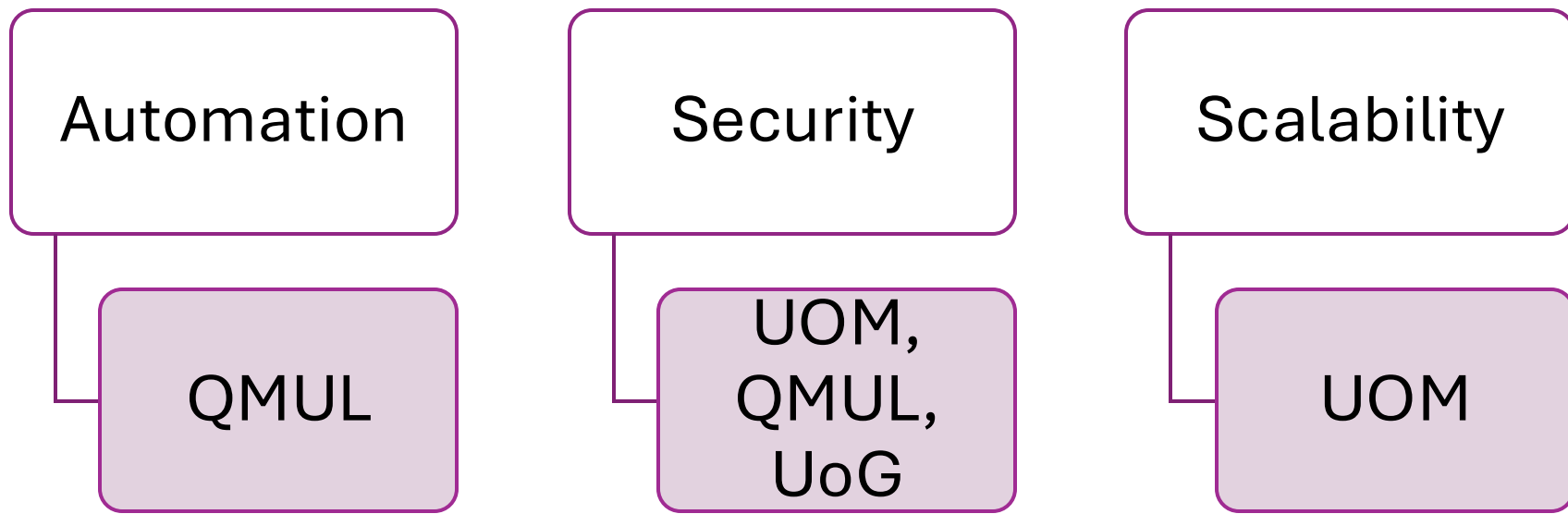
University Place, 176 Oxford

Manchester, M13 9PL

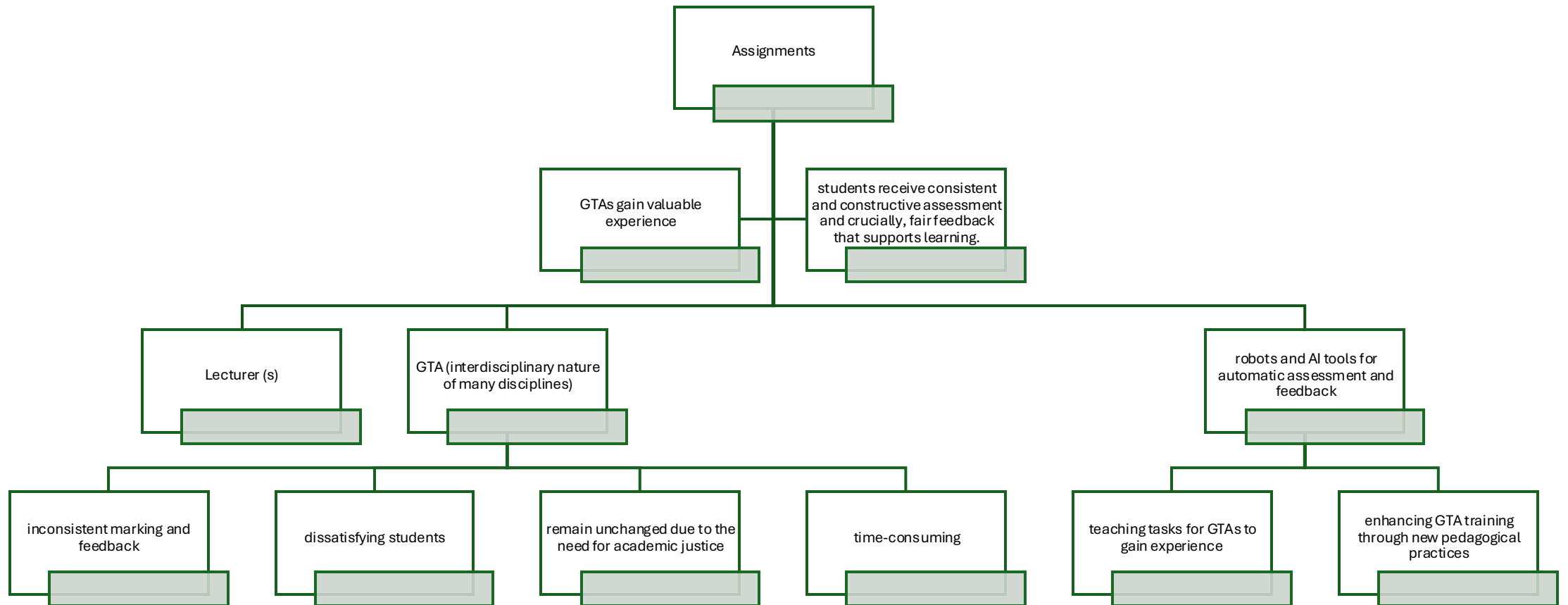
How Research Problems Emerge and Develop

- Our experience navigating the "ban AI vs. embrace AI" debate between UoM computer science teams
- Finding middle-ground solutions
- Building evidence-based consensus
- Leading cultural change through workshops and safe spaces for discussion
- Policy briefing

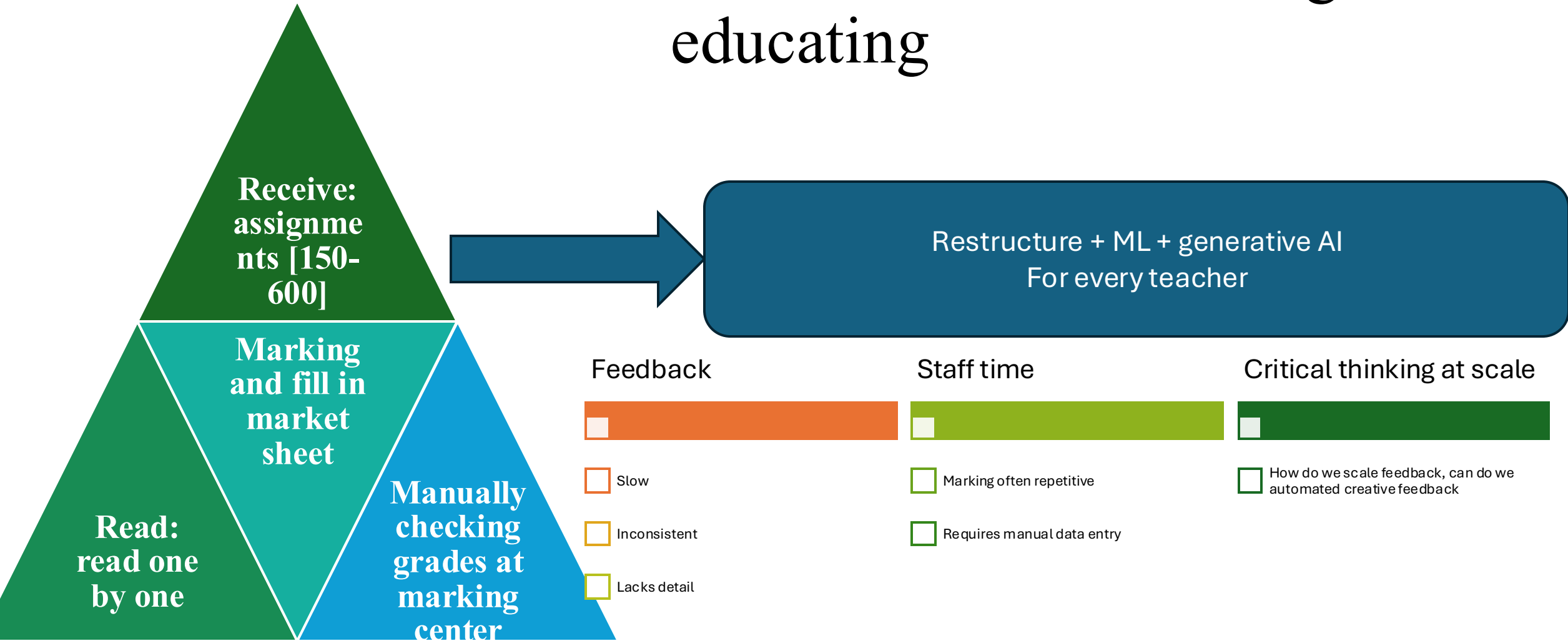




- Outsourcing thinking
- **Design AI-resistant, not AI-free, assessments**
- The assessment is no longer measuring **learning**, only **tool access**.
- AI excels at output. It does not struggle unless we test **process**.
- students will optimise for **efficiency**, not learning.



TA/TF roles: transition to AI-driven learning and educating



Story >

AI in education: ensuring ethical and human-centered integration

Over the course of an afternoon, three experts and more than 500 French-speaking educators gather to examine the role of AI in teaching practices.



Design a global framework

AI competency framework for teachers (AI CFT) UNESCO

Beyond Automation: Processual Assessment Integration Model (P-AI-M)

The AI Assessment Scale (AIAS)

How do we prepare teachers for the AI era?

What level of AI use is appropriate for this assignment?

Where and how should AI be integrated throughout the assessment process?

AI CFT (UNESCO) = Teacher AI competencies

P-AI-M = Student AI use in assessments

AIAS = AI integration in assessment processes

Research Question

How can institutions design AI-enabled assessment ecosystems that maintain integrity while scaling, automating assessment across diverse student populations safely?

Expert Delphi Panel (50 experts)

Round 1

Objective

Generate framework components.

Open-ended questions :

What threatens assessment integrity in AI-enabled education?

What safeguards should be mandatory?

Which assessment activities may be safely automated?

What role should students play?

What evidence is needed to demonstrate authentic learning?

Round 2

Experts rate each statement.

Scale: 1–7 Likert

Criteria:

- Importance
- Feasibility
- Relevance

Round 3

Experts receive:

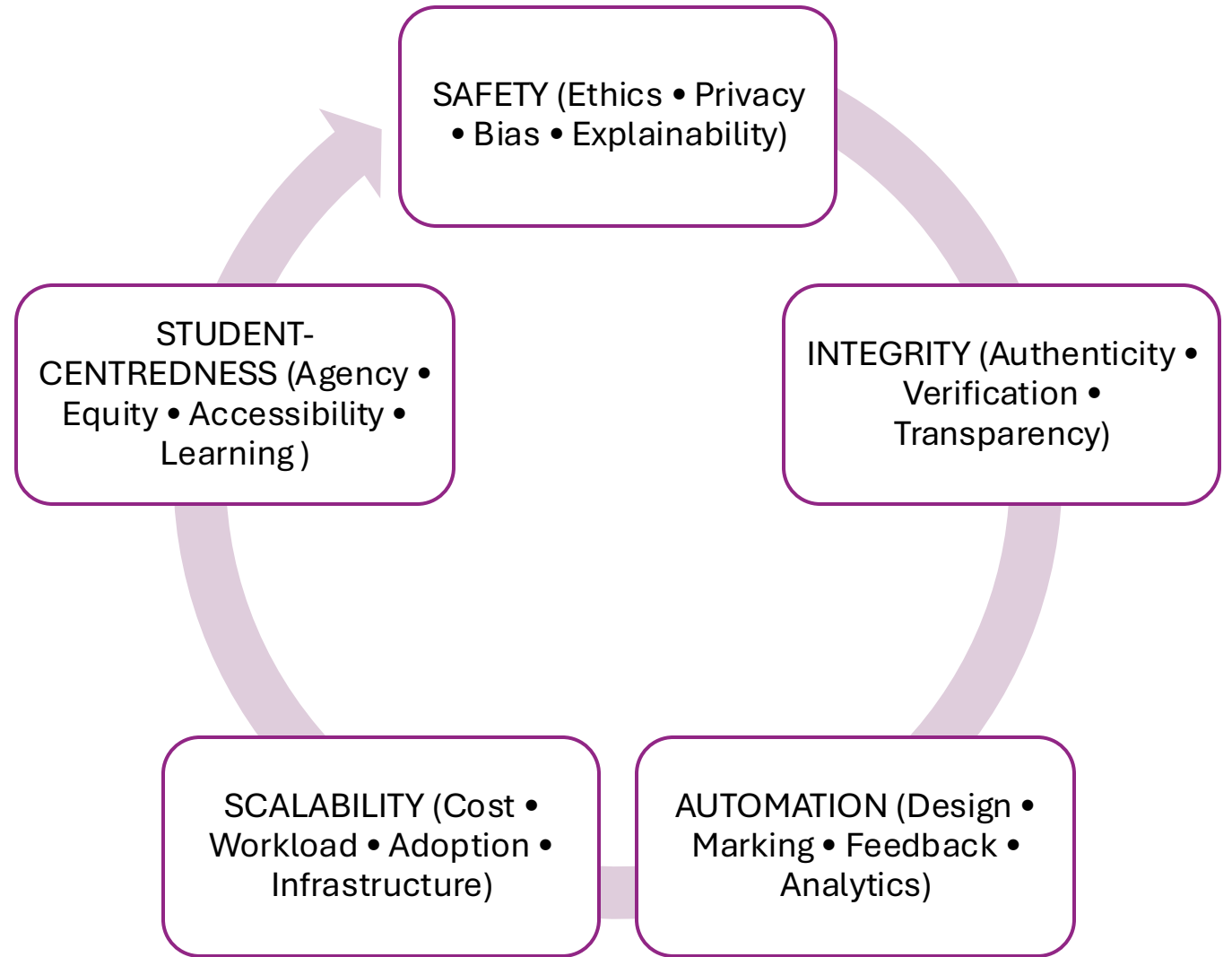
- Group median
- Their previous score

Experts revise scores.

This round typically stabilizes consensus.

outcome: 30–50 high-consensus principles.

DESIGNING FOR INTEGRITY AT SCALE (DIS)



Theoretical Contribution

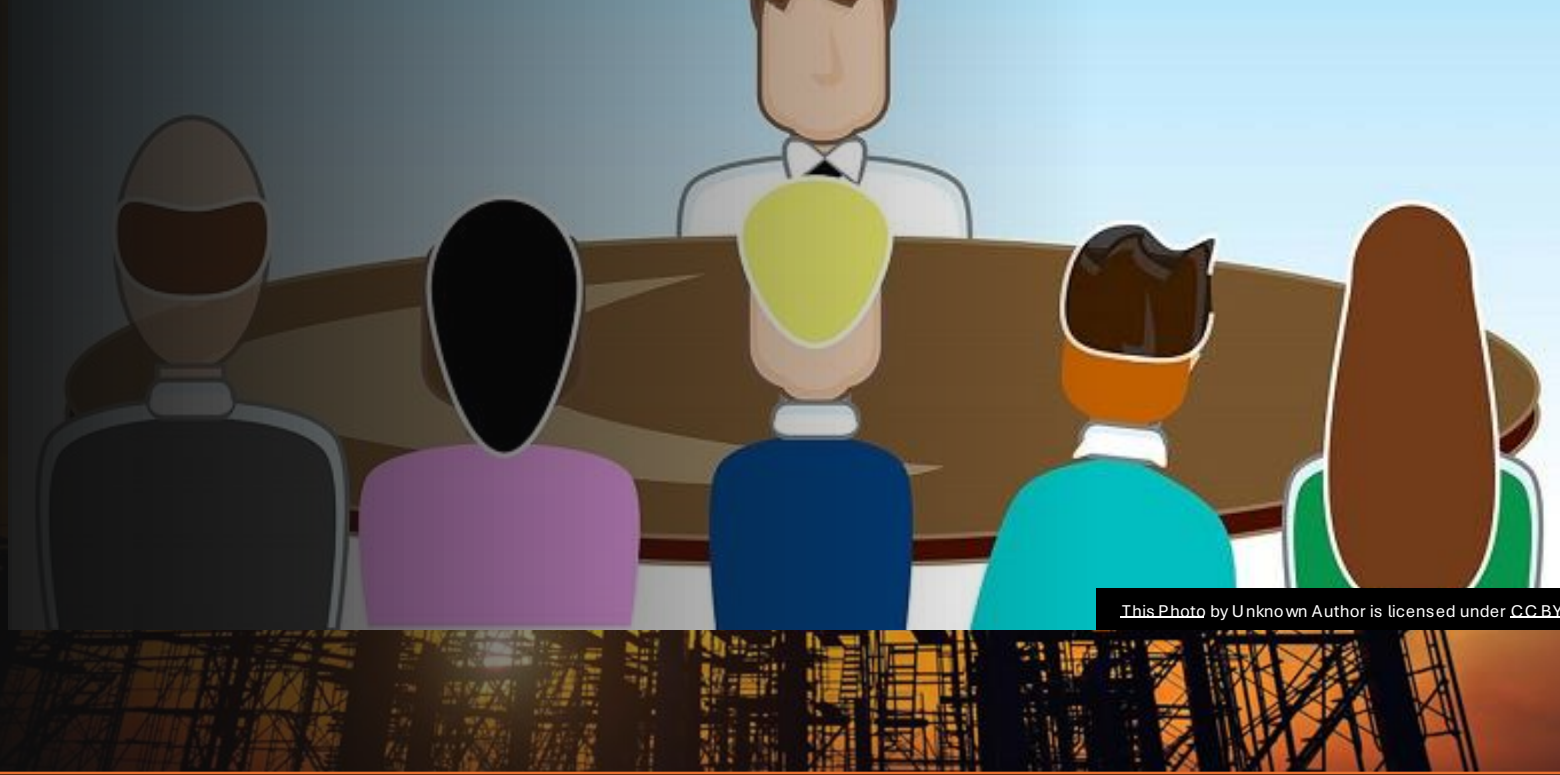
This moves the conversation from:

"Should AI be used in assessment?"

To

“How should higher education systems design assessment ecosystems that remain valid, trustworthy, scalable, and educationally meaningful when AI use becomes normal?”

That shift, from individual assessment tasks to cross-institutional assessment governance, is where the strongest theoretical contribution lies.



The work
continues....

Thank you for listening.

