



Shaping the Future of AI



UCLAIT - Team

Taico
Teacher-AI Complementarity

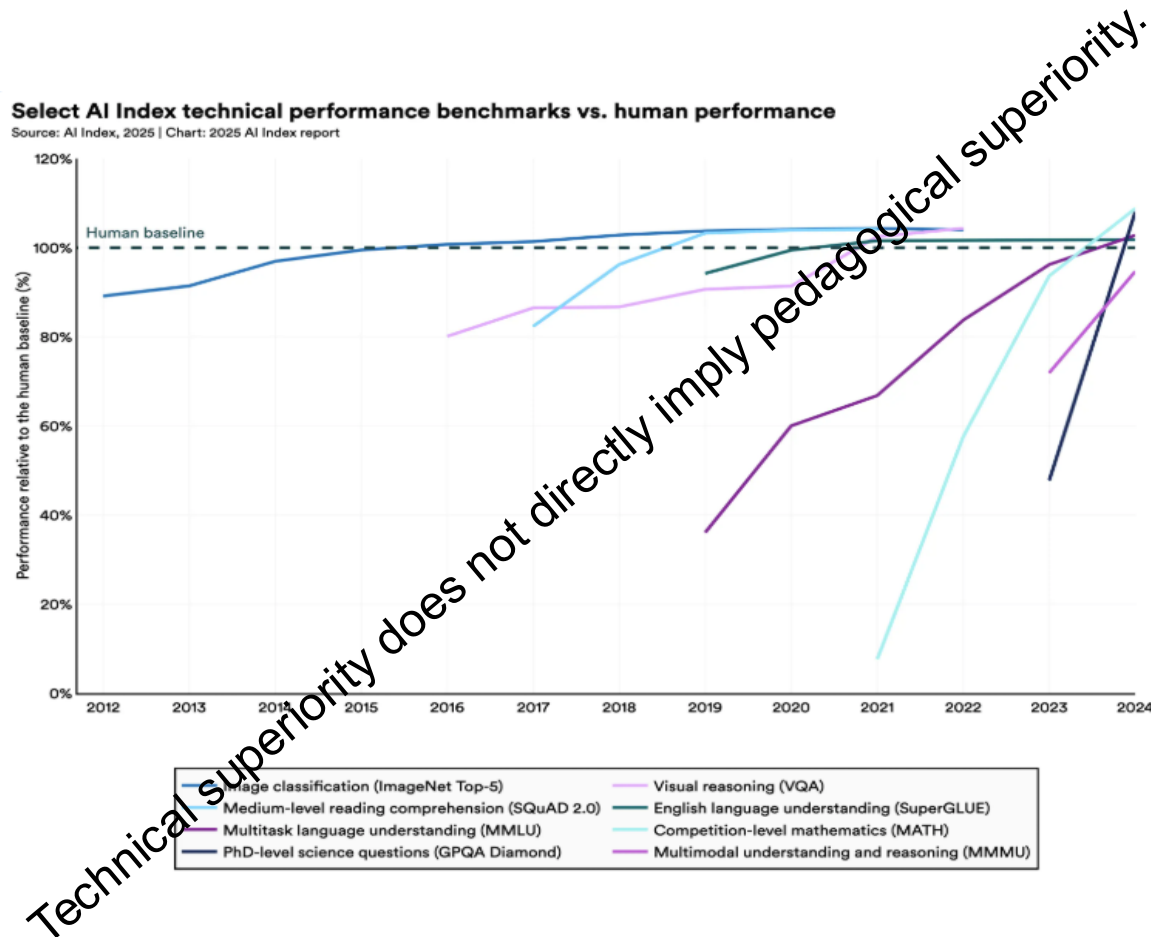


A Conceptual Framework for Teacher-AI Teaming in Education

Professor Mutlu Cukurova

UCL Knowledge Lab & UCL Centre for Artificial Intelligence,
University College London

AI Benchmarks are improving: But what are we actually measuring?



1) AI outperforms expert human baselines on increasingly complex benchmark tasks.

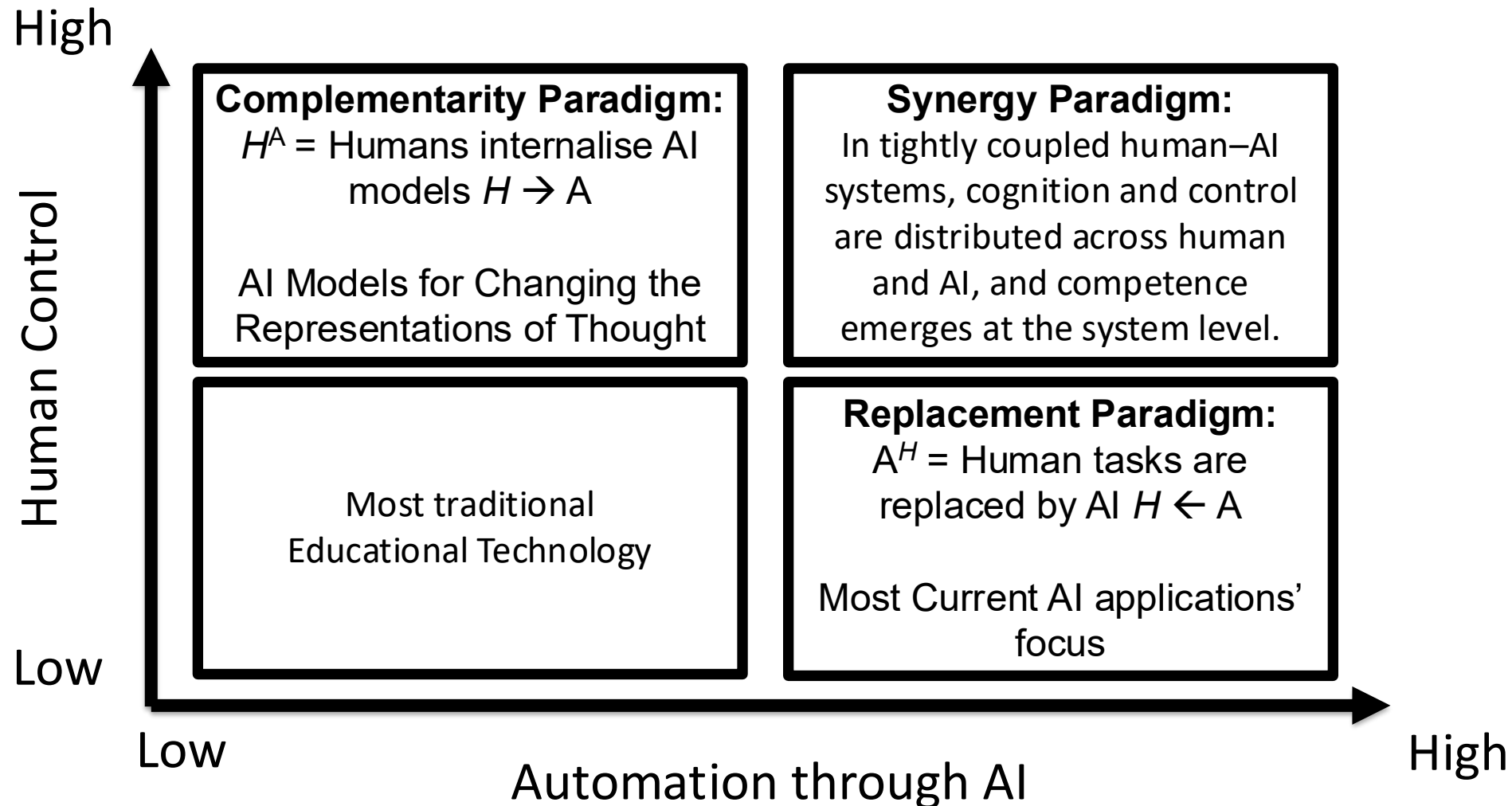
2) This does not tell us whether humans become more capable when working with them.

3) We need benchmarks that evaluate whether AI improves human capability over time.

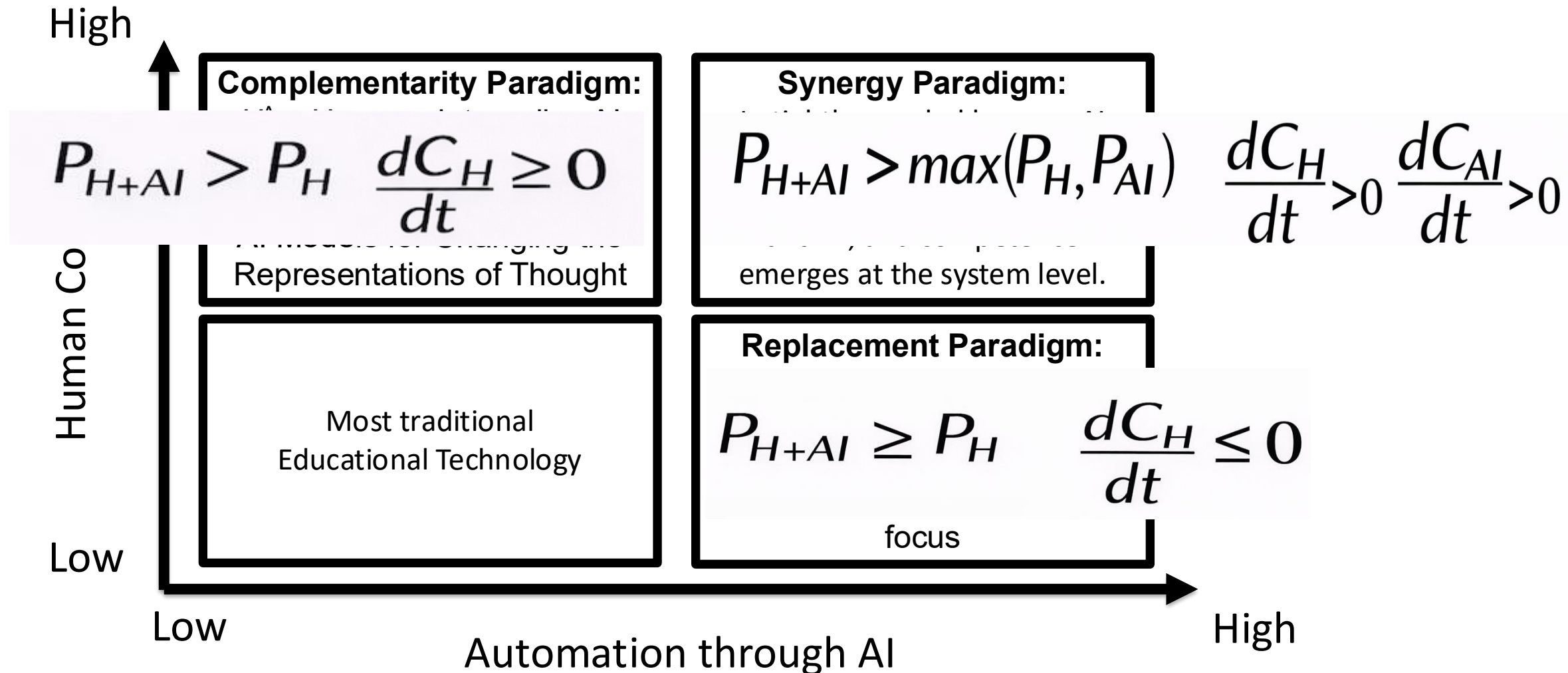
In AI, progress is often defined by automating tasks once performed by humans.

In AI in education, a paradigm shift is needed.
How do we design human–AI systems that increase human competence rather than displace it?

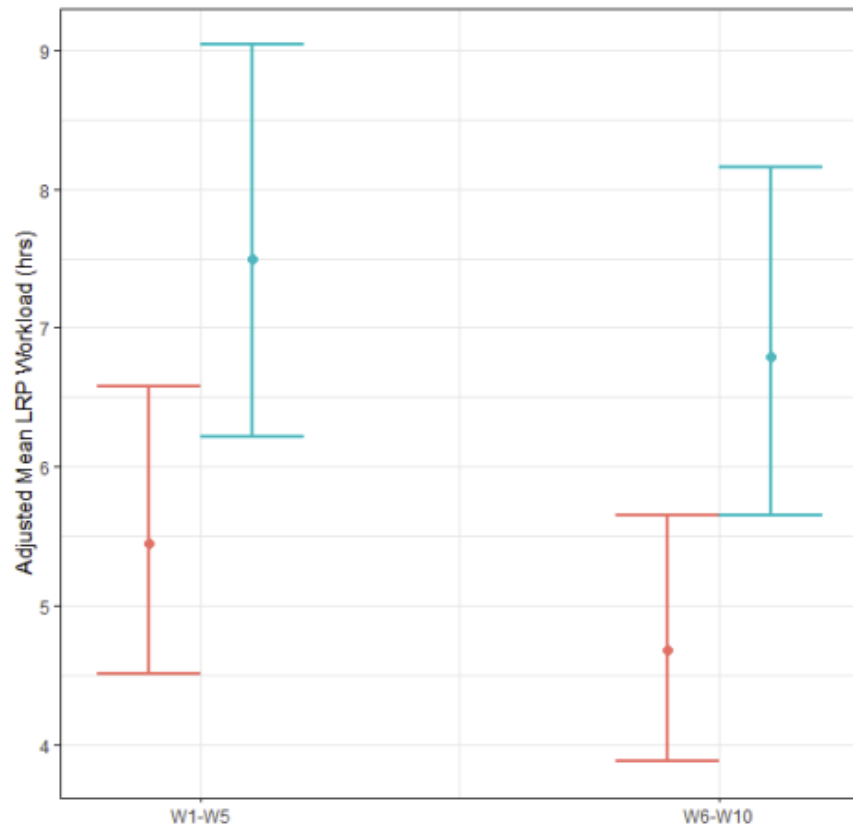
Three Paradigms of AI in Education



Three Paradigms of AI in Education



AI increases teacher performance, but performance gains $\neq$ learning gains



- RCT with 68 representative schools across the UK,

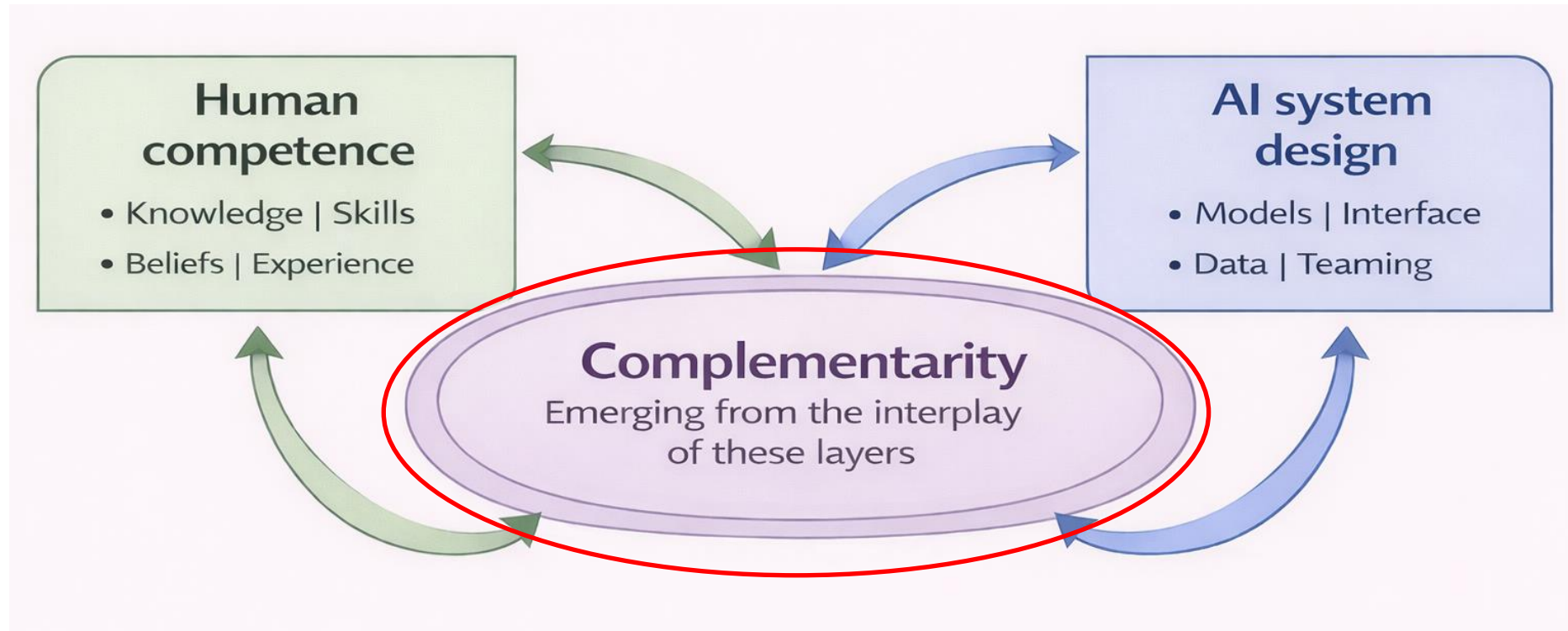
$$P_{H+AI} > P_H \text{ but } \frac{dC_H}{dt} \leq 0$$

- **No statistically significant difference in the quality of resources.**

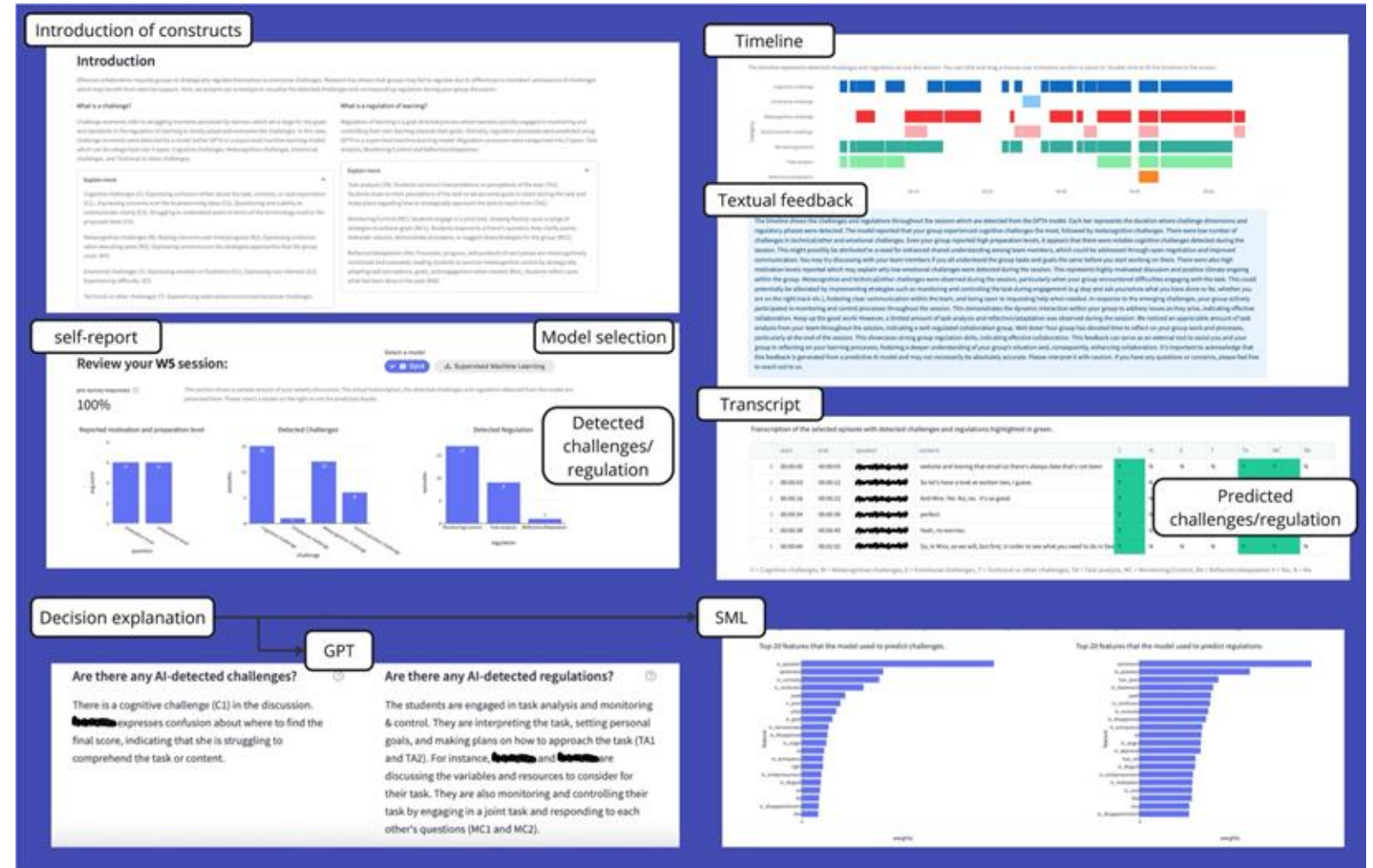
Currently, most teacher–genAI use operates in the replacement paradigm; it can improve performance while leaving teacher competence flat or declining.

From Replacement to Complementarity:

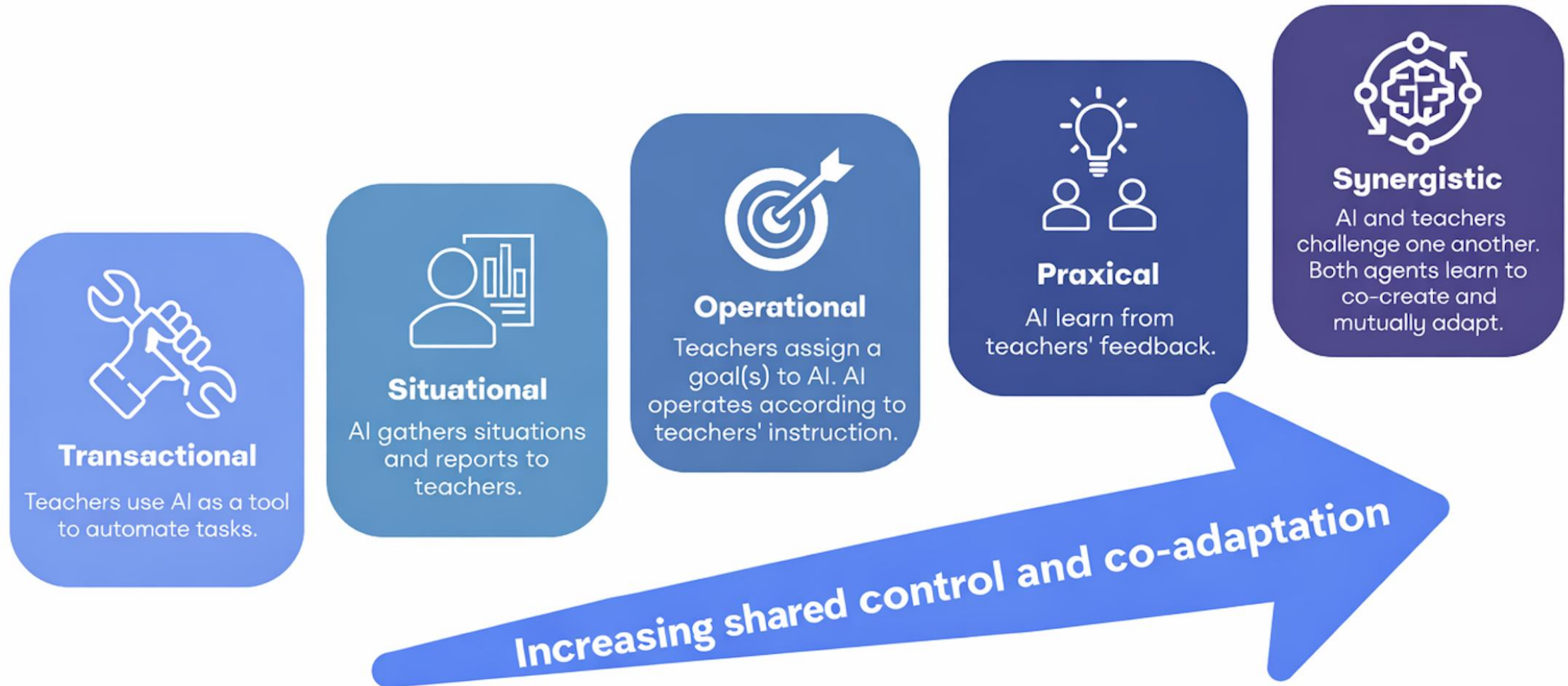
Designing Human–AI Systems that Grow Human Competence



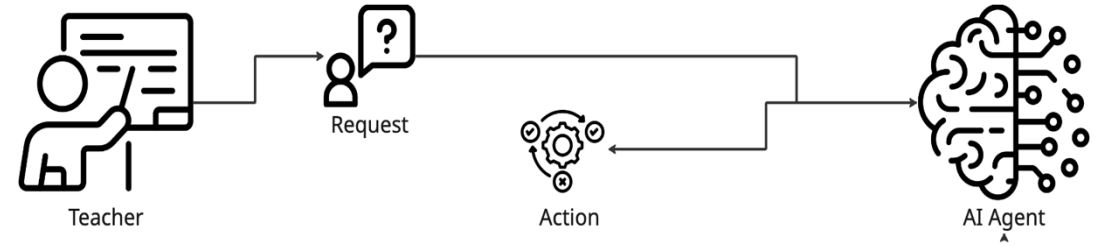
Collaborative Learning Feedback



Complementarity is not binary, it unfolds across (at least) five levels of human–AI teaming

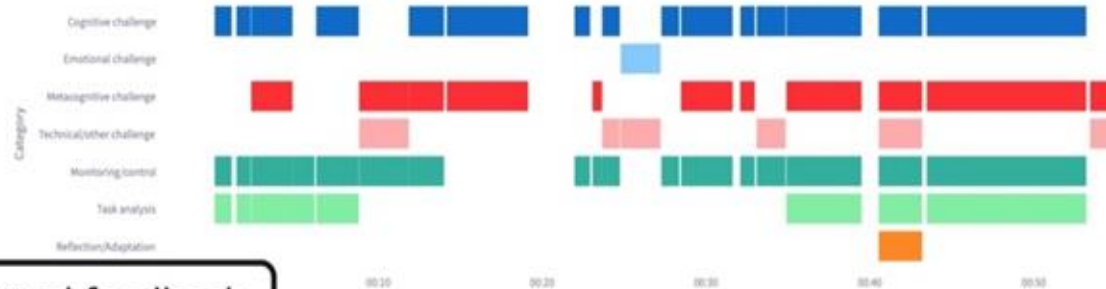


Transactional Teaming



Timeline

The timeline represents detected challenges and regulation across the session. You can click and drag a mouse over a timeline section to zoom in. Double click to fit the timeline to the screen.



Textual feedback

The timeline shows the challenges and regulations throughout the session which are detected from the GPT4 model. Each bar represents the duration where challenge dimensions and regulatory phases were detected. The model reported that your group experienced cognitive challenges the most, followed by metacognitive challenges. There were low number of challenges in technical/other and emotional challenges. Even your group reported high preparation levels, it appears that there were notable cognitive challenges detected during the session. This might possibly be attributed to a need for enhanced shared understanding among team members, which could be addressed through open negotiation and improved communication. You may try discussing with your team members if you all understand the group tasks and goals the same before you start working on them. There were also high motivation levels reported which may explain why low emotional challenges were detected during the session. This represents highly-motivated discussion and positive climate ongoing within the group. Metacognitive and technical/other challenges were observed during the session, particularly when your group encountered difficulties engaging with the task. This could potentially be alleviated by implementing strategies such as monitoring and controlling the task during engagement (e.g stop and ask yourselves what you have done so far, whether you are on the right track etc.), fostering clear communication within the team, and being open to requesting help when needed. In response to the emerging challenges, your group actively participated in monitoring and control processes throughout the session. This demonstrates the dynamic interaction within your group to address issues as they arise, indicating effective collaboration. Keep up the good work! However, a limited amount of task analysis and reflection/adaptation was observed during the session. We noticed an appreciable amount of task analysis from your team throughout the session, indicating a well-regulated collaboration group. Well done! Your group has devoted time to reflect on your group work and processes, particularly at the end of the session. This showcases strong group regulation skills, indicating effective collaboration. This feedback can serve as an external tool to assist you and your group in reflecting on your learning processes, fostering a deeper understanding of your group's situation and, consequently, enhancing collaboration. It's important to acknowledge that this feedback is generated from a predictive AI model and may not necessarily be absolutely accurate. Please interpret it with caution. If you have any questions or concerns, please feel free to reach out to us.

Control: AI-led processing
Teacher role: Delivery gatekeeper

AI role: Create and deliver feedback based on model parameters

Competence growth: Minimal (mainly from the tool use)

Situational Teaming

Define feedback for students:

Feedback for groups with silence time over: 15%

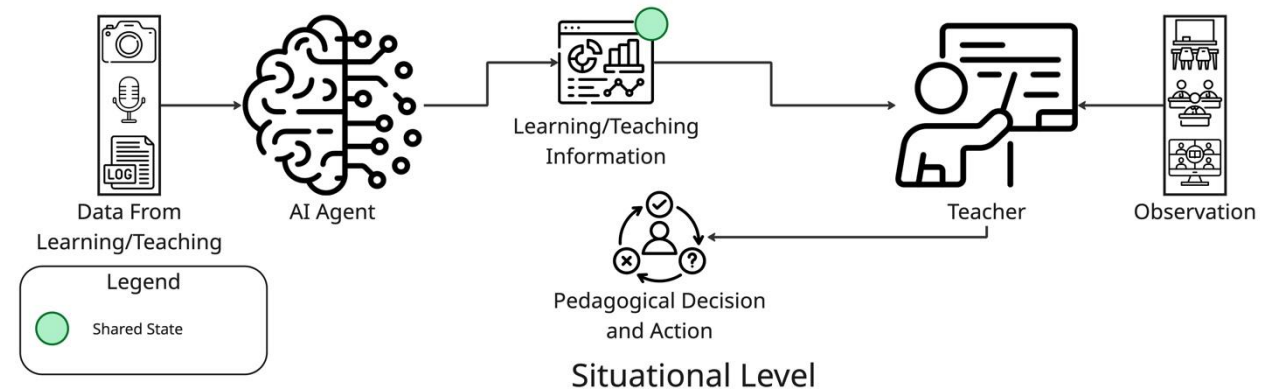
Notable silence time was observed during the session. While this may reflect your resource management, it may also suggest a low coordination through discussion. It is good to communicate with each other to align your group thinking.

AI-generated explanation

(evidence + theory)

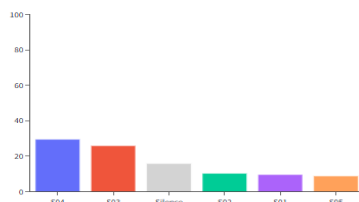
Evidence from the graph: the bar chart shows a non-trivial **Silence** portion relative to individual speaking bars, and speaking contributions are not evenly distributed (a small number of students account for a larger share of talk time). These patterns suggest limited group-level verbal exchange and/or uneven participation.

Theory consideration (collaborative learning): in collaborative learning, verbal interaction supports shared understanding, joint problem solving, and co-/socially shared regulation. More balanced participation and active turn-taking can increase opportunities for elaboration, monitoring, and coordination. When silence is prolonged and participation is uneven, groups may miss chances to externalise reasoning and align perspectives—hence the feedback encouraging more communication to coordinate thinking.

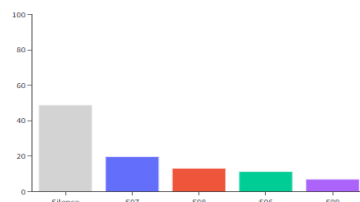


Control: AI-led detection, teacher-led interpretation
Teacher role: Transactional + Contextual interpreter
AI role: Process situational signals and create feedback based on model parameters
Competence growth: Increased situational awareness

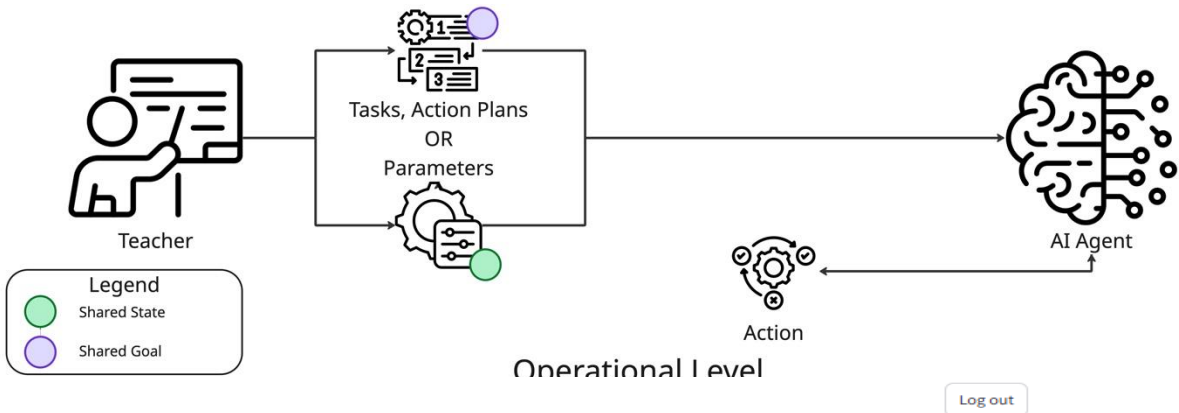
Group 1: ⚠️



Group 2: ⚠️



Operational Teaming



Classroom Report

Area to explore

Area to explore:

Select a feature:

speaking time

The total duration that a particular student spoke during the session.

Select a type of visualisation:

bar chart

Select a week:

5

Alerting criteria:

silence time

The proportion of time during which all group members remain silent.

exceeds the threshold of

(from 0% to 100%)

15

Alert: Groups with silence time over 15%.

Define feedback for students:

Define feedback for students:

Feedback for groups with silence time over: 15%

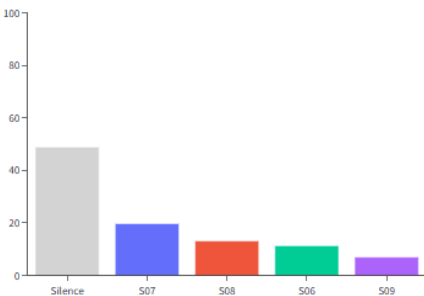
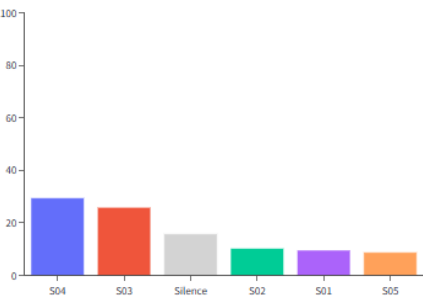
Notable silence time was observed during the session. While this may reflect your resource management, it may also suggest a low coordination through discussion. It is good to communicate with each other to align your group thinking.

Feedback for groups with silence time under: 15%

Only small portion of silence time was observed during the session. Well done on coordinating with peers through discussion.

Group 1:

Group 2:



Control: Shared but teacher-directed.

Teacher role: Situational + System configurator & pedagogical rule-setter.

AI role: Operationalise teacher input in model outputs

Competence growth: Teachers develop operational AI literacy, formalises pedagogical judgement with models

Praxical Teaming

Define feedback for students

Define feedback for students:

Feedback for groups with silence time over: 15%

Notable silence time was observed during the session. While this may reflect your good resource management, it may also suggest low coordination through discussion. Consider communicating more to align group thinking.

Teacher edit trace

Tracks saved versions of your feedback. When you edit the feedback and click Save, the system records the version and summarises what it learned from your edits, applying these insights to future feedback.

Versions saved: 1

Last saved: 11:50:41

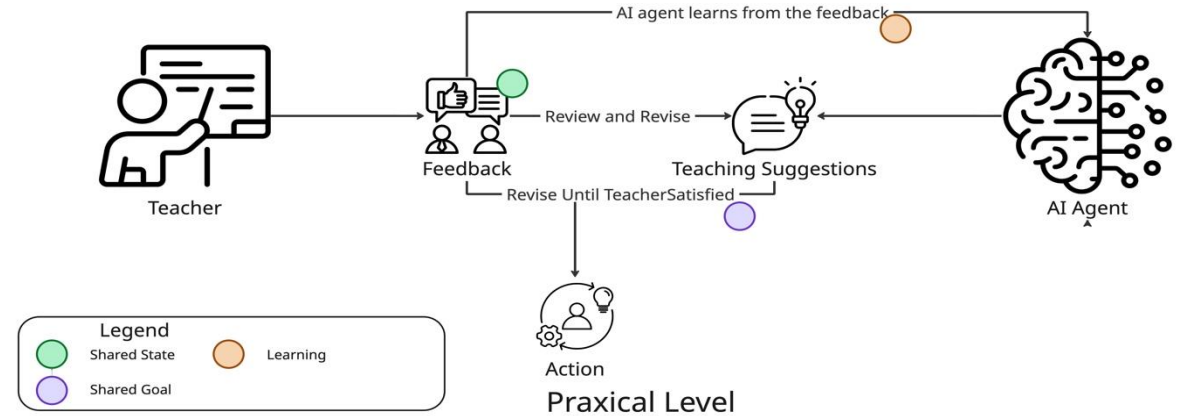
Clear History

Save

What the system learned

- You added more supportive/affirming wording.

Tip: Try adding actionable keywords (e.g., 'try', 'consider', 'make sure'), questions, hedging statements (e.g., 'could', 'might'), or affect-related keywords (e.g., 'good', 'great') to see the learned summary. This demo stores saved versions and summaries in your browser (localStorage) to simulate record-keeping.



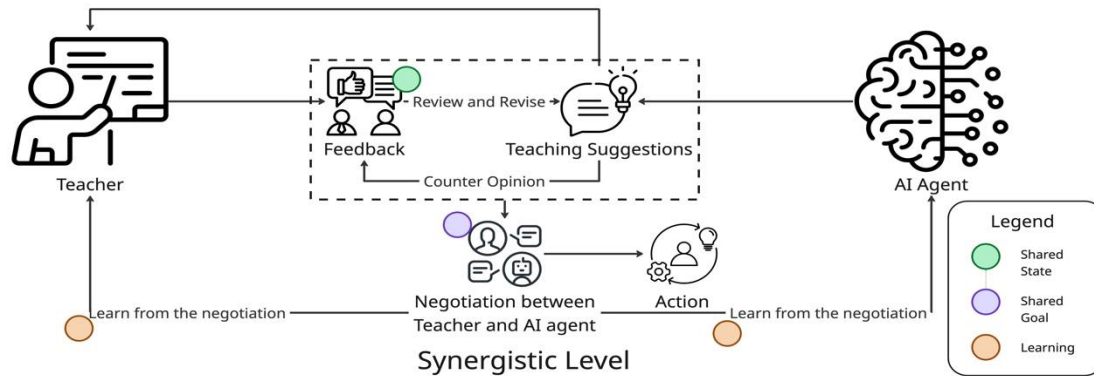
Control: Reciprocal and adaptive. Becomes co-regulated, teacher edits shape model behaviour, and model suggestions shape teacher reflection.

Teacher role: Operational + Co-designer and feedback calibrator.

AI role: The system models teacher professional judgement as the main signal input.

Competence growth: Meta-pedagogical and reflective competence. Teachers externalise tacit knowledge through edits, observe how the system internalises their preferences, and refine both their own practice and the AI's behaviour.

Synergistic Teaming



Control: Negotiated and co-regulative. Neither agent has unilateral authority; control emerges through structured negotiation.

Teacher role: Praxical + Epistemic co-constructor. Not only refines outputs but defends, revises, and re-theorises pedagogical decisions in response to AI challenges.

AI role: Critical co-agent, evaluating teacher proposals against learning sciences principles and historical data. Can push-back and introduce counter-positions as well as updating internal models through iterative engagement.

Competence growth: Emergent & bidirectional.

DEFINE FEEDBACK FOR STUDENTS

Define feedback:

For groups with silence time over 15% Edited

Your group's silence time exceeded 15% this week while some thinking time is healthy. Try using a round-robin approach so every member has a structured opportunity to share their thinking each session.

Try editing the feedback to see different AI discussions:

- Specificity** Remove the percentage number (e.g. delete "exceeded 15%" or replace with "is high")
- Actionability** Remove the round-robin strategy (e.g. delete "round-robin approach")
- Pedagogy** Remove the causal explanation (e.g. delete "unclear roles or hesitation to contribute")

Discuss with AI Deploy feedback

Discuss with AI

Current feedback:

Your group's silence time exceeded 15% this week while some thinking time is healthy. Try using a round-robin approach so every member has a structured opportunity to share their thinking each session.

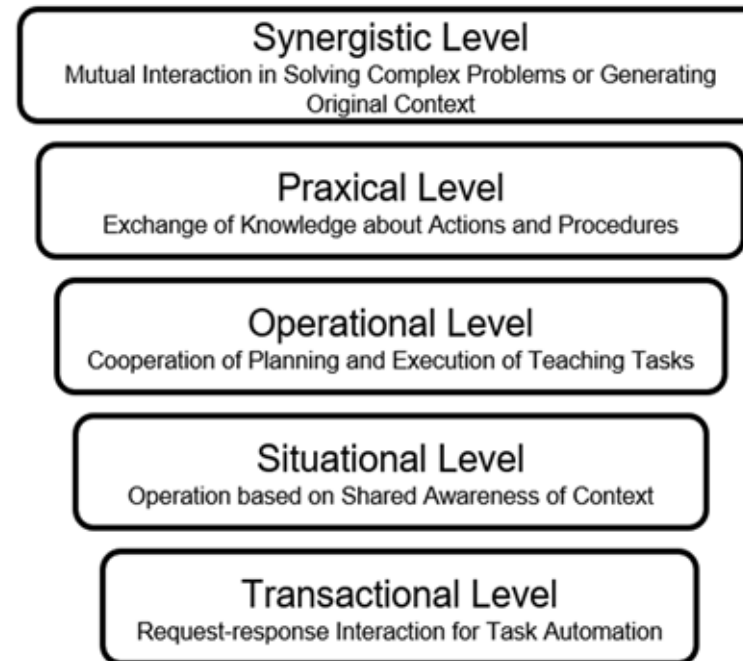
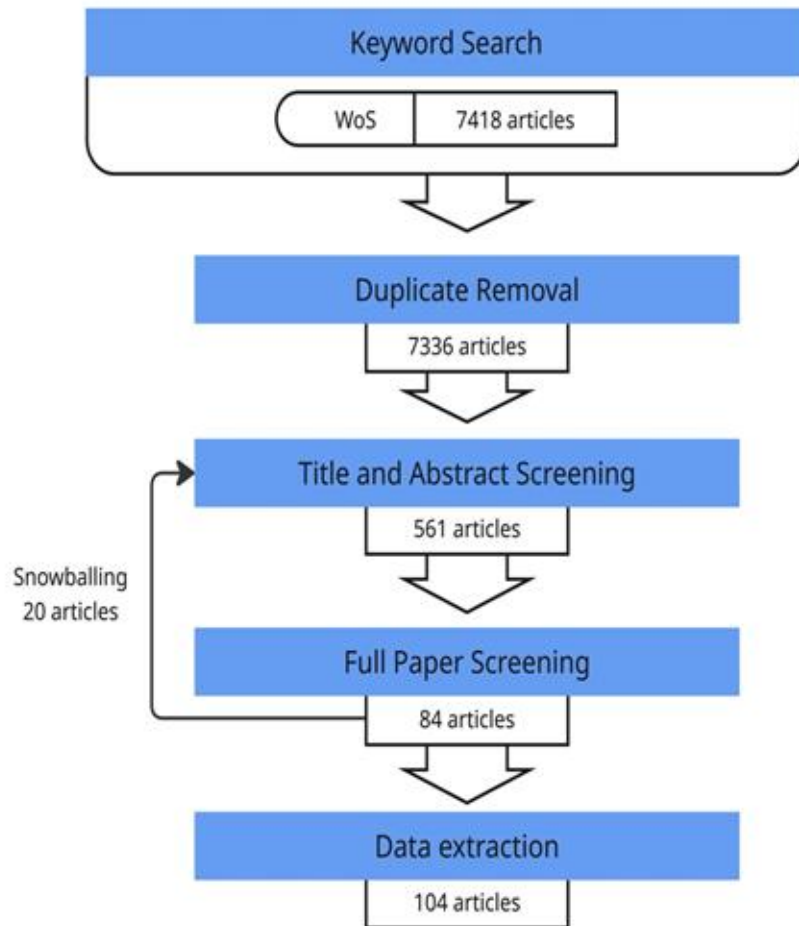
The most notable change in the updated scores is in **Pedagogical grounding** (4 → 3/5).

The original feedback offered a causal explanation for why silence is a signal worth attending to — linking it to unclear roles and hesitation to contribute. Your revision flags the silence but removes that explanatory layer. Järvelä et al. (2010) suggest that disengagement and silence often signal a loss of shared direction, whether due to unclear expectations, low confidence, or uncertainty about roles.

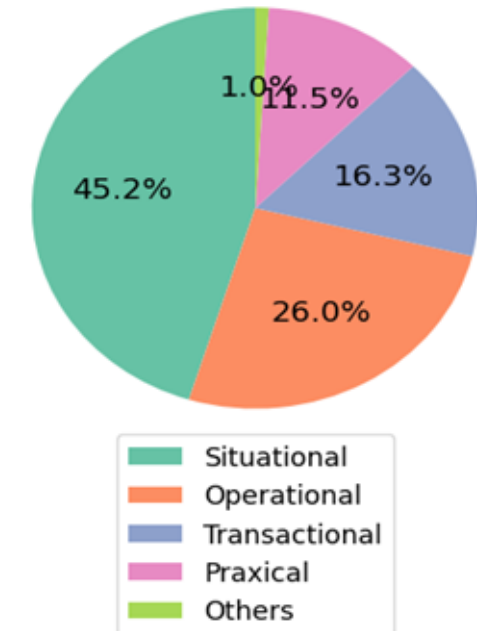
Without that framing, students may read the feedback as a criticism of their behaviour rather than a cue to reflect on their group dynamics. Do you feel your version still helps students understand why silence is a concern, or was the causal explanation removed deliberately?

I was unsure how to explain why silence matters — it may be context-dependent.

Synergistic human–AI teaming remains largely absent in teacher-facing systems



$$P_{H+AI} > \max(P_H, P_{AI}) \quad \frac{dC_H}{dt} > 0 \quad \frac{dC_{AI}}{dt} > 0$$



The future of AI in education is not in better AI models for replacement, but better teacher–AI teaming.



Shaping the Future of AI



UCLAIT - Team

Taico
Teacher-AI Complementarity



Mutlu Cukurova

**Professor of Learning and Artificial Intelligence
UCL Knowledge Lab & UCL Centre for Artificial Intelligence
University College London**

m.cukurova@ucl.ac.uk

