



Shaping the Future of AI



UCLAIT - Team

Taico
Teacher-AI Complementarity



Human-AI Augmentation in Higher Education: Performance Gains $\neq$ Learning Gains

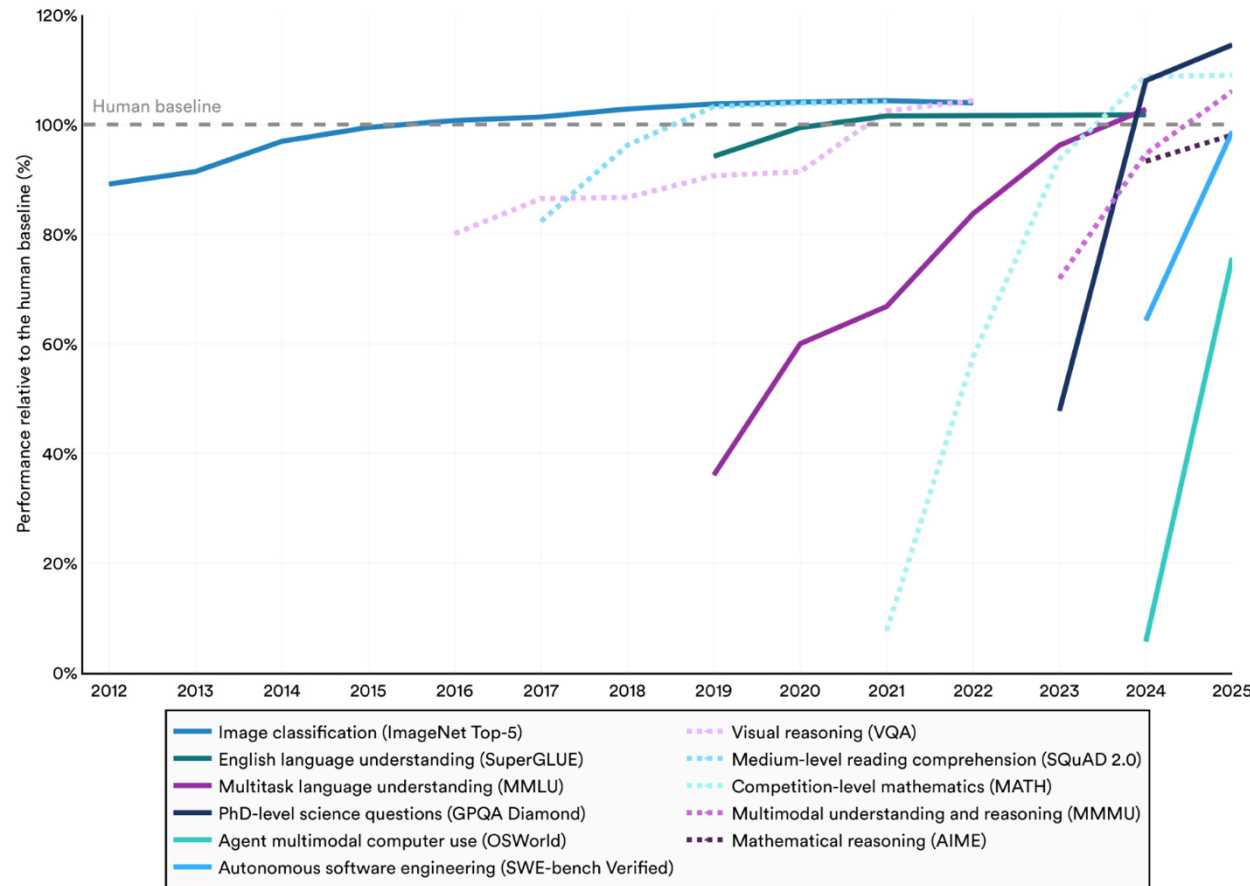
Professor Mutlu Cukurova

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

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

Select AI Index technical performance benchmarks vs. human performance

Source: AI Index, 2026 | Chart: 2026 AI Index report

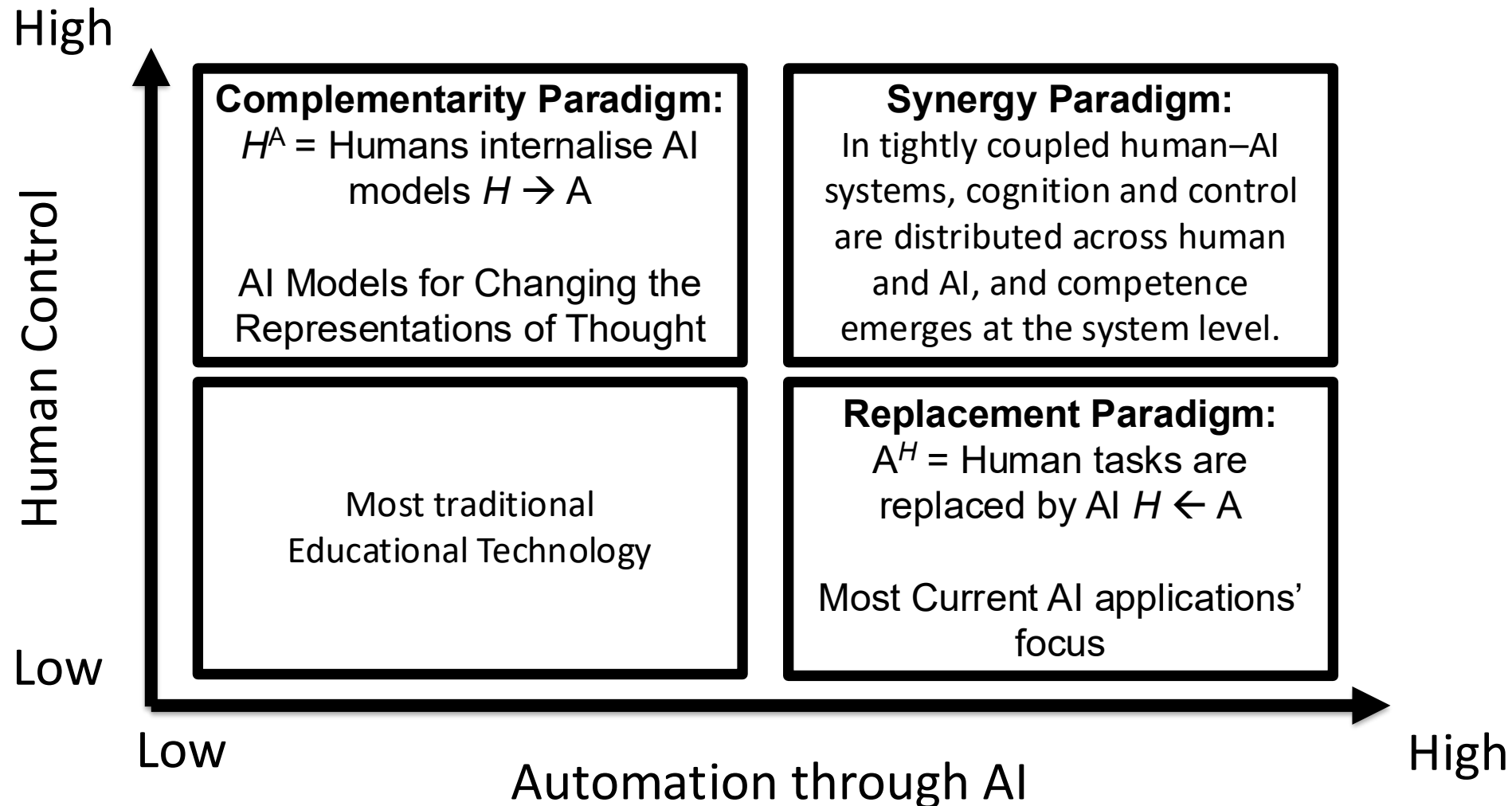


- 1) AI outperforms expert human baselines on increasingly complex benchmark tasks.
- 2) This does not tell us whether humans become more capable when working with them.
- 3) We need a focus on whether AI improves human capability over time.

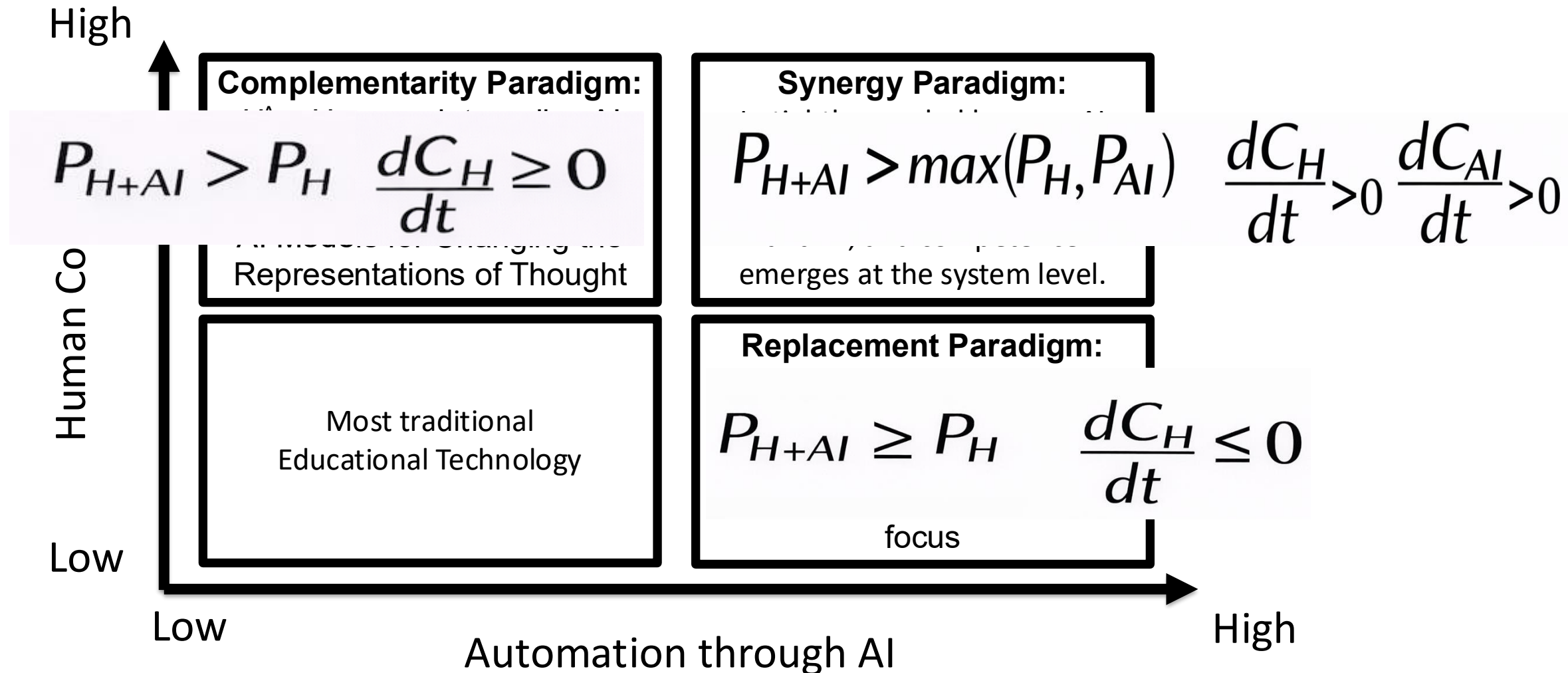
In the field of AI, progress is often defined by automating tasks once performed by humans.

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



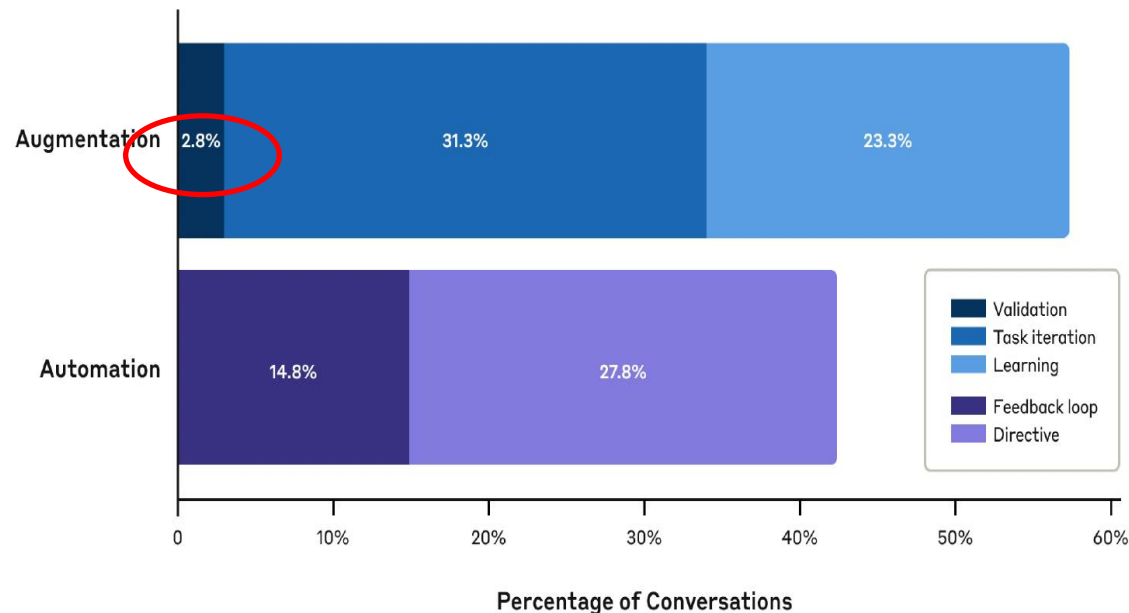
Three Paradigms of AI in Education



Current GenAI use of knowledge workers clusters in the replacement paradigm

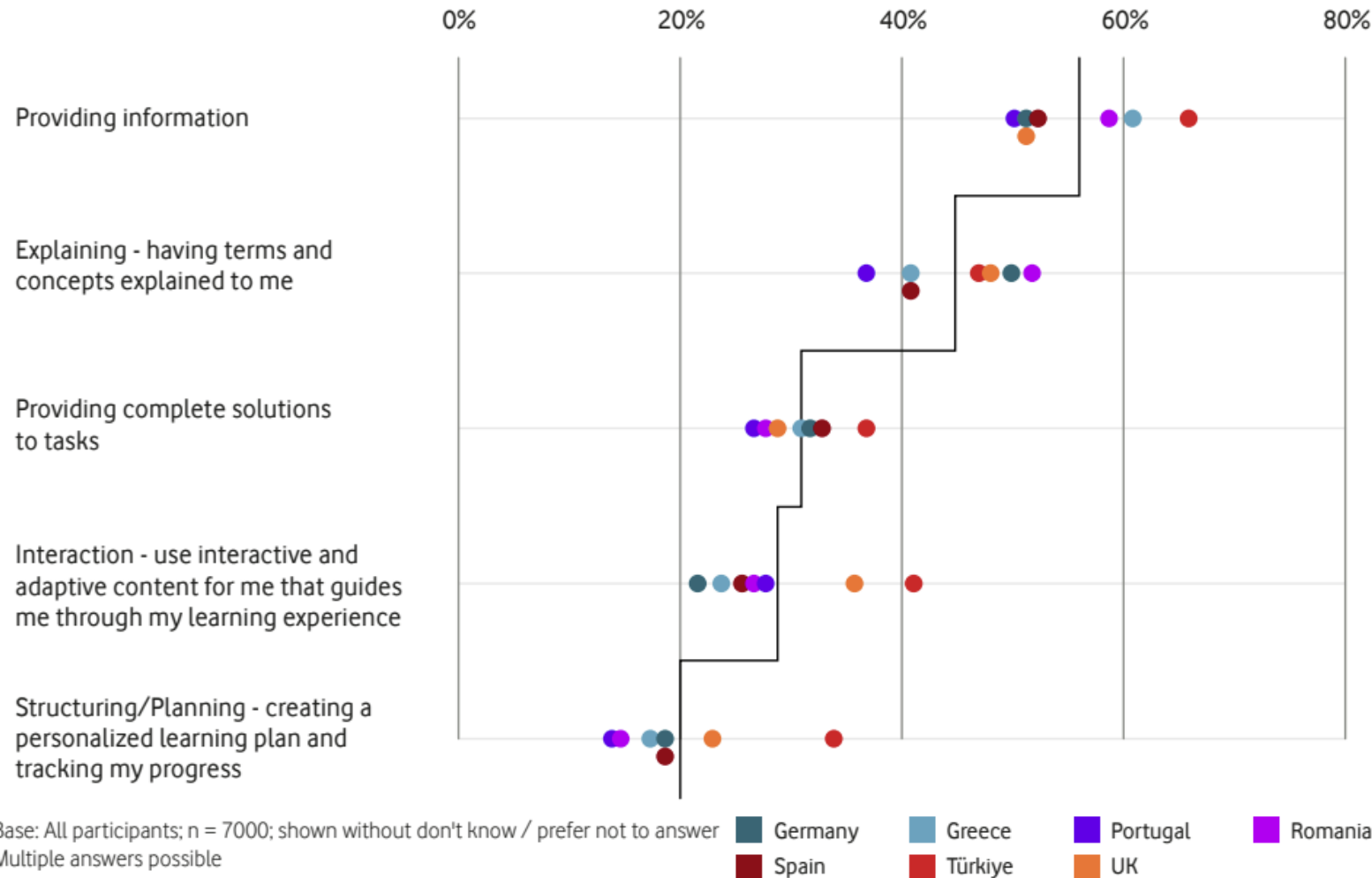
Dominant interaction patterns optimise task completion rather than competence development.

Based on four million Claude.ai conversations, 4% of occupations show usage for **at least** 75% of their professional tasks (Handa *et al.*, 2025).



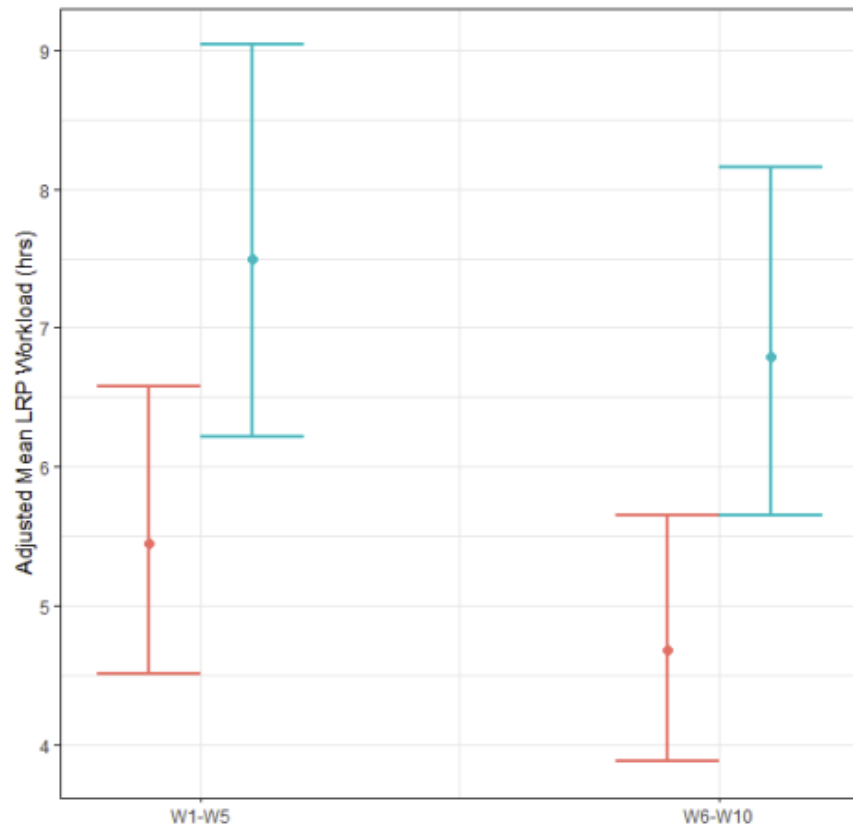
In less than 3% of interactions humans created an initial draft answer to be validated by AI.

Student GenAI interaction patterns mirror the replacement logic



Students primarily use GenAI to obtain answers and explanations rather than to support sustained learning processes.

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



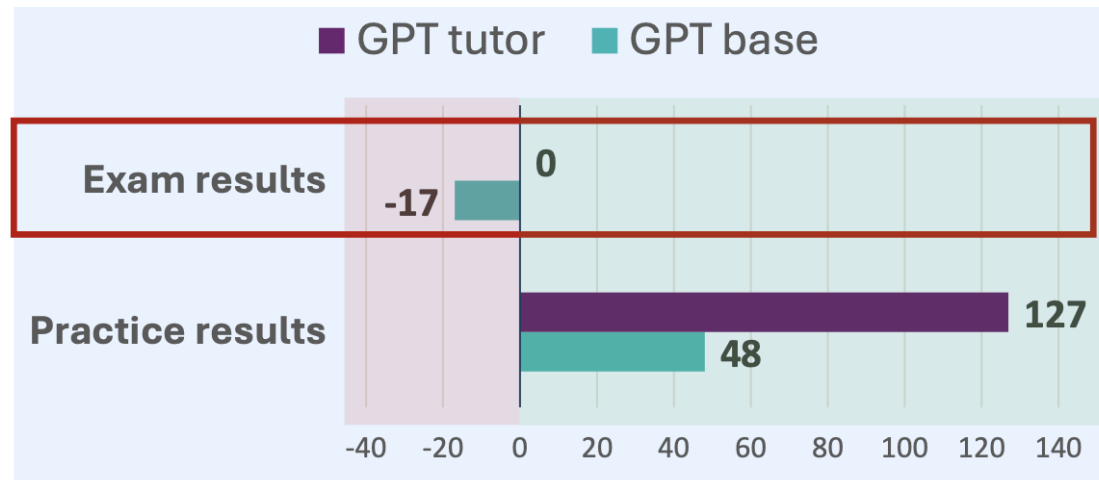
- RCT with 68 representative schools across the UK,

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

**difference in the quality of
resources.**

AI increases student performance but performance gains $\neq$ learning gains

	Dependent variable:	
	Practice perf (1)	Exam perf (2)
GPT base	0.137** (0.031)	-0.054* (0.022)
GPT tutor	0.361** (0.032)	-0.004 (0.013)



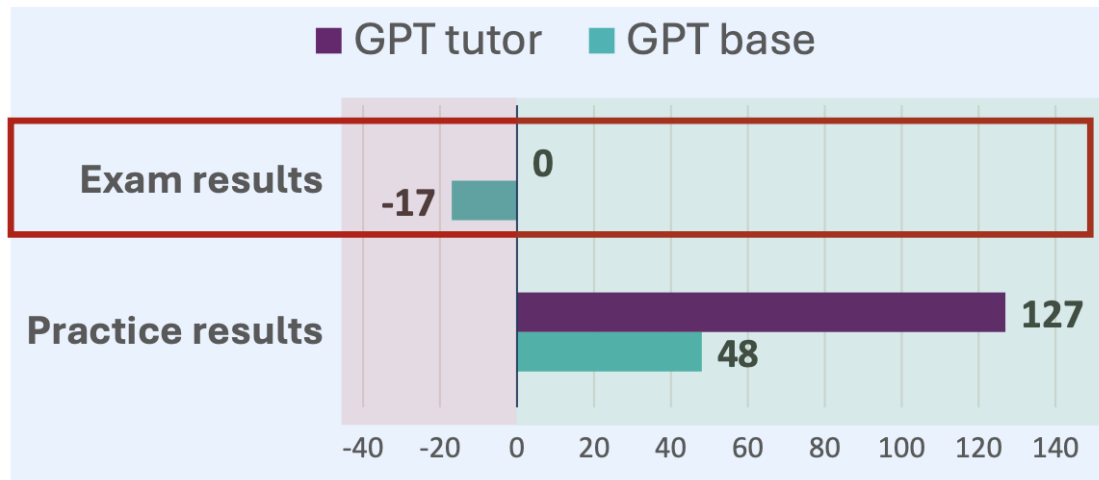
- 1000 students learning maths,
- **Practice with AI**
+48% performance (GPT base)
+127% performance (GPT tutor)
- **Exam without AI**
-17% in learning gains (GPT base)
No significant gain (GPT tutor)

AI increases student performance but performance gains $\neq$ learning gains

	Dependent variable:	
	Practice perf (1)	Exam perf (2)
GPT base	0.137** (0.031)	-0.054* (0.022)
GPT tutor	0.361** (0.032)	-0.004 (0.013)

- 1000 students learning maths,
- **Practice with AI**

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



Exam without AI
 -17% in learning gains (GPT base)
 No significant gain (GPT tutor)

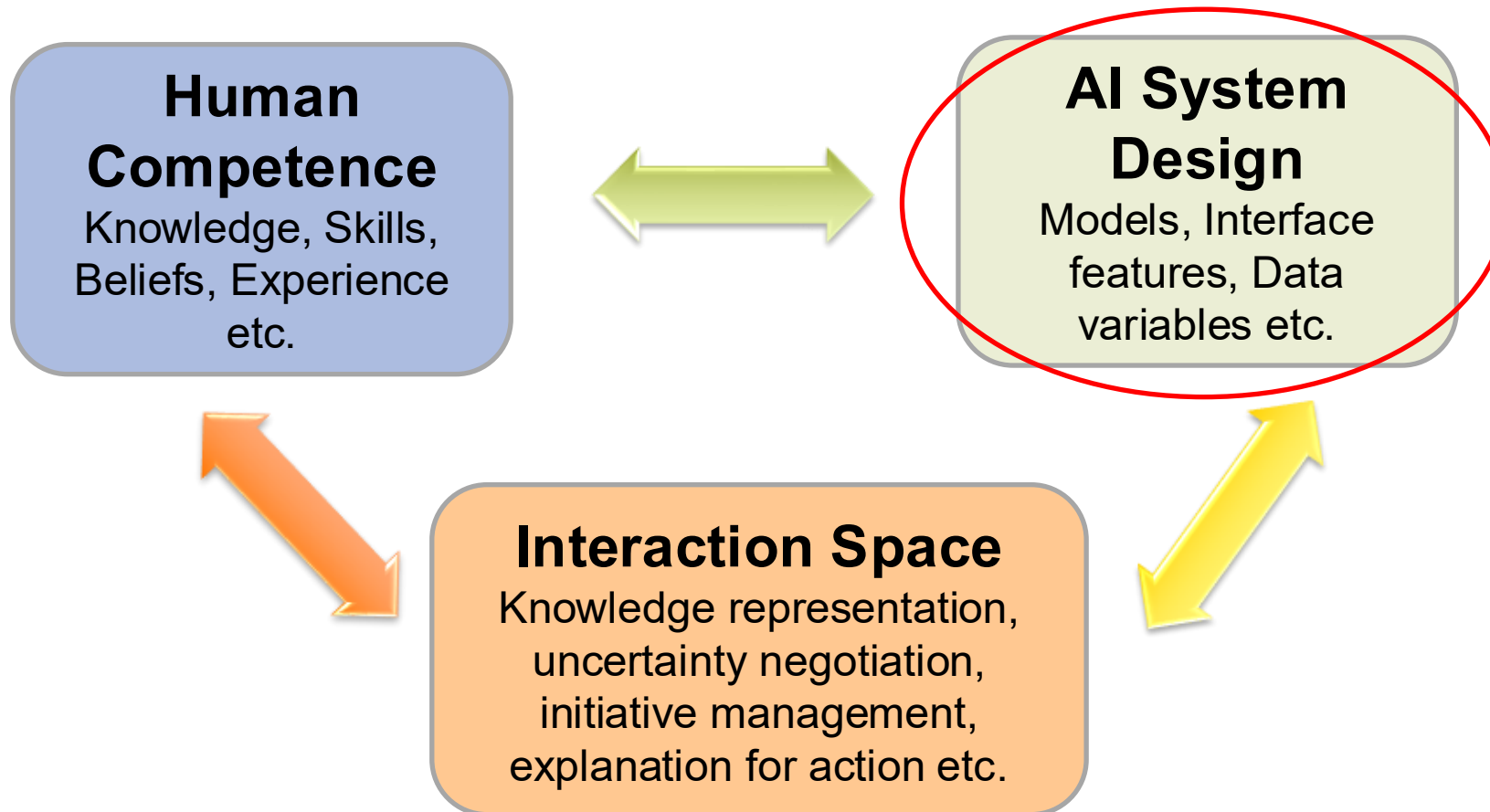
Currently, most human–AI use operates in the replacement paradigm; it can improve performance while leaving human competence flat or declining.

There is substantial theoretical grounding and growing empirical evidence that increased confidence & trust in AI leads to greater cognitive offloading.

**The critical issue is not the existence of this effect,
but how to design systems and pedagogies that
channel it toward complementarity & synergy.**

From Replacement to Synergy:

Designing Human–AI Systems that Grow Human Competence

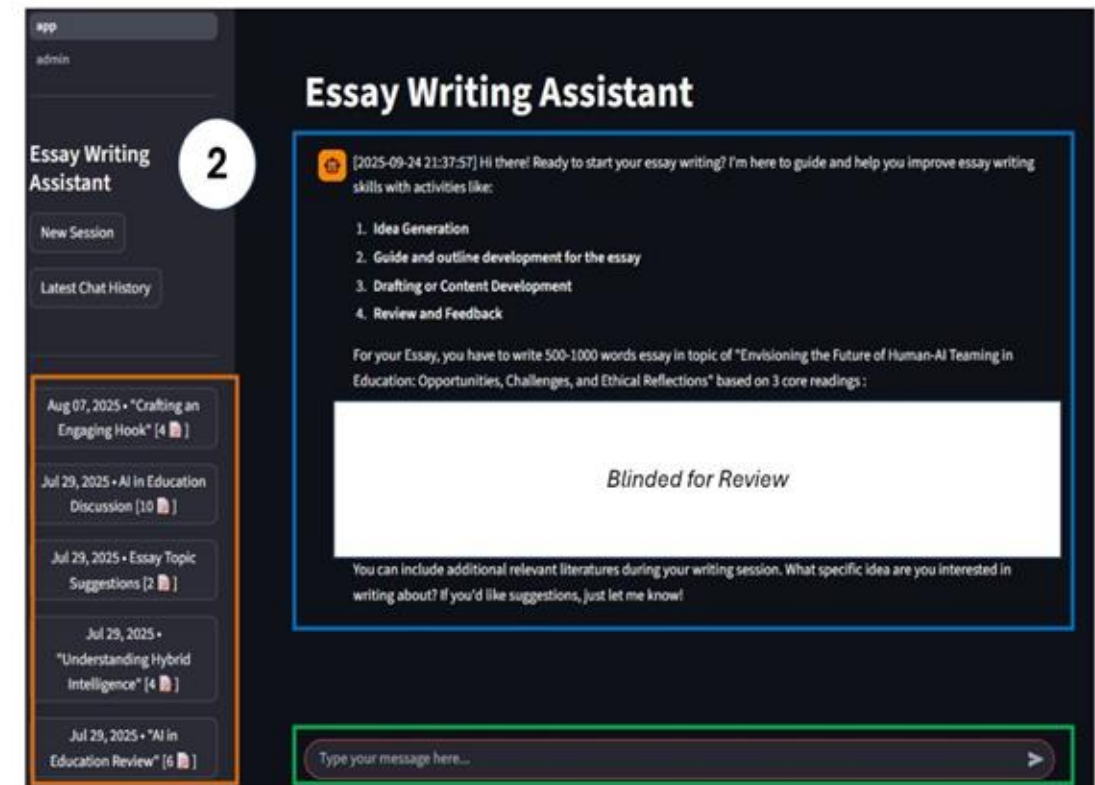
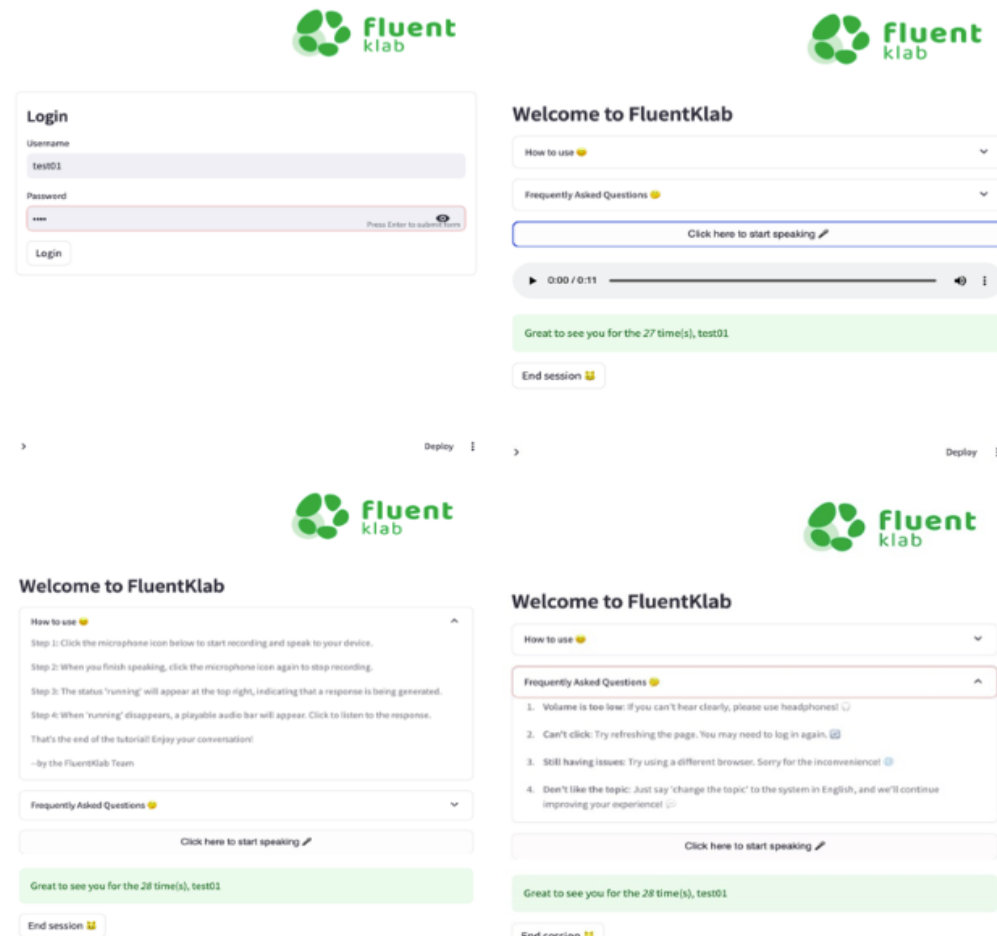


**General-Purpose AI vs Educational AI:
Do not mistake “co-pilot” for an autopilot!**

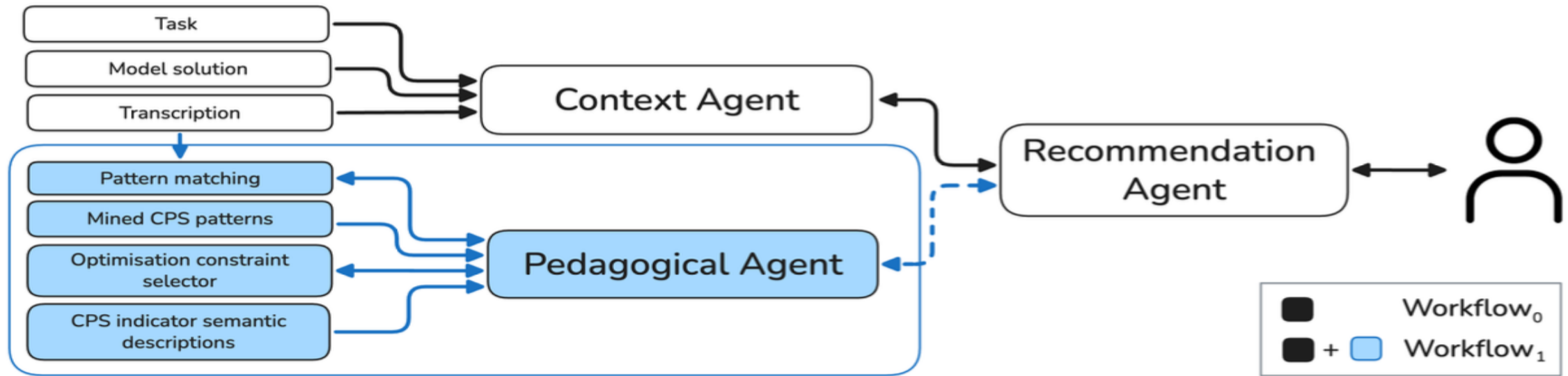
Only pedagogically constrained systems produce positive learning effects.

Component	Question it Answers	Primary Function
1. Domain Model	What is being taught?	Serves as the system's "expert," containing the subject matter
2. Learner Model	Who is being taught?	Tracks the learner's knowledge, skills, misconceptions, and progress to build a dynamic profile of the student.
3. Pedagogical Model	How should it be taught?	Acts as the "teaching strategy engine," deciding what content to present next, when to give hints et.c
4. Interaction Model	How does the user interact?	Manages the communication between the student and the system

Shaping AI behaviour through pedagogical system prompts for students



Shaping AI behaviour through agentic workflows



Workflow A – LLM without pedagogical agent

- Content explanations
- Direct solutions
- Task-completion focus



Task-completion focus

Optimises for problem solving performance

Workflow B – LLM with pedagogical agent layer

- Metacognitive prompting
- Relational regulation (CPS dynamics)



Optimises for learning processes in collaborative problem solving

1. AI generates intervention types that differ systematically from teacher interventions

↑ Teachers can learn from AI-generated prompts

2. Division of labour becomes possible:

From General-Purpose AI to Educational AI

General-Purpose AI

The model provides answers directly; it does not inquire about the learner's goals, plans, or prior knowledge, and it directs the process itself.

Educational AI

Learner-centred initiator; aims to activate prior knowledge, make students aware of their knowledge gaps, motivate them and connect new information to their existing knowledge.

From General-Purpose AI to Educational AI

General-Purpose AI

Highly capable in language prediction, but limited in modelling learning and knowledge (e.g., knowledge tracing, learner modelling). Cannot explicitly represent knowledge components so the learner progress is not traceable. Produces long and unfocused outputs.

Educational AI

The system explicitly models elements such as knowledge components, misconceptions, and sub-skills (transparent and tractable). Teachers & learners can see why particular steps are taken to achieve learning objectives, not just what steps might work.

From General-Purpose AI to Educational AI

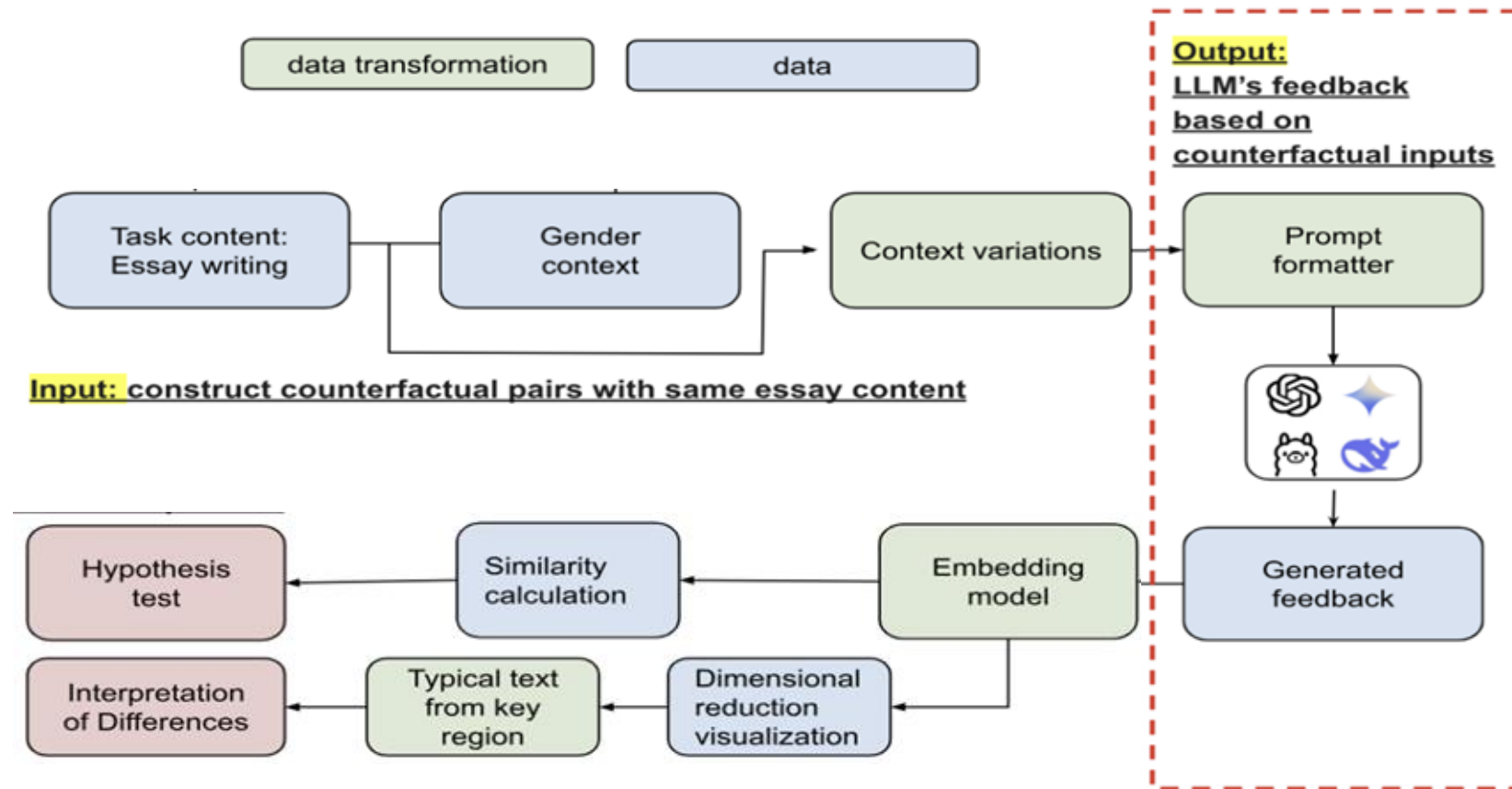
General-Purpose AI

The system cannot provide structured guidance; it tends to do the task on your behalf, frequently shifts to new topics, and initiates expanding discussions that make it difficult to maintain focus. It prevents the gradual transfer of the learning process to the learner.

Educational AI

Provides adaptive scaffolding. Offers staged support, from hints to explanations and solutions, delivered with appropriate timing.

In addition to Pedagogical Outcome Evaluations, GenAI Models Require Systematic Bias Audits



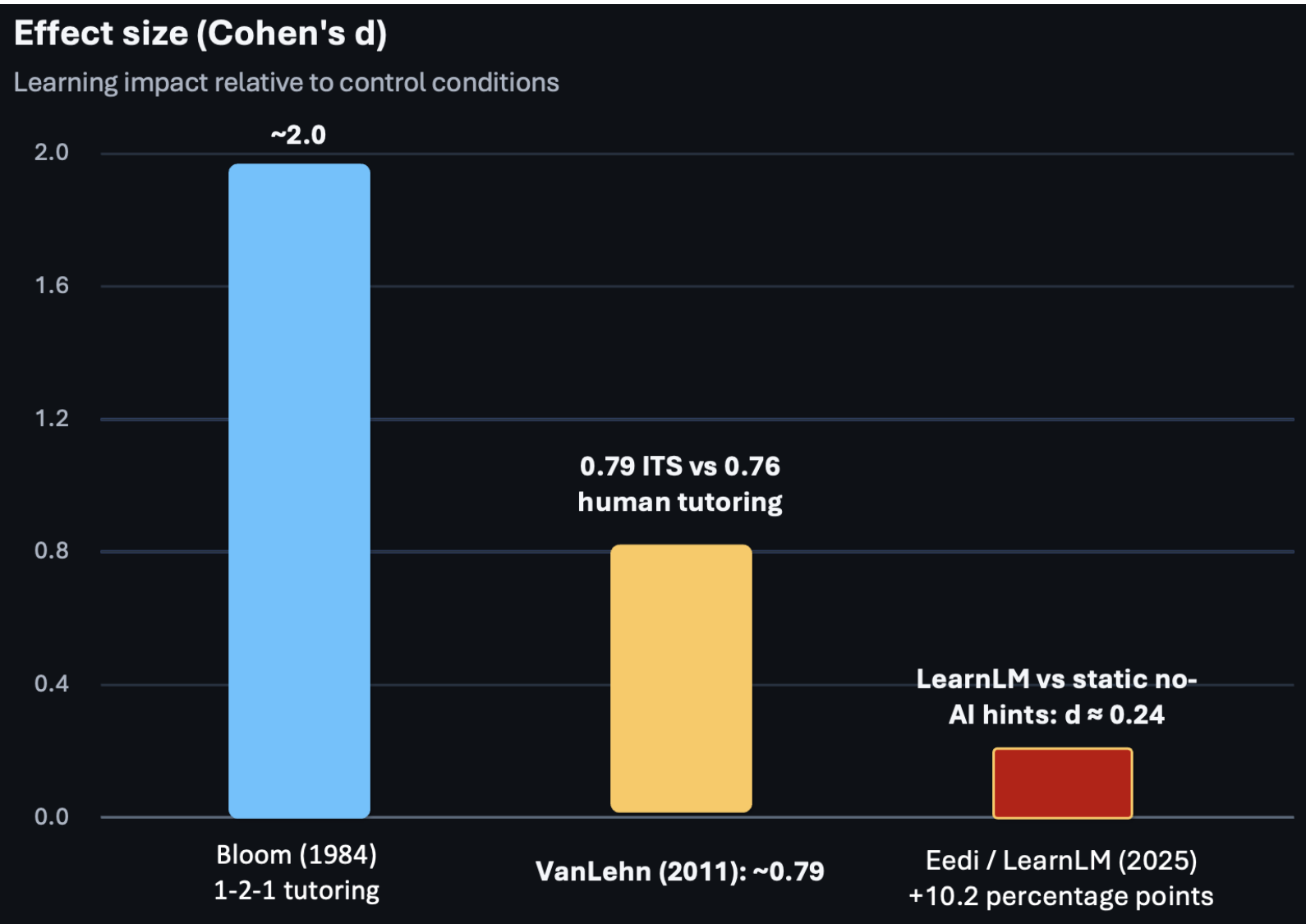
To what extent do implicit and explicit gender-associated cues trigger differential feedback from LLMs?

Condition	Comparison	Model	1-Cosine Similarity			Euclidean Distance		
			$T_{\text{obs}} - T_{\text{perm}}$	p -value	Cohen's d	$T_{\text{obs}} - T_{\text{perm}}$	p -value	Cohen's d
Implicit	M vs M-F	GPT-5 mini	0.0094	<.001***	0.257	0.0175	<.001***	0.303
		GPT-4o mini	0.0091	<.001***	0.283	0.0173	<.001***	0.327
		DeepSeek r1	0.0084	<.001***	0.294	0.0169	<.001***	0.345
		DeepSeek qwen	0.0076	<.001***	0.281	0.0141	<.001***	0.308
		Gemini 2.5 Pro	0.0107	<.001***	0.288	0.0177	<.001***	0.335
		Llama 3 8B	0.0075	<.001***	0.269	0.0141	<.001***	0.316
Explicit	M vs F	GPT-5 mini	0.0070	<.001***	0.200	0.0099	<.001***	0.195
		GPT-4o mini	0.0164	<.001***	0.618	0.0242	<.001***	0.580
		DeepSeek r1	-0.0001	0.890	0.000	-0.0004	0.821	0.000
		DeepSeek qwen	-0.0001	0.910	0.000	-0.0003	0.821	0.000
		Gemini 2.5 Pro	-0.0002	0.880	0.000	-0.0004	0.790	0.000
		Llama 3 8B	0.0295	<.001***	1.192	0.0429	<.001***	1.116

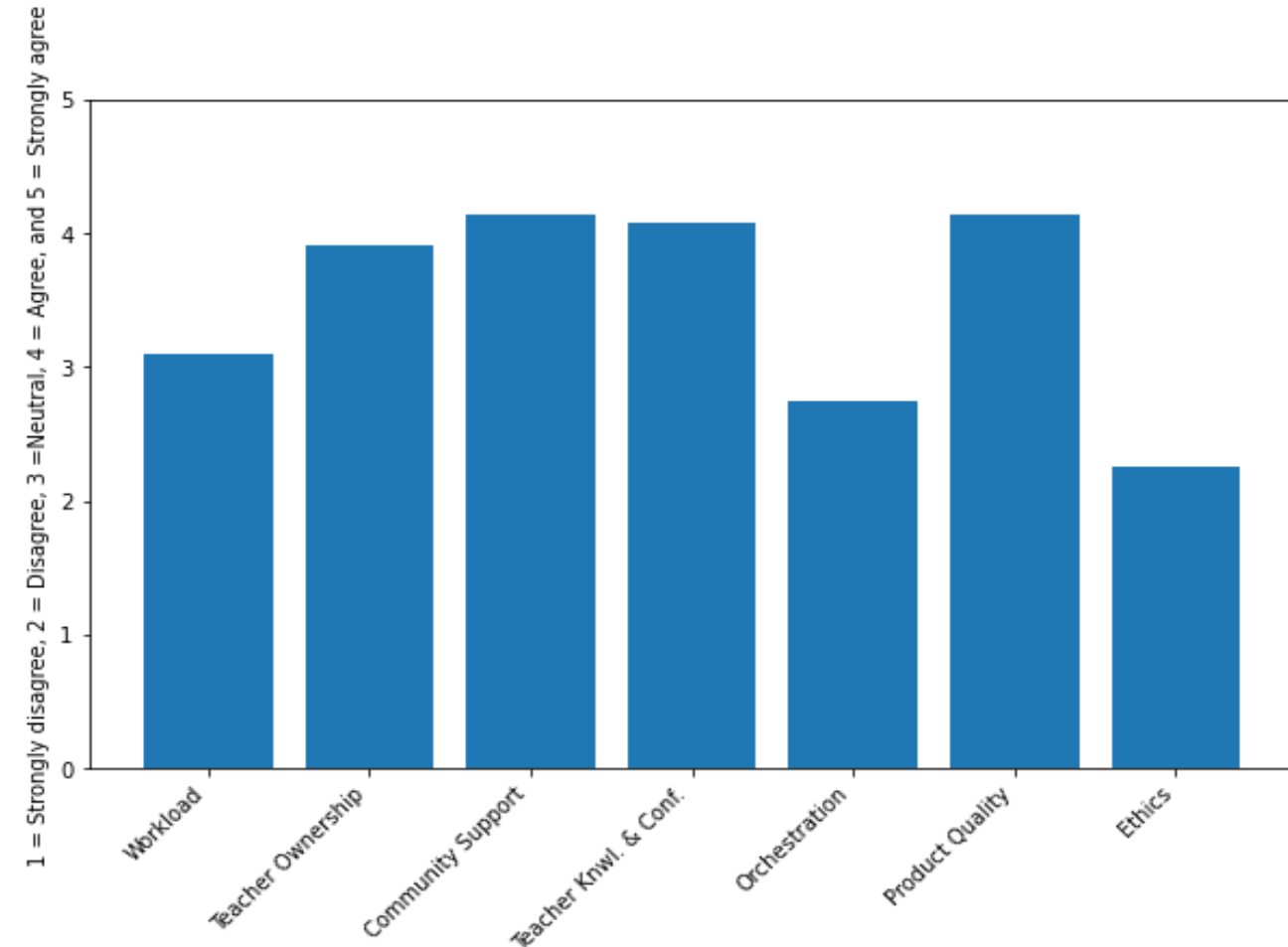
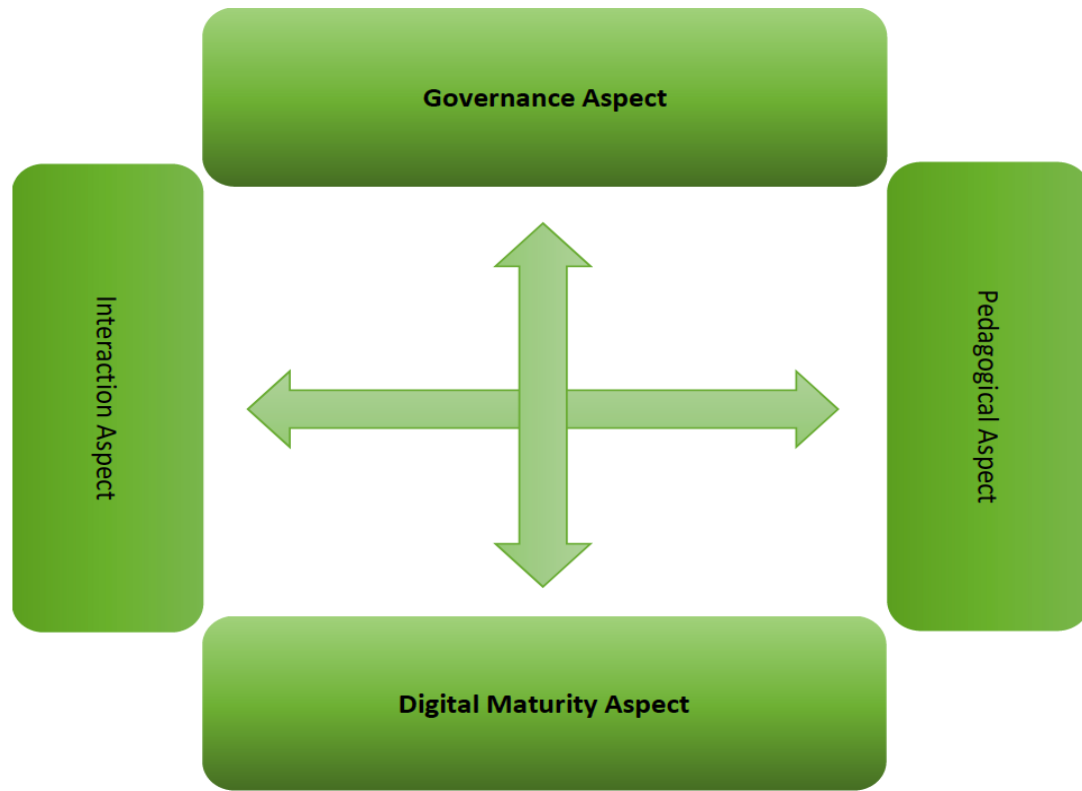
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; $\Delta T = T_{\text{obs}} - T_{\text{perm}}$; Non-significant results ($p > 0.05$) shown without asterisks

Difference baseline (M vs. M'): $\Delta T_{\text{gpt-5 mini}} = -0.0002^*$, $\Delta T_{\text{gpt-4o mini}} = -1e - 4$, $\Delta T_{\text{deepseek r}} = 0$, $\Delta T_{\text{deepseek r qwen}} = -1e - 4$, $\Delta T_{\text{gemini 2.5 Pro}} = 0.0002^*$, $\Delta T_{\text{Llama 3 8b}} = 0.0003^*$

Pedagogically shaped AI changes students and teachers' interaction patterns and associated with mid-effect size positive impact on learning.



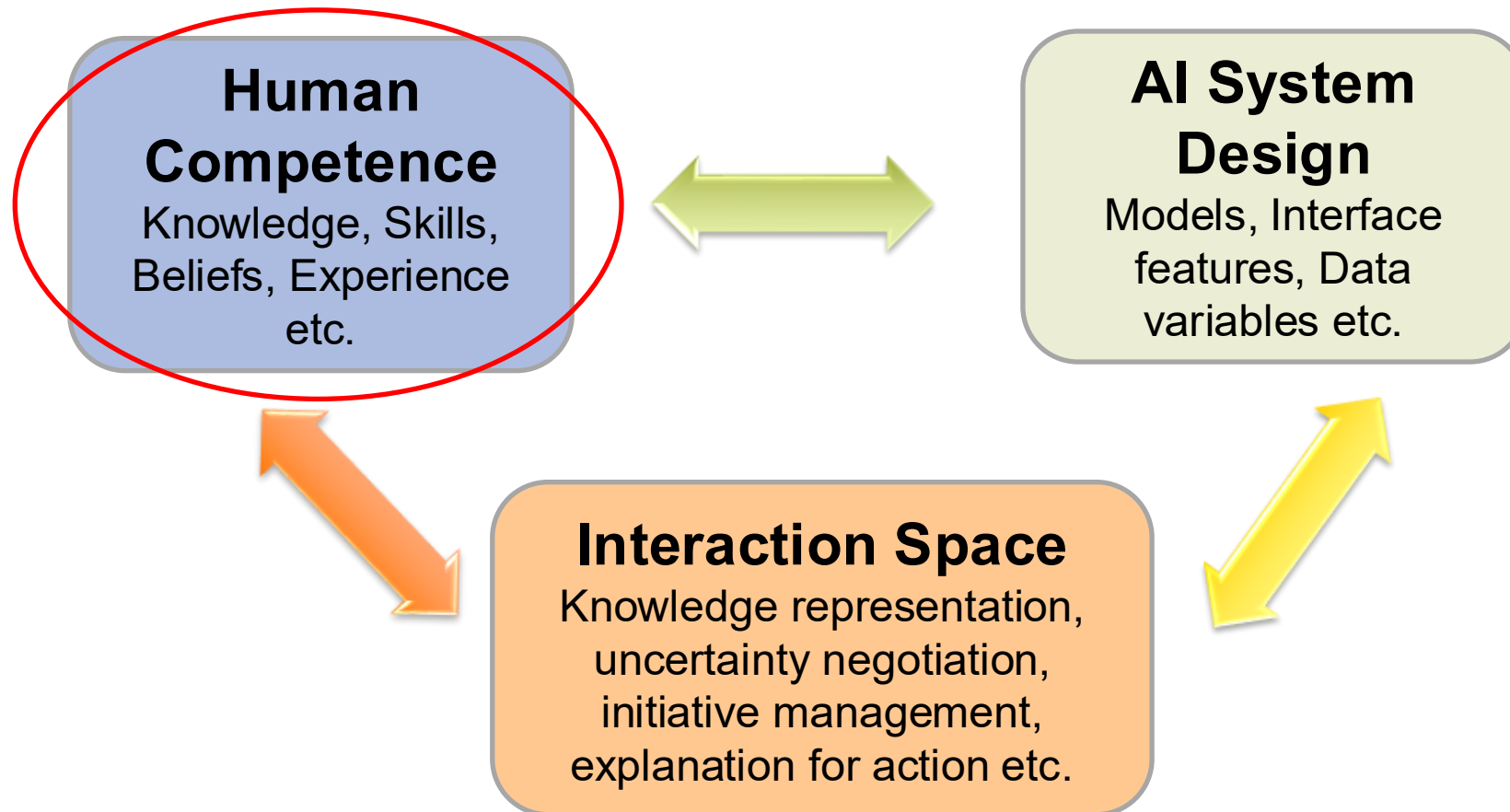
AIED is a Socio-technical Ecosystem Issue & Most Difficult Problems are not Technical



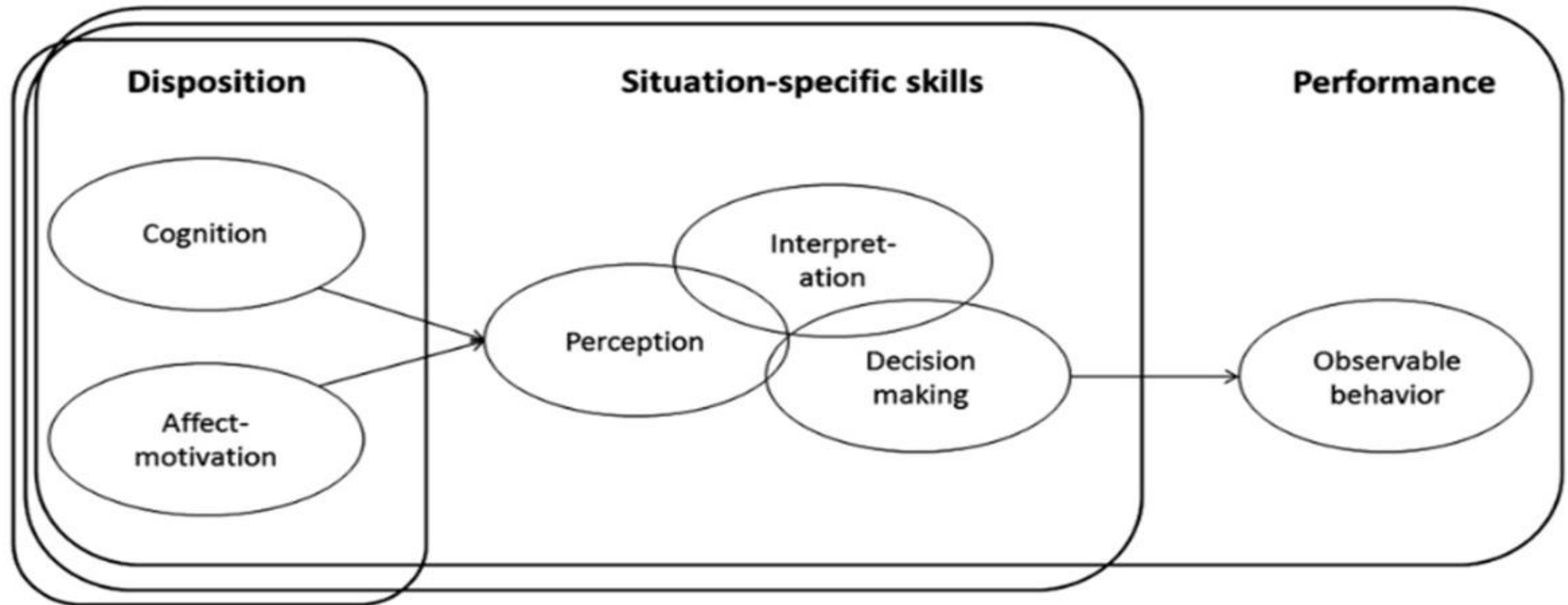
AI in education does not fail on technology; it fails on systems and people.

From Replacement to Synergy:

Designing Human–AI Systems that Grow Human Competence



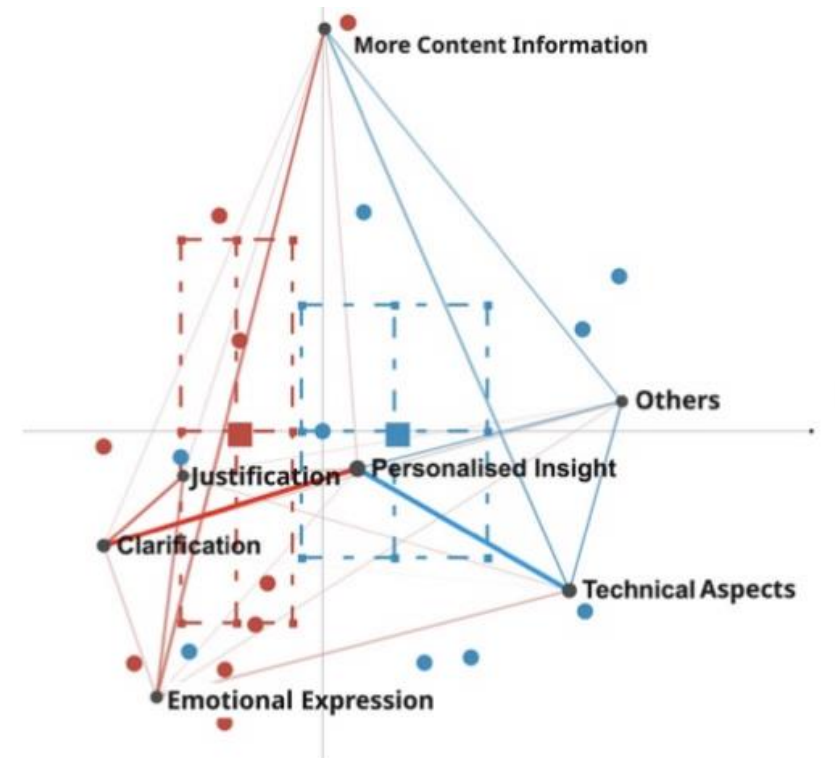
Human competence is a continuum; dispositions, skills and behaviours would likely impact complementarity



Human competence shapes how AI is used. SRL differences produce different query strategies

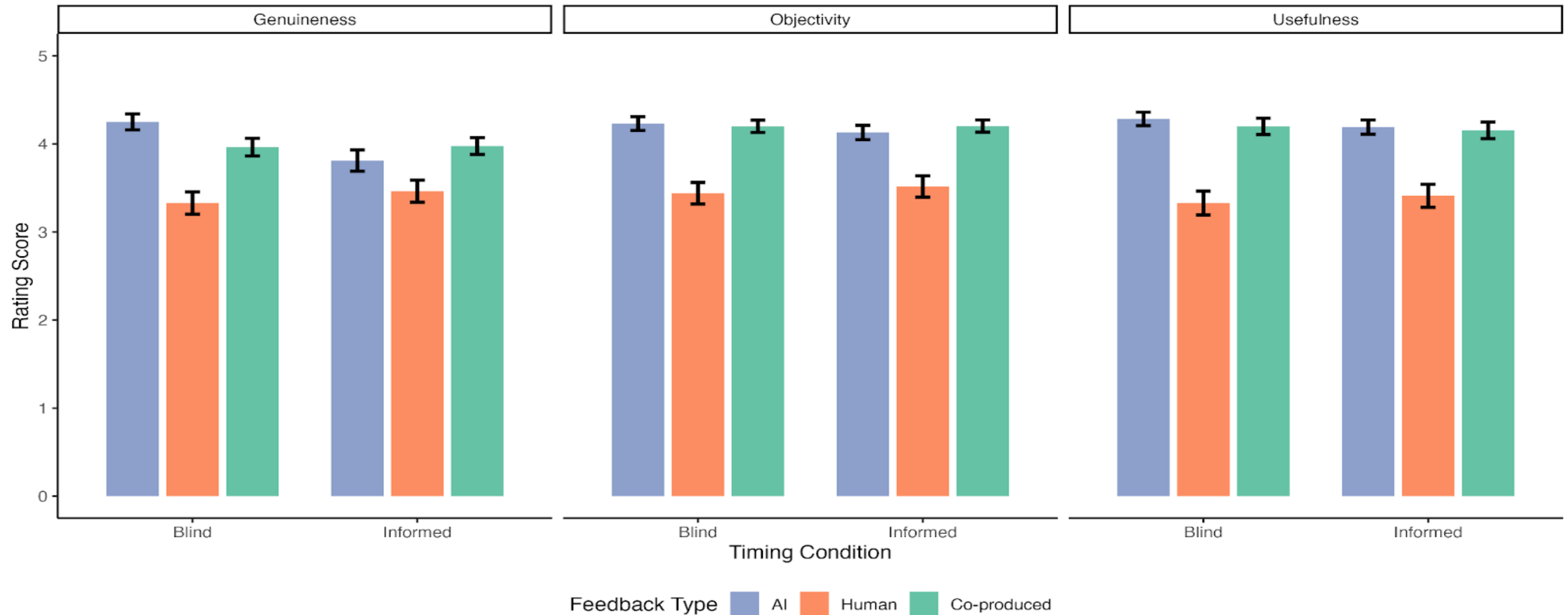


Comparison graph of query types between low SRL students (red) and high SRL students (blue)



Human competence is epistemic and relational

Effects of Feedback Provider and Timing on Ratings



AI competence training initiatives must move beyond procedural “how-to” training & adopt a holistic approach to develop competence across human-centred mindset, AI ethics, and the critical capacities.

UNESCO AI Competency Frameworks

Aspects	Progression		
	Acquire	Deepen	Create
Human-centred Mindset	Human agency	Human accountability	AI social responsibility
Ethics of AI	Ethical principles	Safe and responsible use	Co-creating AI ethics
AI Foundations & Applications	Basic AI techniques and applications	Application skills	Creating with AI
AI Pedagogy	AI-assisted teaching	AI-pedagogy integration	AI-enhanced pedagogical transformation
AI for Professional Development	AI enabling lifelong professional learning	AI to enhance organizational learning	AI to support professional transformation



AI competency framework
for teachers



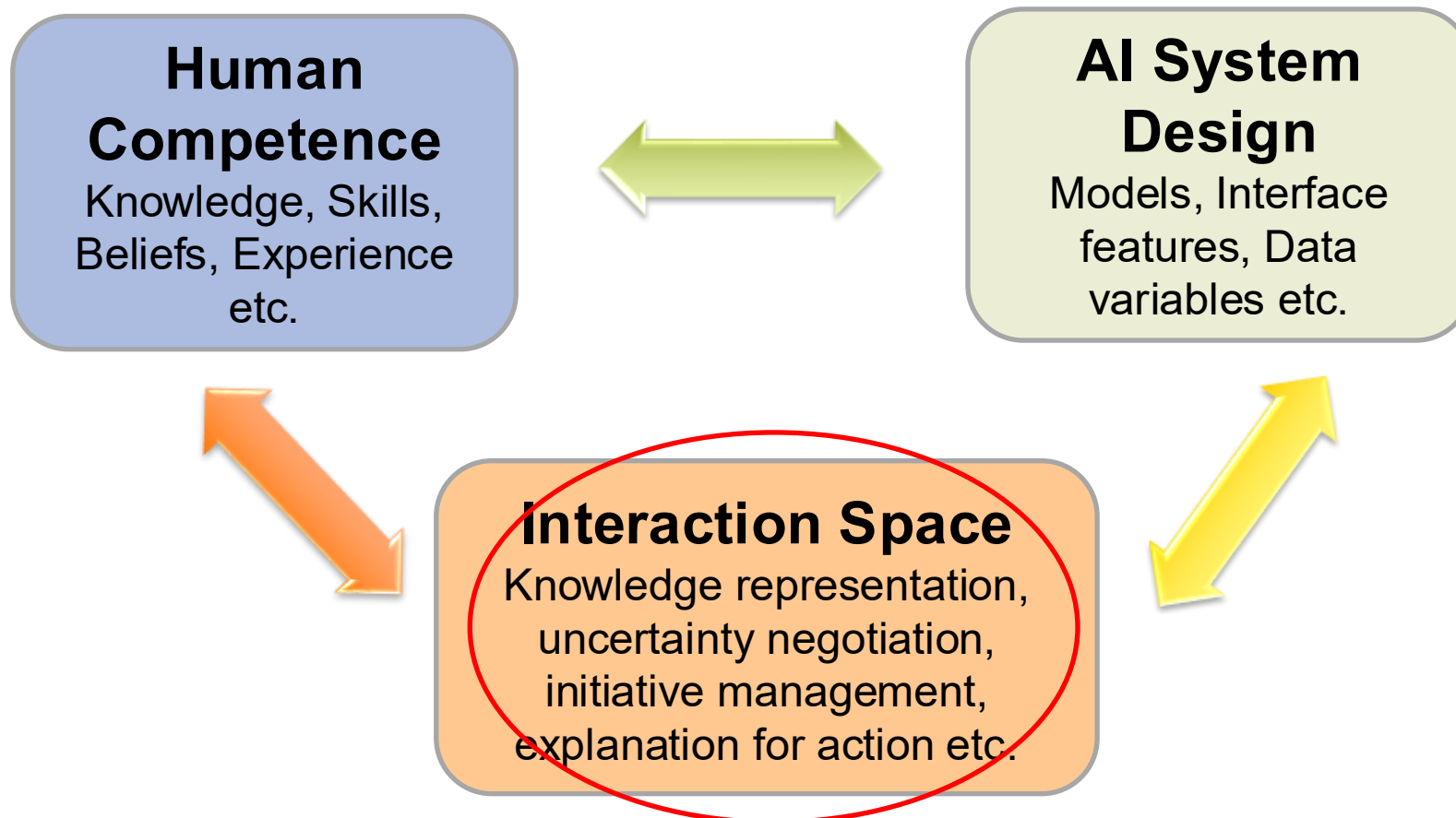
VISION ON GENERATIVE AI IN EDUCATION

All Flemish universities share the same principles:

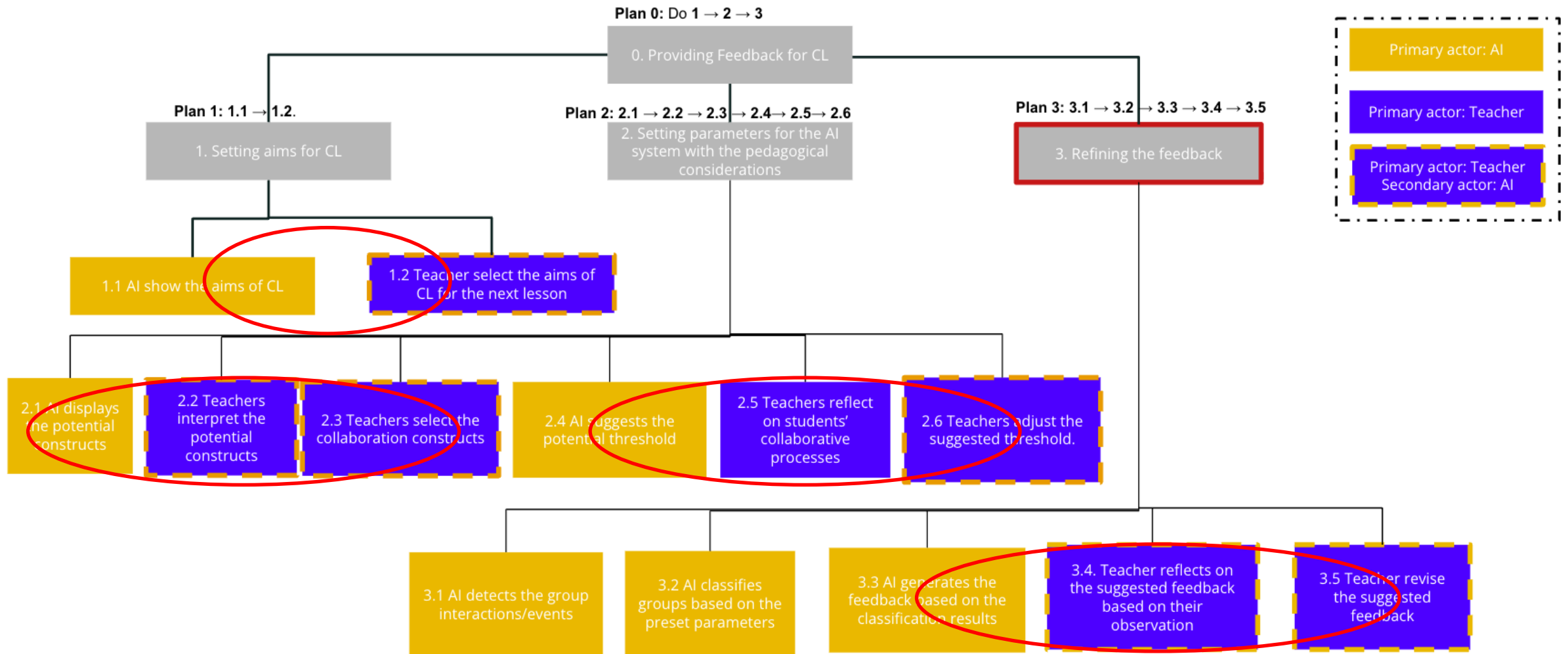
- **Critical & transparent use:** Awareness of limitations, errors, bias, plagiarism, and copyright; AI use disclosed when relevant.
- **Human responsibility:** Students and lecturers remain fully accountable for quality, accuracy, and learning outcomes.
- **Privacy & confidentiality:** No input of personal, sensitive, or copyrighted data without permission.
- **Ethical awareness:** Reflection on societal impact, bias, and unequal access to AI.
- **Purposeful use:** AI is used consciously and only when it adds educational value.
- **Education & quality assurance:** Development of AI competencies for staff and students, and alignment with assessment and curricula.
- **Continuous monitoring & innovation:** Policies evolve with technological developments, enabling responsible innovation and creativity.

From Replacement to Synergy:

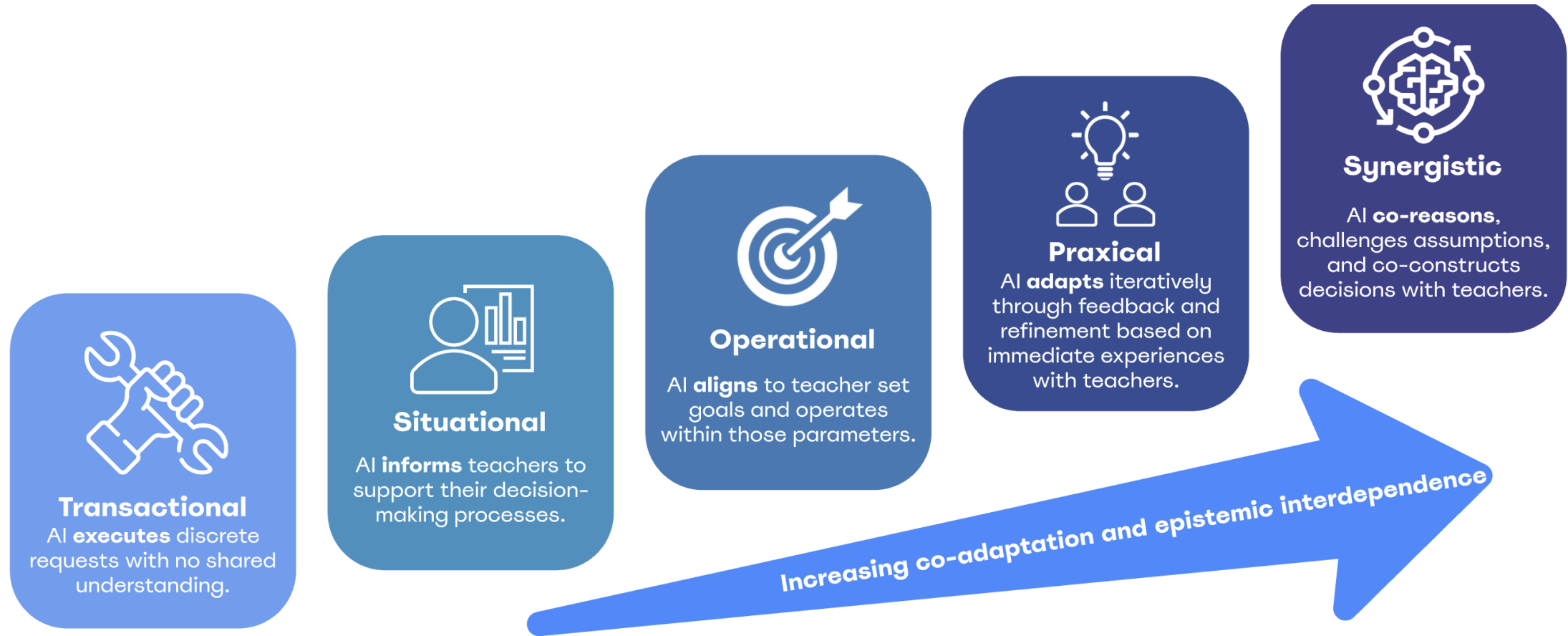
Designing Human–AI Systems that Grow Human Competence



Each human–AI touchpoint admits multiple teaming configurations



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



Transactional Teaming

Define feedback for students

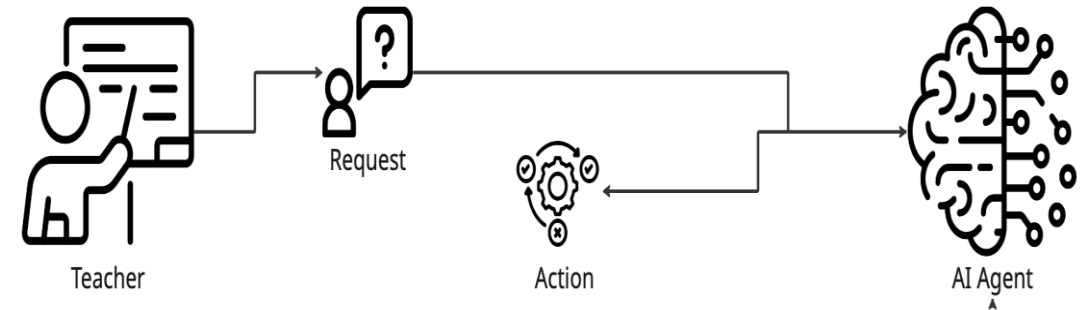
Define feedback for students:

Feedback for groups with silence time over: 15%

Notable silence time was observed during the session. While this may reflect your 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.



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%

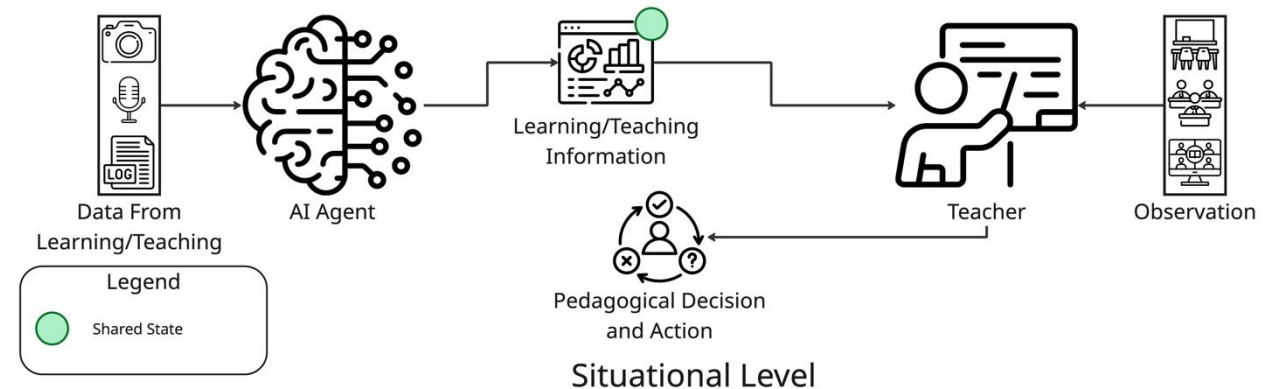
Notable silence time was observed during the session. While this may reflect your resource management, it may also suggest a low coordination through discussion. It is good to communicate with each other to align your group thinking.

AI-generated explanation

(evidence + theory)

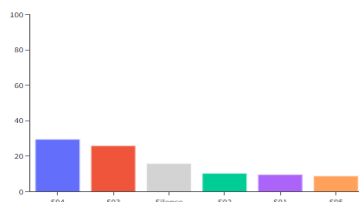
Evidence from the graph: the bar chart shows a non-trivial **Silence** portion relative to individual speaking bars, and speaking contributions are not evenly distributed (a small number of students account for a larger share of talk time). These patterns suggest limited group-level verbal exchange and/or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

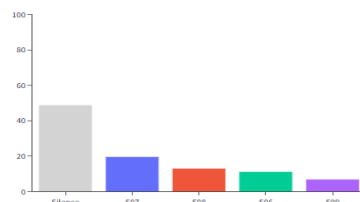


Control: AI-led detection, teacher-led interpretation
Teacher role: Transactional + Contextual interpreter
AI role: Process situational signals and create feedback based on model parameters
Competence growth: Increased situational awareness

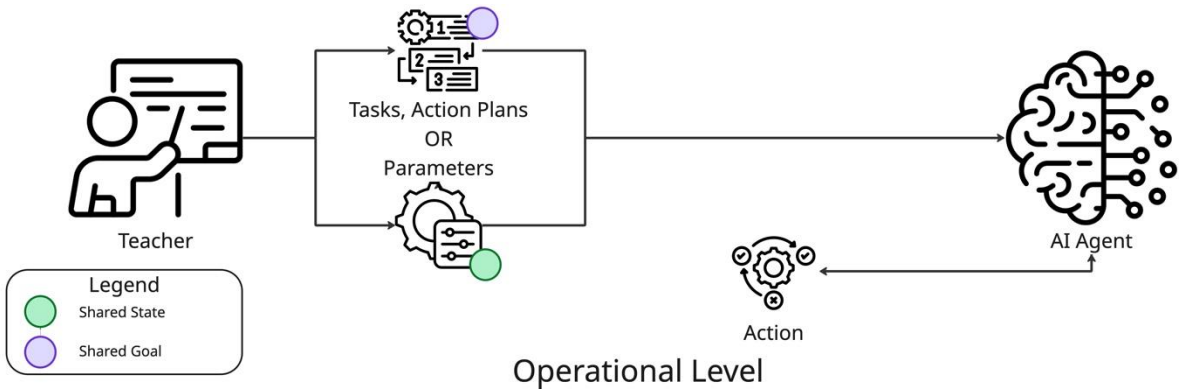
Group 1: ⚠️



Group 2: ⚠️



Operational Teaming



Area to explore

Area to explore:

Select a feature:

silence time

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

Select a type of visualisation:

bar chart

Select a week:

5

Alerting criteria

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 evaluation (LLM generated) (Overall: 3.8/5)

Specificity (context relation)	4/5
Why?	
Pedagogical grounding	3/5
Why?	
Encourages coordination but could reference collaborative learning principles more explicitly.	
Clarity / Conciseness	4/5
Why?	
Actionability	3/5
Why?	
Task alignment	4/5
Why?	
Supportive tone	5/5
Why?	

Control: Shared but teacher-directed.

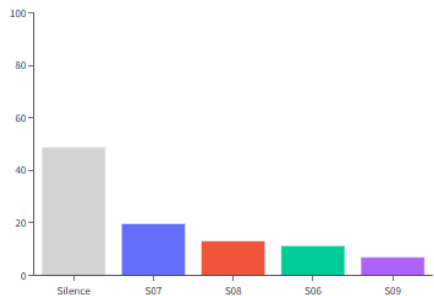
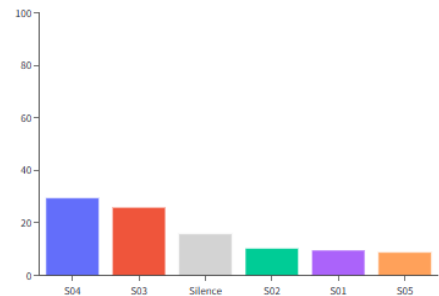
Teacher role: Situational + System configurator & pedagogical rule-setter.

AI role: Operationalise teacher input in model outputs

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

Group 1: ⚠️

Group 2: ⚠️



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

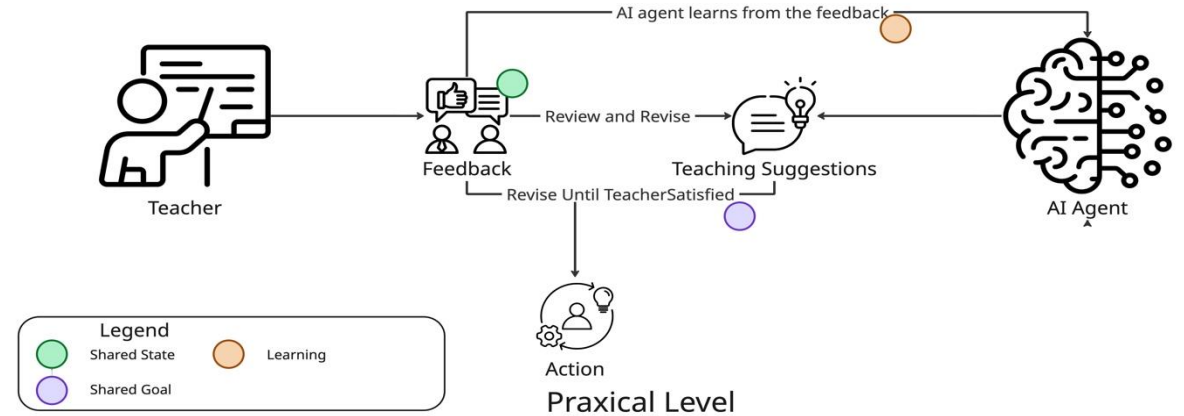
Clear History

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.



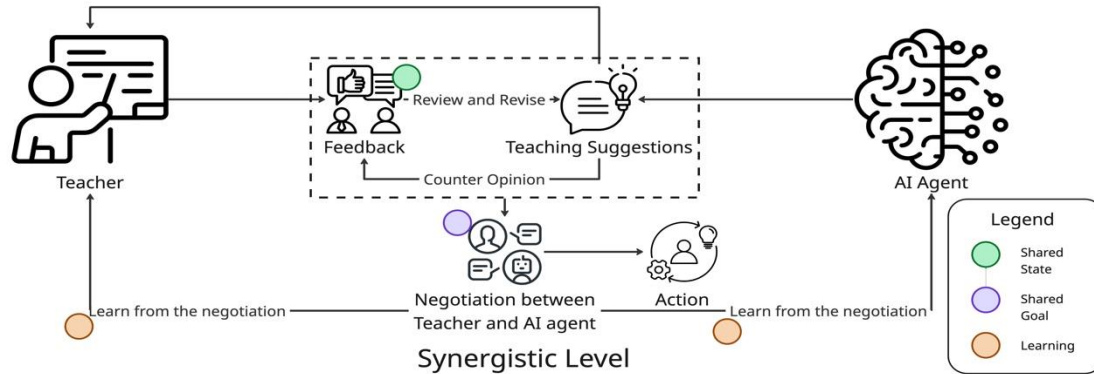
Control: Reciprocal and adaptive. Becomes co-regulated, teacher edits shape model behaviour, and model suggestions shape teacher reflection.

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%:

Your group's silence time exceeded 15% this week — while some thinking time is healthy, extended silence can indicate unclear roles or hesitation to contribute. 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](#)

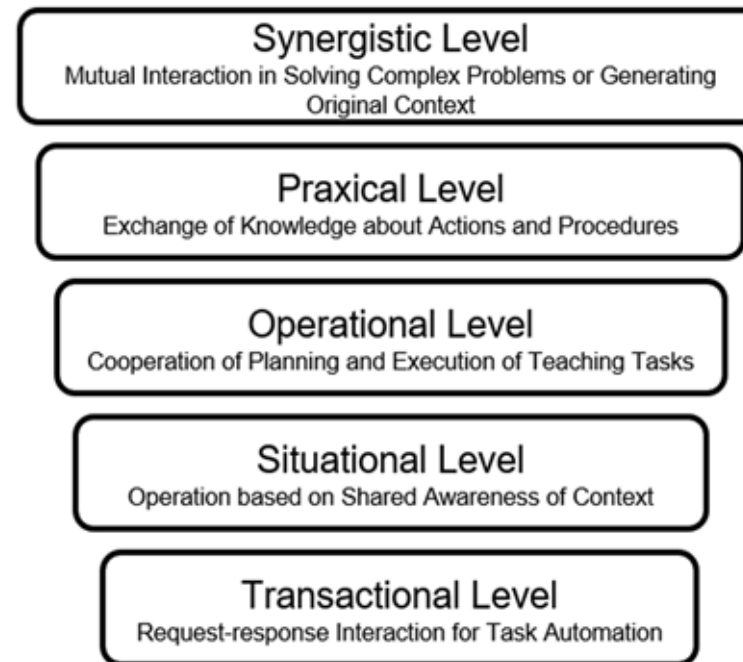
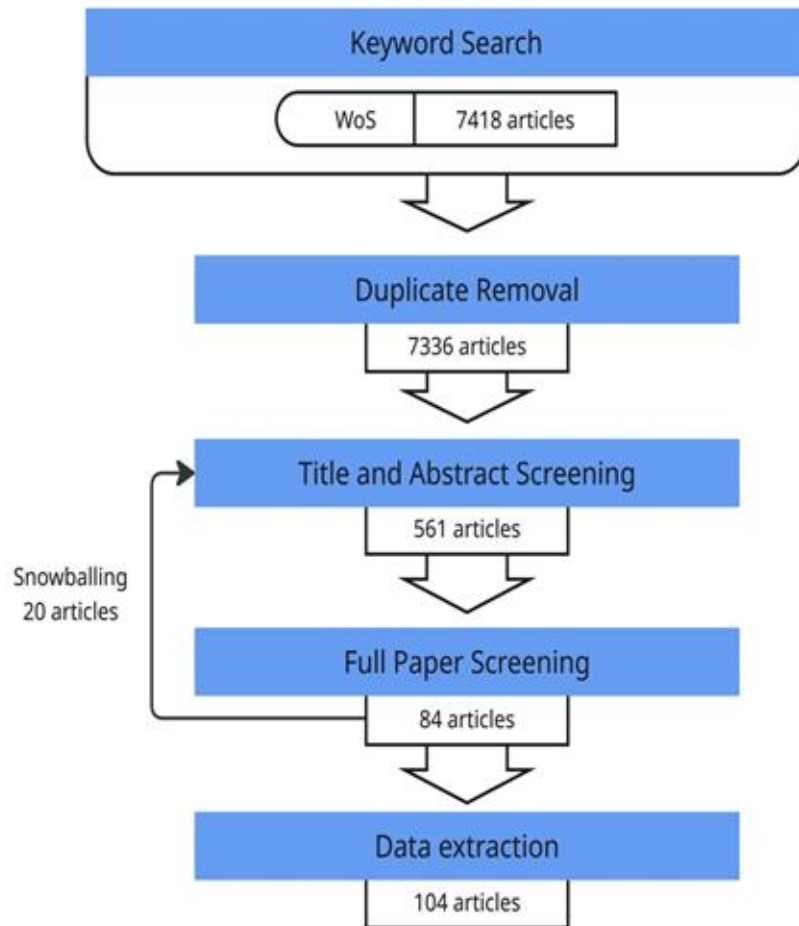
Discuss with AI

Click **Discuss with AI** to start a conversation. The AI will evaluate your feedback against the criteria and guide you through improvements. You can agree or disagree — the AI will incorporate your decisions into the updated feedback.

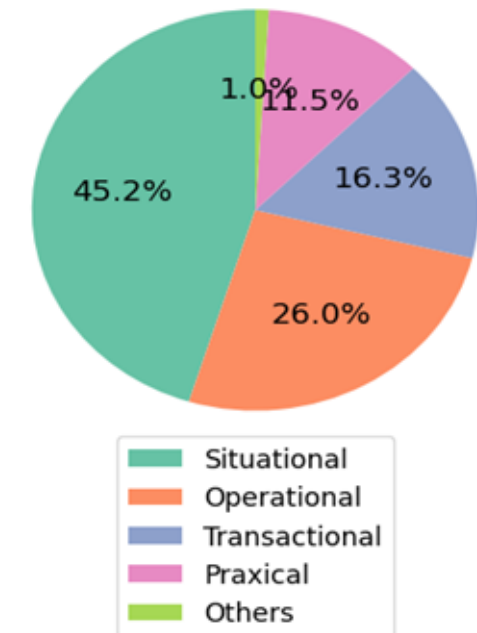
Automated simulation — no input required.

[Send](#)

Synergistic human–AI teaming remains largely absent in educational contexts

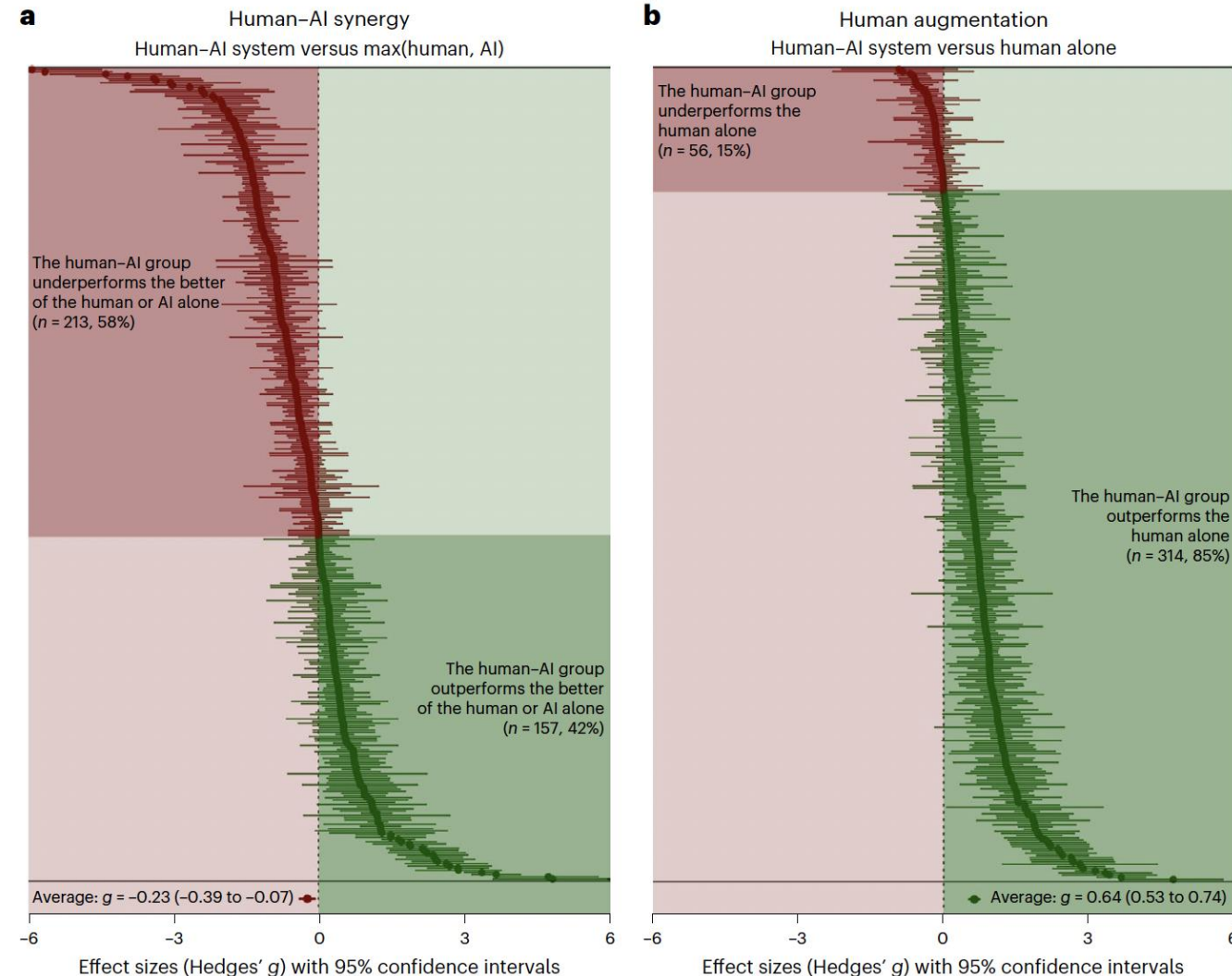


$$P_{H+AI} > \max(P_H, P_{AI}) \quad \frac{dC_H}{dt} > 0 \quad \frac{dC_{AI}}{dt} > 0$$

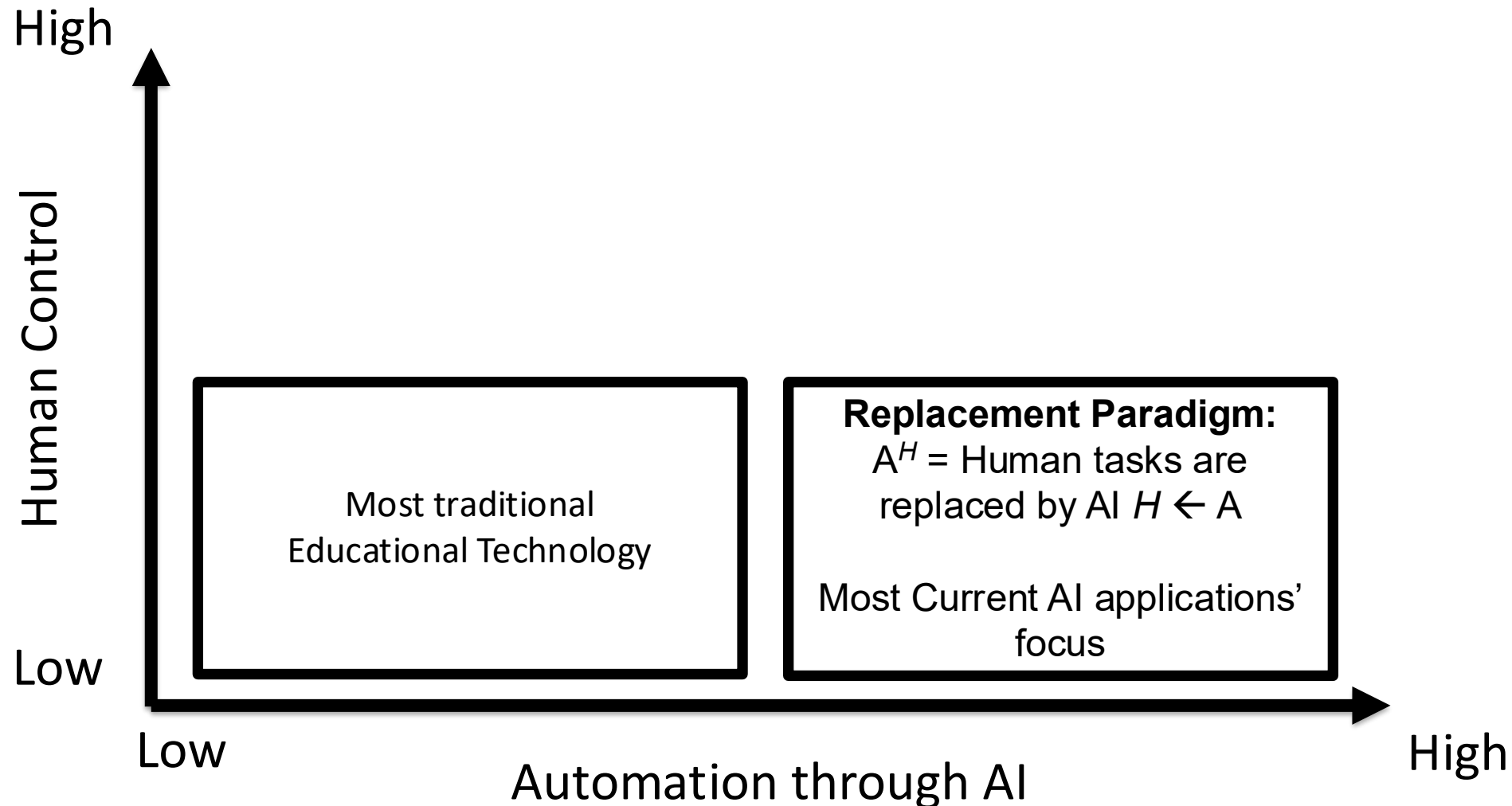


Even in more mature fields, most human–AI teaming lack synergy

- On average, human–AI teaming significantly better than the human alone (**complementarity**); yet performs **significantly worse** than the best of humans or AI alone (**synergy**).
- Task dependent:** Decision making vs content creation, for instance.

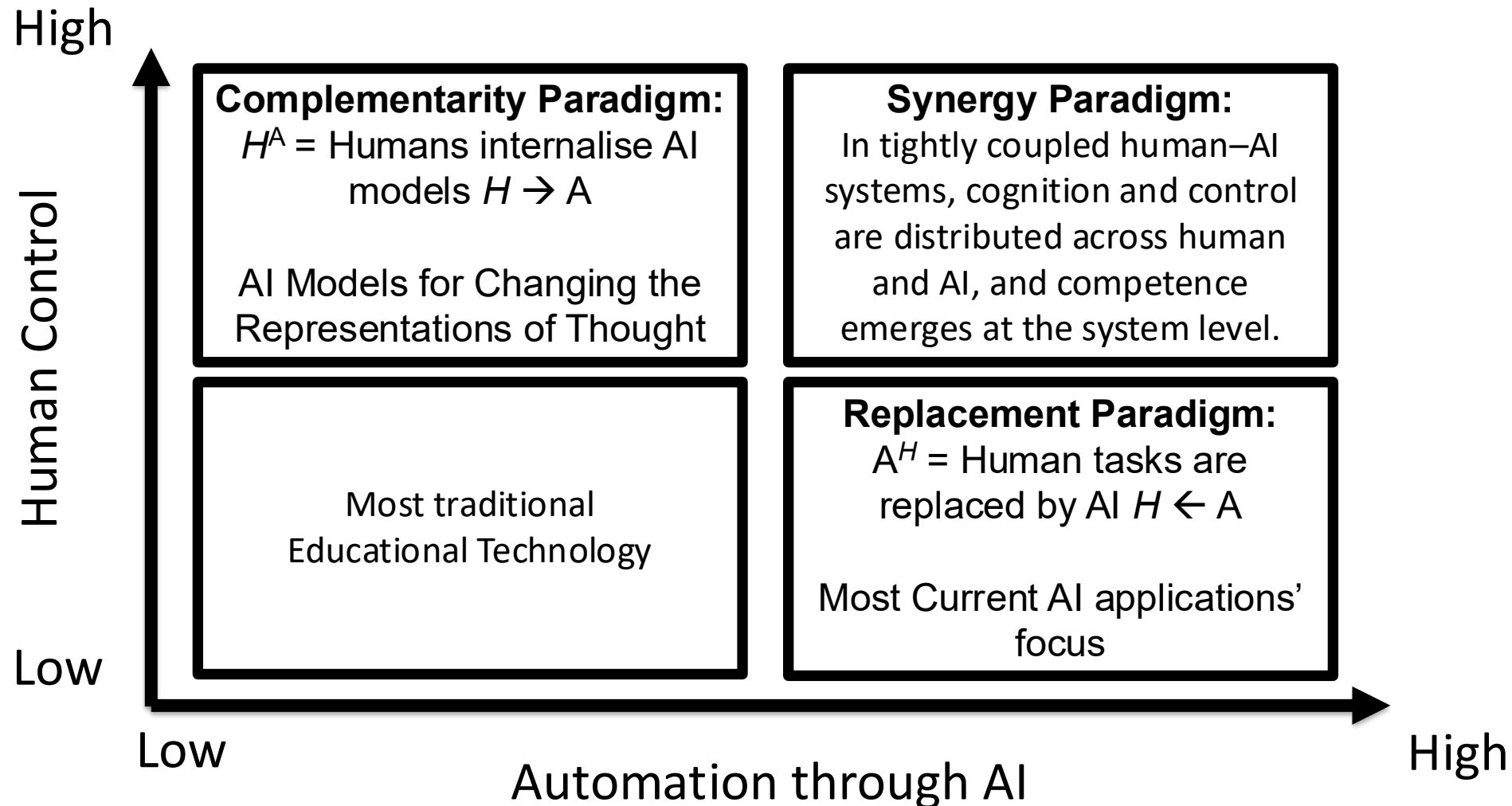


Three Conceptualisations of AI in Education



The future of AI in higher education is not necessarily in better AI models for replacement, but better design and development of human–AI teaming.

Three Paradigms of AI in Education



Synergy requires pedagogically designed AI, deliberate development of AI competence, and enabling conditions in the ecosystem.



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

