

Artificial Intelligence in Education

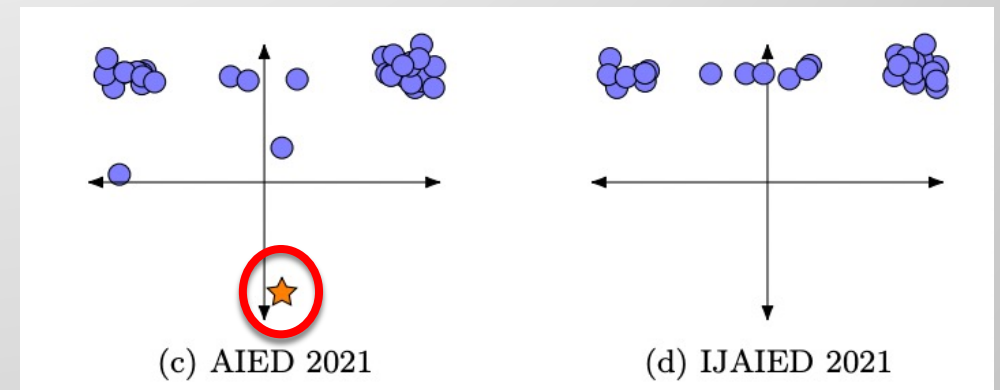
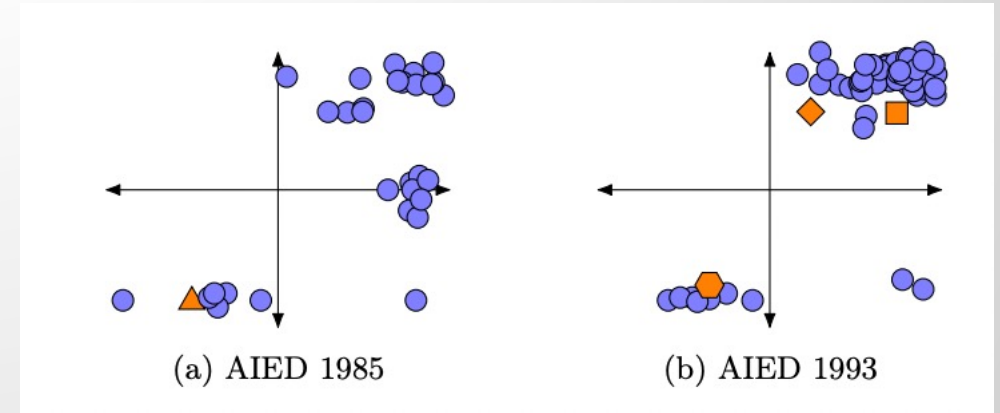
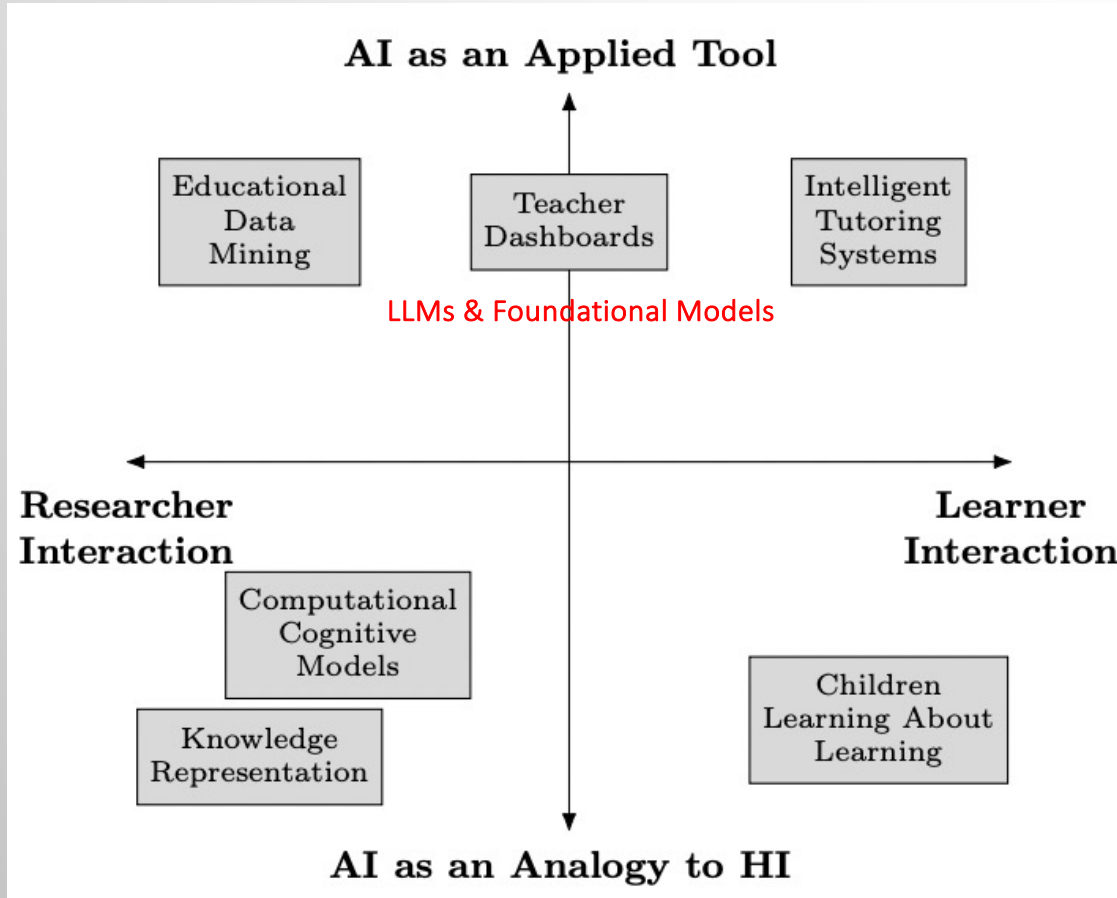
Professor Mutlu Cukurova
University College London

m.cukurova@ucl.ac.uk

@mutlucukurova



AI is More than Applied Tools



Conceptualisations of AI in Education

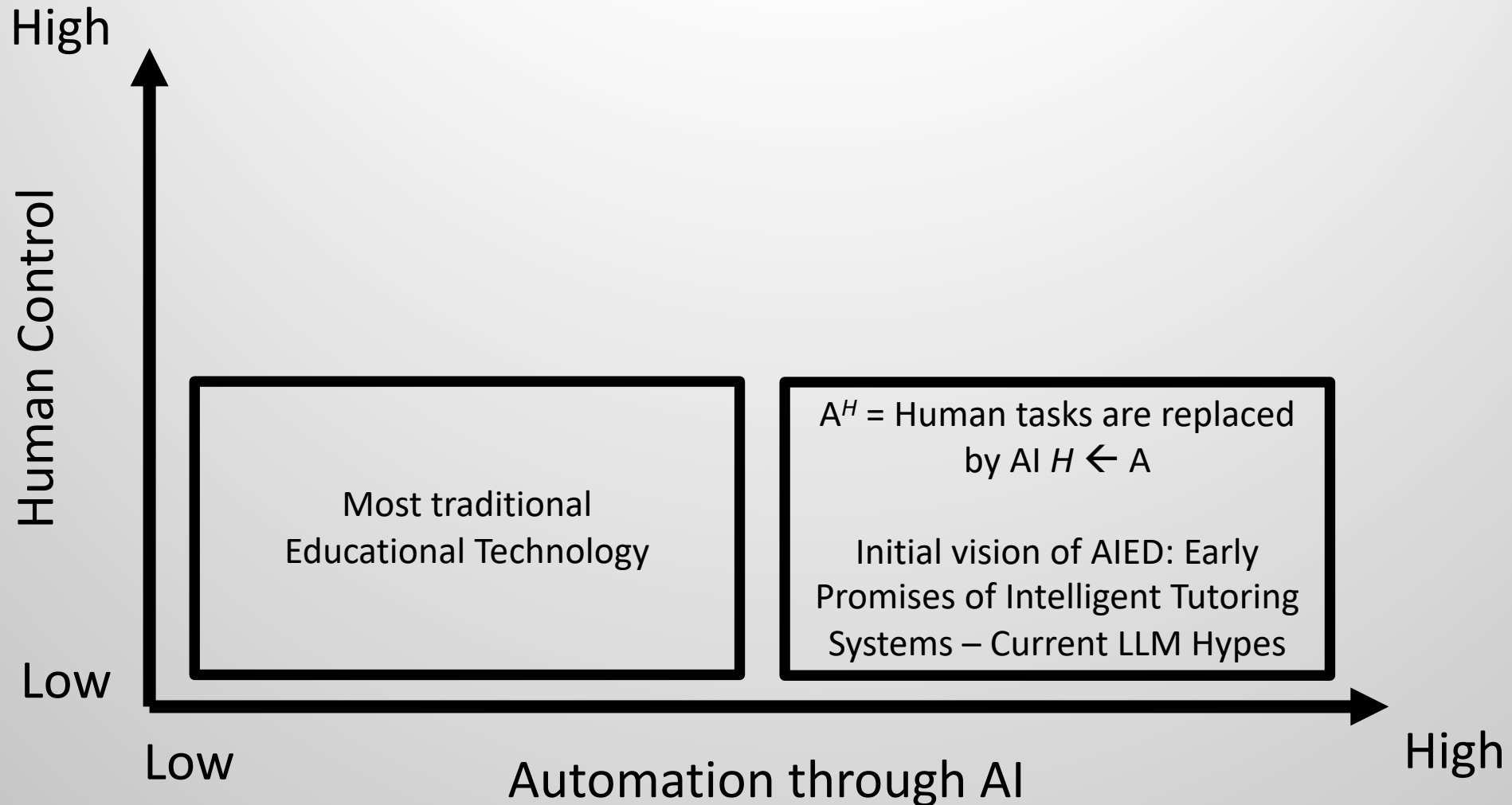
- AI can be conceptualised to externalize, be internalized or extend human cognition.
- A^H = Human tasks are replaced by AI $H \leftarrow A$
- H^A = Humans can internalise AI models $H \rightarrow A$
 - Changing the operations and representations of thought (GOFAI)
- $H[A]$ = Human (H) extended with an AI (A), **tightly coupled human and artificial systems.**
- $H[A] \neq H + A$
 - The whole should be more than the sum of its parts.
 - Change in H, perhaps also in A, is observed.

Cukurova, M. (2019). Learning Analytics as AI Extenders in Education: Multimodal Machine Learning versus Multimodal Learning Analytics. *Proceedings of the Artificial Intelligence and Adaptive Education Conference*, xx1-xx3.

Cukurova, M. (2024). The Interplay of Learning, Analytics, and Artificial Intelligence in Education. arXiv preprint arXiv:2403.16081.

<https://doi.org/10.48550/arXiv.2403.16081>

AI in Education: A vision for the future



The most common applications of AI in Education were focusing on pedagogical task automation with ITSs

Cognitive Algebra I Tutor has multiple tools
www.carnegielearning.com

Problem

A rock climber is currently on the side of a cliff 57 feet off the ground. She can climb on average about two and one-half feet per minute.

- 1 When will she be 92 feet off the ground?
- 2 In twenty minutes, how many feet above the ground will she be?
- 3 In 75 seconds, how far above the ground will she be?
- 4 Ten minutes ago, how far above the ground would she be?

Step: Label a column

Quantity	CLIMBING TIME	HEIGHT
Unit	MINUTES	FEET
Expression	T	$57 + 2.5T$
Question 1	10	92
Question 2	20	117
Question 3	1.25	79.125
Question 4	-10	42

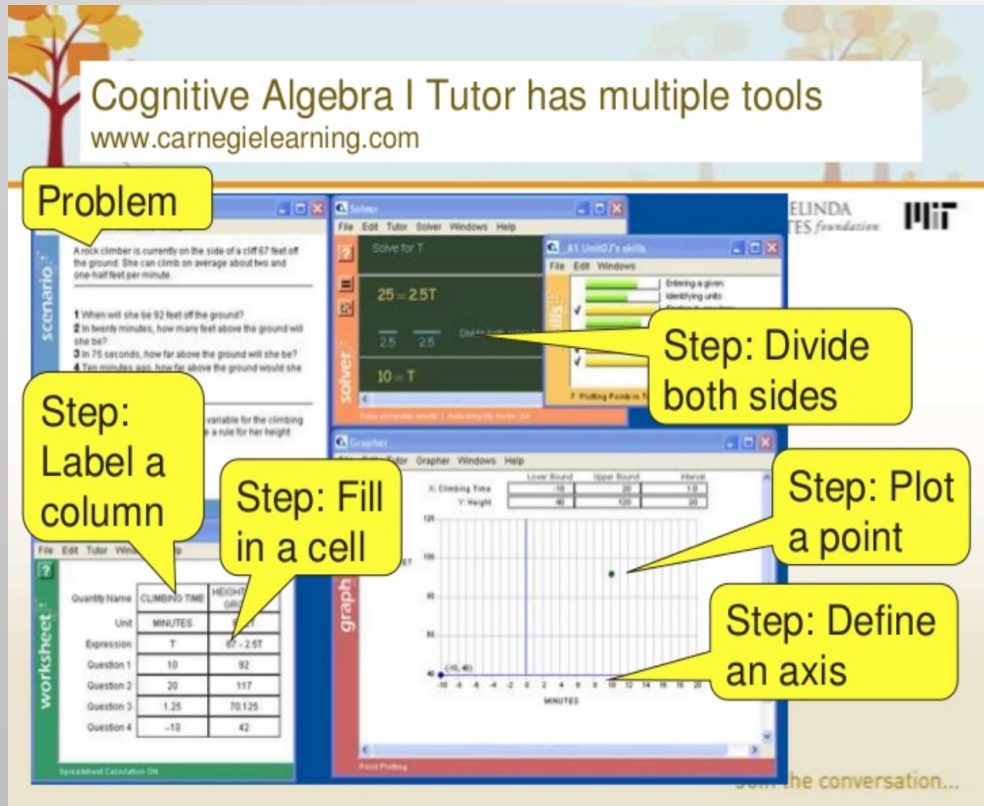
Step: Fill in a cell

Step: Divide both sides

$$25 = 2.5T$$
$$2.5 = 2.5$$
$$10 = T$$

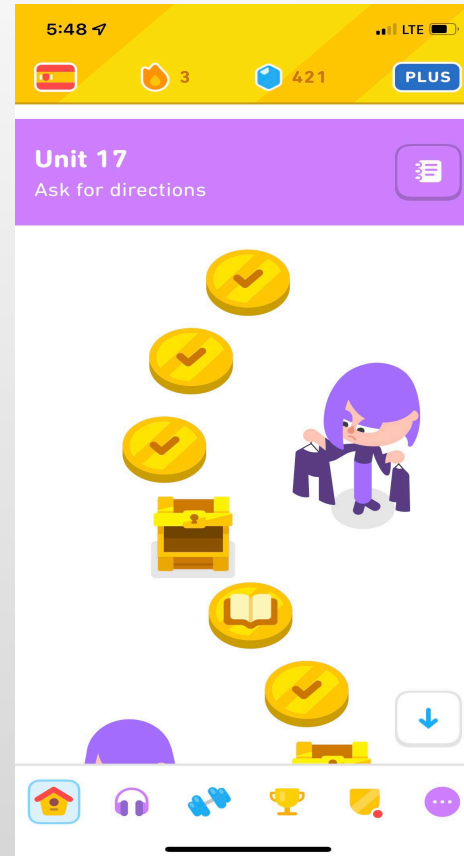
Step: Plot a point

Step: Define an axis



5:48 LTE

Unit 17
Ask for directions



duolingo

Khanmigo

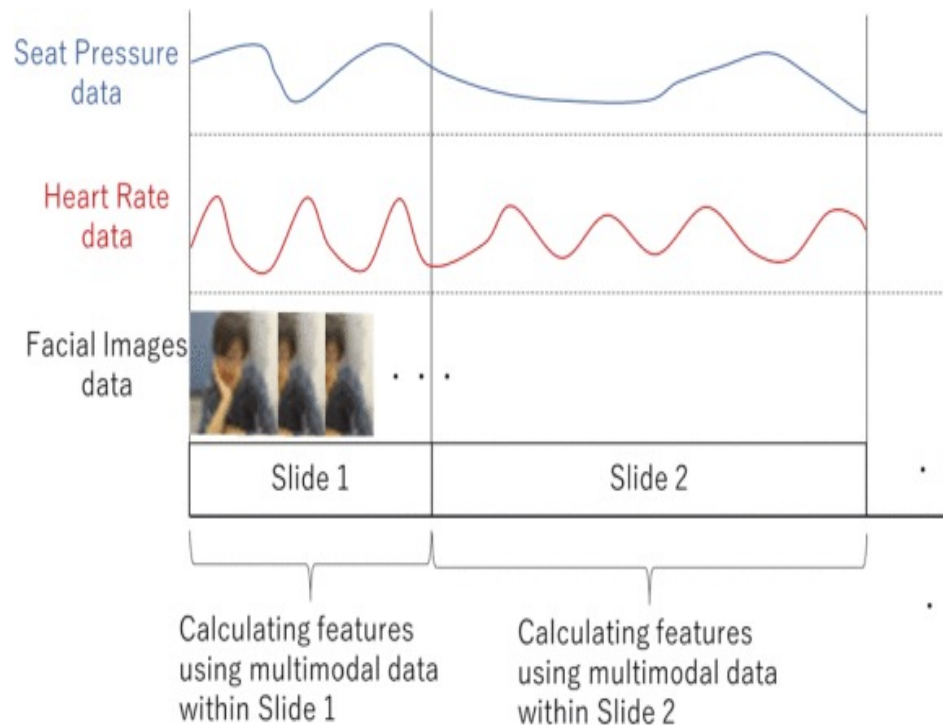


What topic would you like to explore in your lesson plan? Try something like "Add fractions with common denominators"

Amazing, I'd like to includes topics that cover surface area word problems

Type message

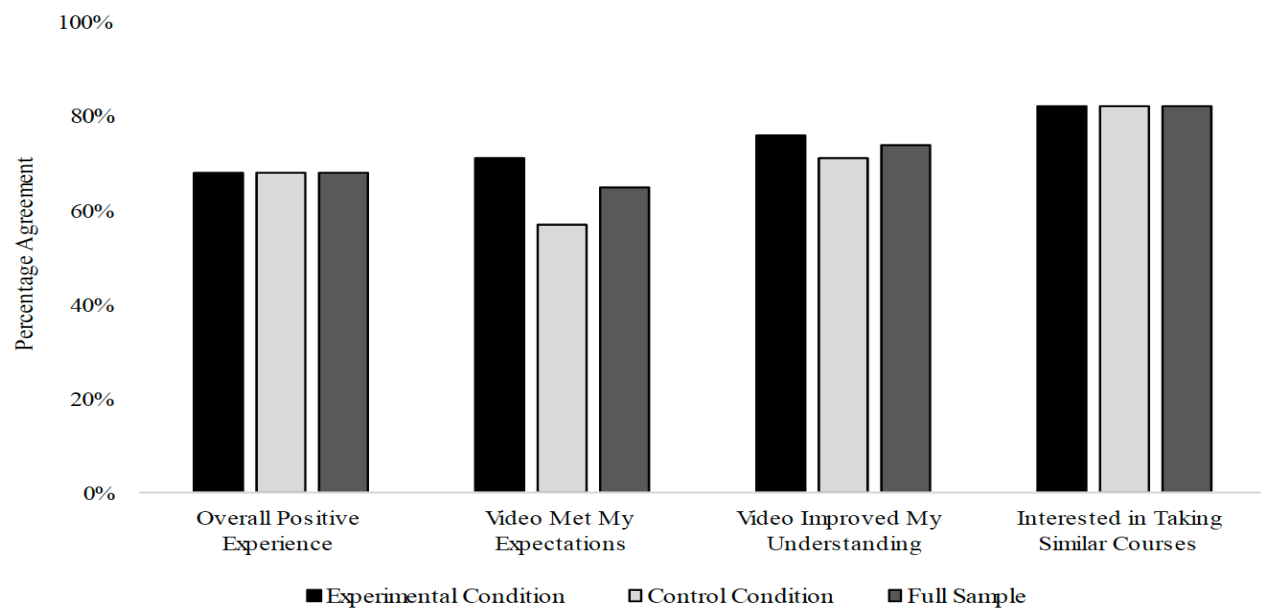
Detecting Drowsy Learners at the Wheel of e-Learning Platforms With Multimodal Learning Analytics



Features	Metrics	Constructs Represented
Heart rate (HR)	Mean/ standard deviation of RRI, HR, LF/HF, Body surface temperature	Heart rate in general represents the activity of the autonomic nervous system. RRI is an index of heart rate variability. HR is the heart rate. Low frequency (LF) power and high frequency (HF) power represent stress and rest states. Body surface temperature is environmental temperature in clothing.
Seat Pressure (SP)	Mean pressure	Mean of each frame's total pressure and mean of pressure per second. They are used to estimate a learner's motions.
	Mean time of MS (moving state) and SS (static state)	Represents how long a learner moves or stays still.
	Ratio of MS (moving state)	Represents how often a learner changes posture.
Facial Expression (Face)	Mean of absolute pressure difference between pressure current and previous frame.	Represents how large and how often a learner changes posture along vertical axis.
	Mean/ standard deviation of AU 2, 15, 26, 45 (occurrence and intensity)	AU2: Outer Brow Raiser, AU15: Lip Corner Depressor, AU26: Jaw drop, AU45: Blink.
	Mean/ standard deviation of head rotation (yaw, pitch, roll)	Represents how large a learner's head rotation is.
	Mean/ standard deviation of head transition along x, y, z	Represents how large a learner's head transition is.

Investigating the potential of AI-generated synthetic learning videos

	Pre-Learning M (SD)	Post-Learning M (SD)	Knowledge Gains M (SD)	p (d)
<i>Experimental</i>	0.45 (0.61)	1.45 (0.95)	1.00 (1.04)	< .001 (0.96)
<i>Control</i>	0.66 (0.70)	1.16 (0.94)	0.94 (1.13)	< .001 (0.83)
Full Sample	0.59 (0.66)	1.55 (1.03)	0.96 (1.11)	< .001 (0.91)



- Condition (experimental vs. control) was not a significant predictor of knowledge gains ($\beta = .03$, $p = .80$, $r = .03$).
- Even when controlled for participants' pre-learning performance ($\beta = -.03$, $p = .79$, $r = .03$) or their self-reported prior knowledge ($\beta = .01$, $p = .92$, $r = .01$).

Evidence of Impact of Intelligent Tutoring Systems

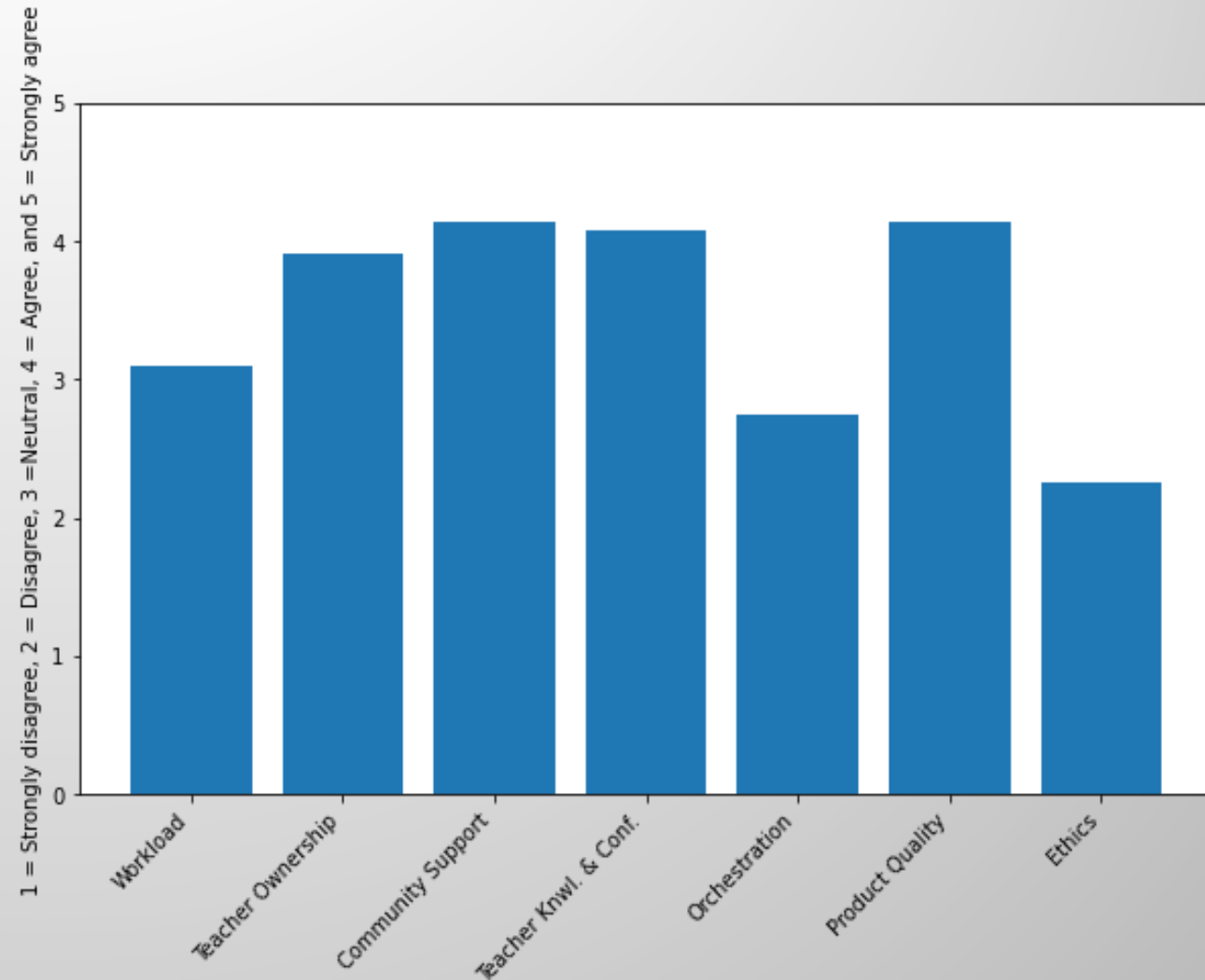
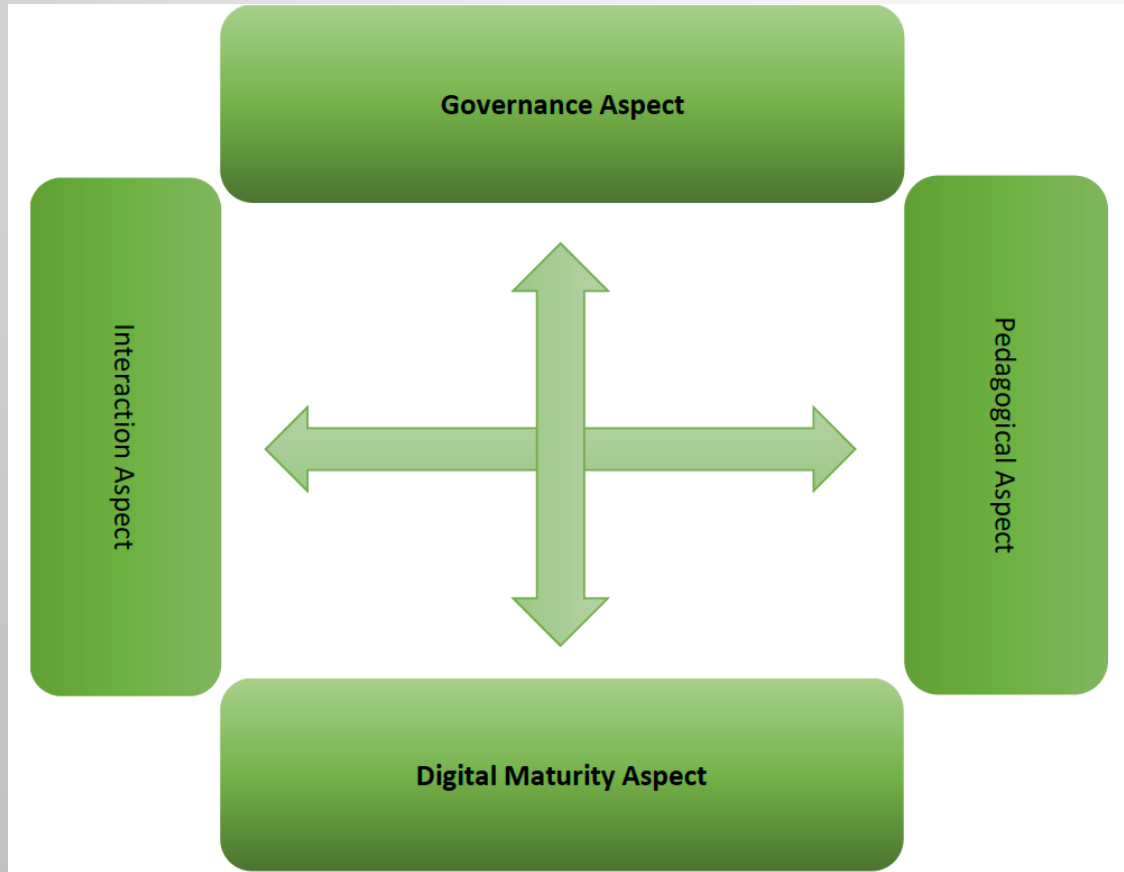
- ITSs can have positive impact on student learning : OLI learning course (Lovett et al., 2008), SQL-Tutor (Mitrovic, & Ohlsson 1999), ALEKS (Craig et al. 2013), Cognitive Tutor (Pane et al. 2014), ASSISTments (Koedinger et al. 2010).

Meta-reviews

- VanLehn (2011) found that the effectiveness of the intelligent tutoring systems were nearly as effective as average human tutors.
- Ma et al. (2014) found similar results both when compared to a no tutoring or to large group human-tutor instruction.
- Pane et al. (2014) found evidence of the relative effectiveness of online tutors over conventional teaching.
- Kulik & Fletcher (2016) median effect was to raise test scores 0.66 standard deviations over conventional levels, or from the 50th to the 75th percentile.
- du Boulay, B. (2016) summary of the metareviews in “Artificial Intelligence As An Effective Classroom Assistant”.

Despite significant advancements in AI and evidence supporting its effectiveness as ITSs, why AI is not prevalent in mainstream education?

Adoption of AI in Education is an Ecosystem Issue



AI might be Considered to Dehumanise Learning and Education

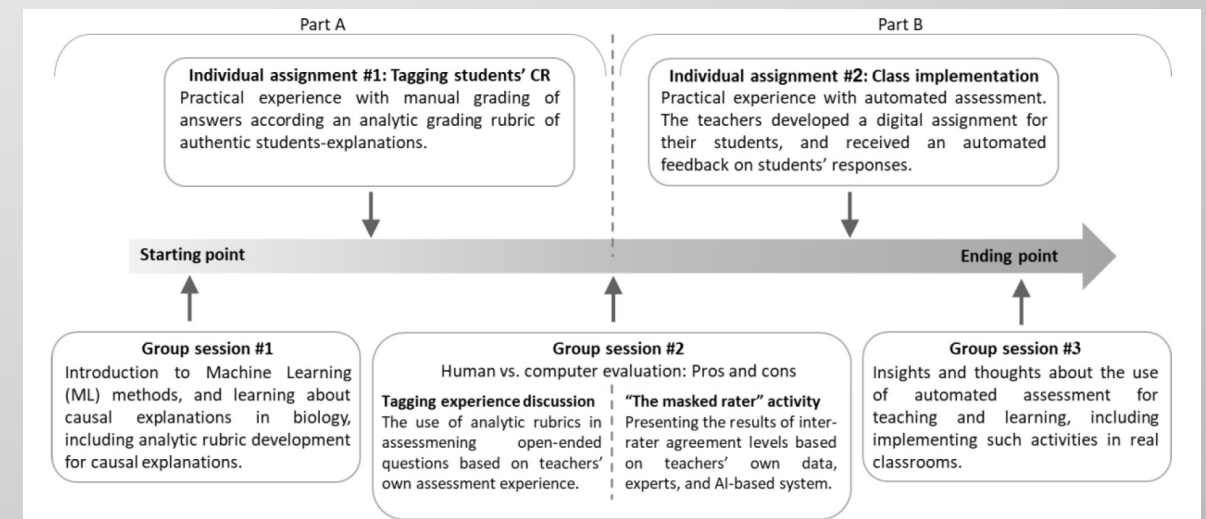
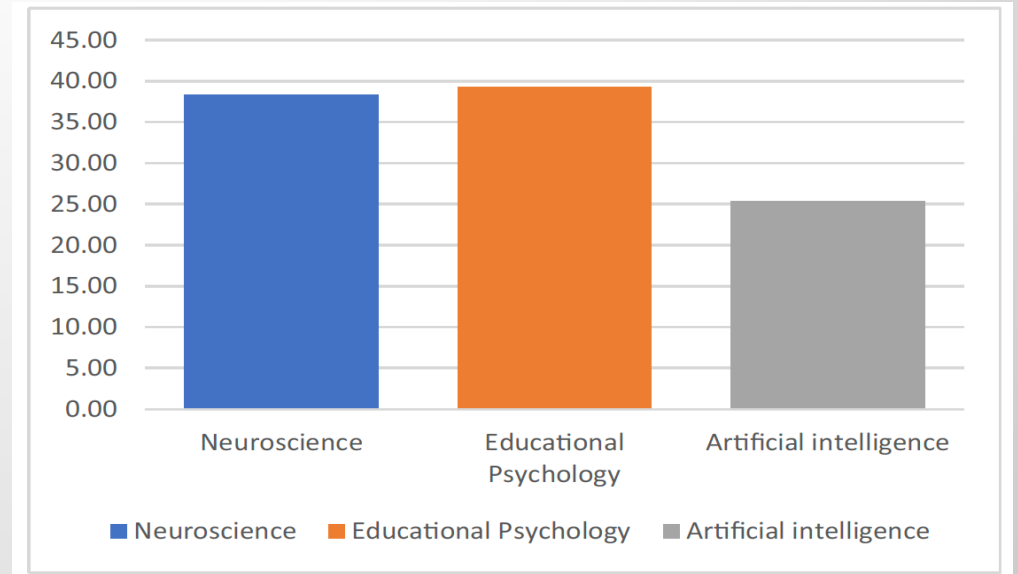


Trust in AI-EdTech is a Significant Concern

Teachers and learners have confirmation biases and unrealistic expectations from AI-EdTech.

"AI framing effect": when people are presented with content framed as coming from AI, they tend to judge it as less credible compared to educational psychology and neuroscience.

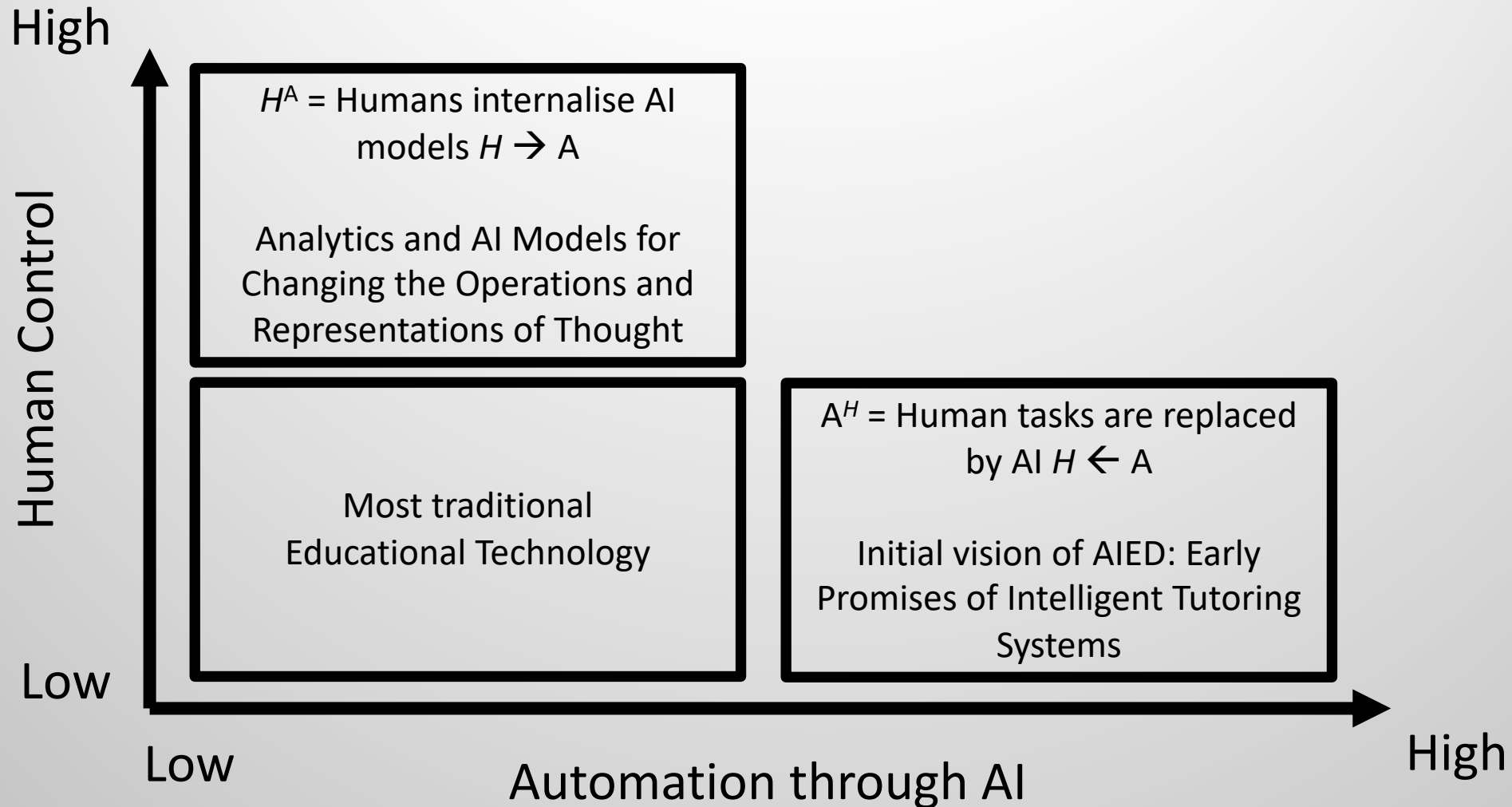
Considerable research is needed to gain end users' critical trust in AI-EdTech.



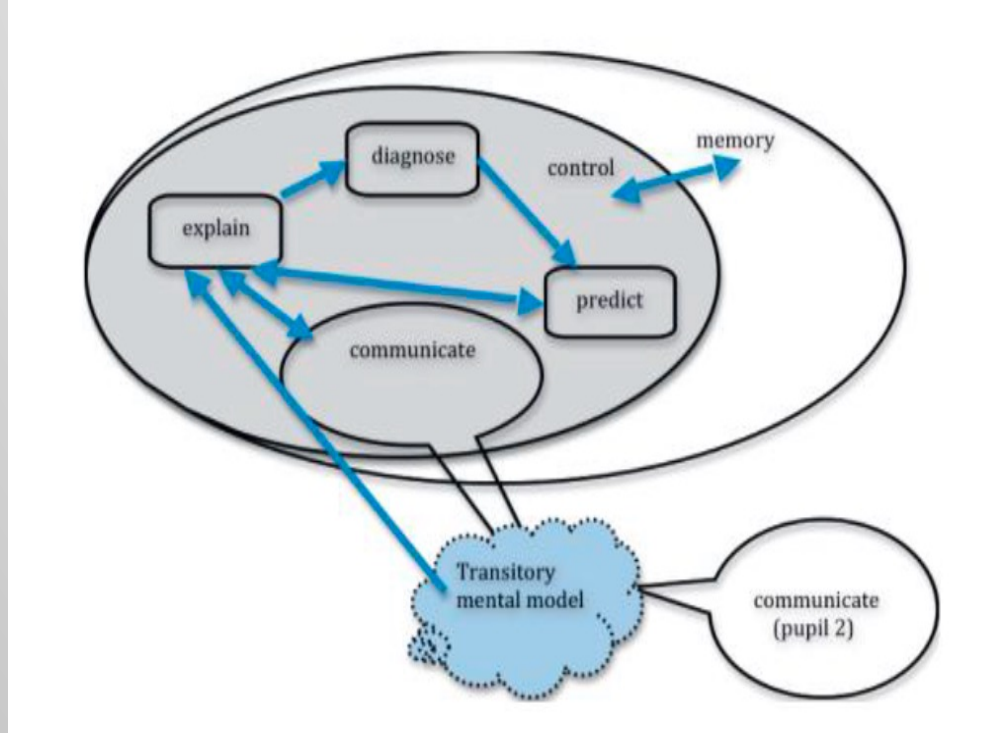
Cukurova, M., Luckin, R., & Kent, C. (2020). Impact of an Artificial Intelligence Research Frame on the Perceived Credibility of Educational Research Evidence. *International Journal of Artificial Intelligence in Education*, 1-31.

Nazaretsky, T., Ariely, M., Cukurova, M., Alexandron, G. (2022). Teachers' Trust in AI-powered Educational Technology and a Professional Development Program to Improve It, *British Journal of Educational Technology*, DOI: 10.1111/bjet.13232

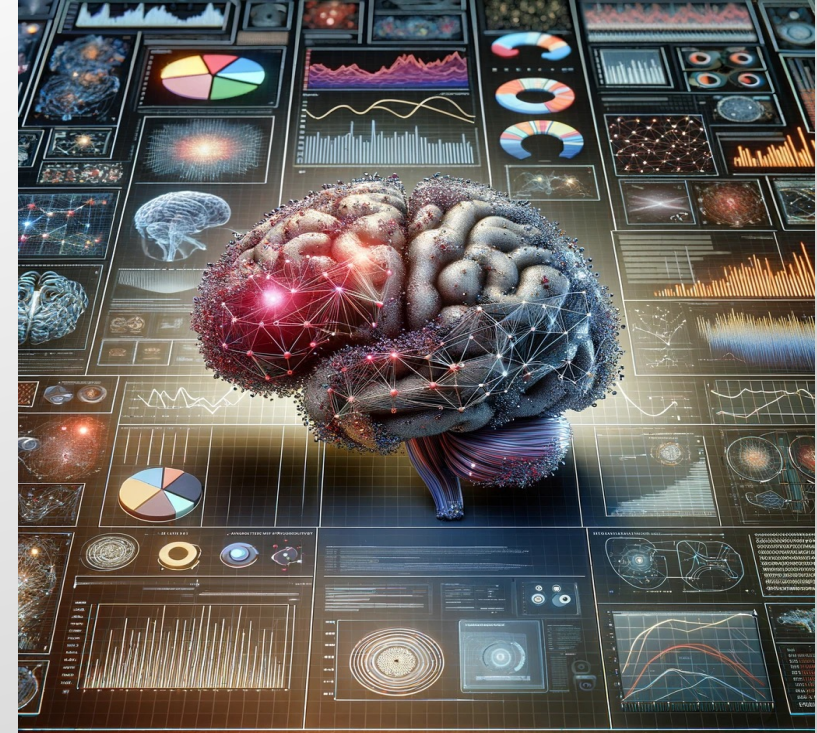
AI in Education: A vision for the future



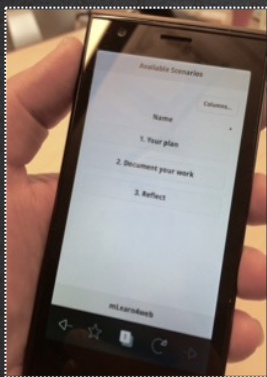
AI Models as Learning Affordances for Humans



Mental model mode: diagram of functionality



Computational and Statistical Models of Learners and Learning Processes



💡 → YAY! I GOT IT !!!
⚡ → ARGH! THAT'S HARD!!



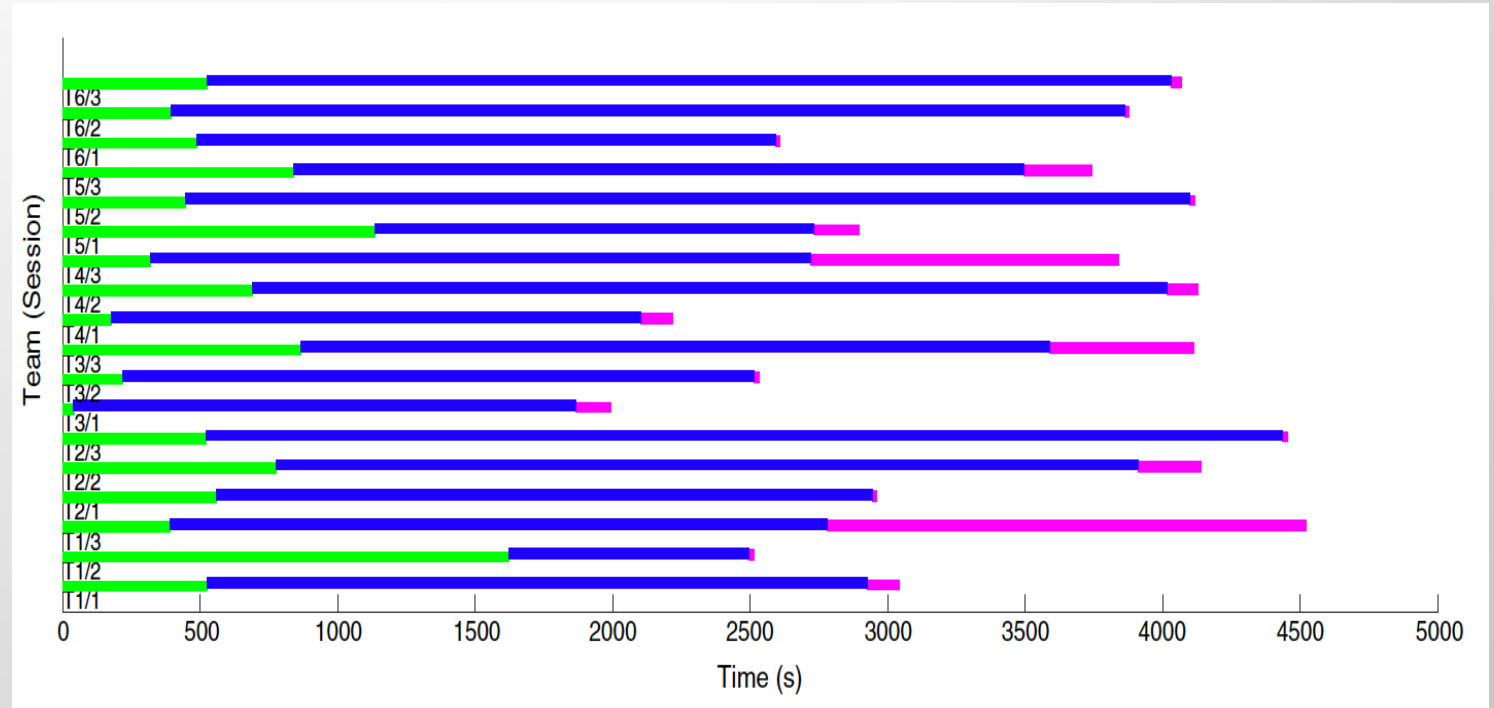
A^H = Human tasks are replaced by AI $H \leftarrow A$

Independent Variables (MMLA Features)

FLS - Number of faces looking at screen
DBL - Mean distance between learners
DBH - Mean distance between hands
HMS - Mean hand movement speed
AUD - Mean audio level

IDEX - Arduino measure of complexity
IDEVHW - Arduino active hardware blocks
IDEVSW - Arduino active software blocks
IDEC - Arduino active blocks

PWR - Student Work Phases



Ground Truth: Expert labelling of video data using CPS frameworks

Machine Learning Classification of CPS Competence

Method	Deep learning	Traditional
Task	Regression	Classification
Input	18 variables	9 variables per window
Output	6 scores over 5 levels	1 score with 3 levels
Metrics	Regression score	Classifier accuracy
Windowing	120,240 and 360 s	10,20,30,90 min
Phase exclusion	Reflection	Reflection
Method	Multiple layers	NB, LR, SVML, and SVMR

Note. NB = naive Bayesian; LR = logistic regression; SVML = support vector machines with linear kernel; SVMR = support vector machines for regression.

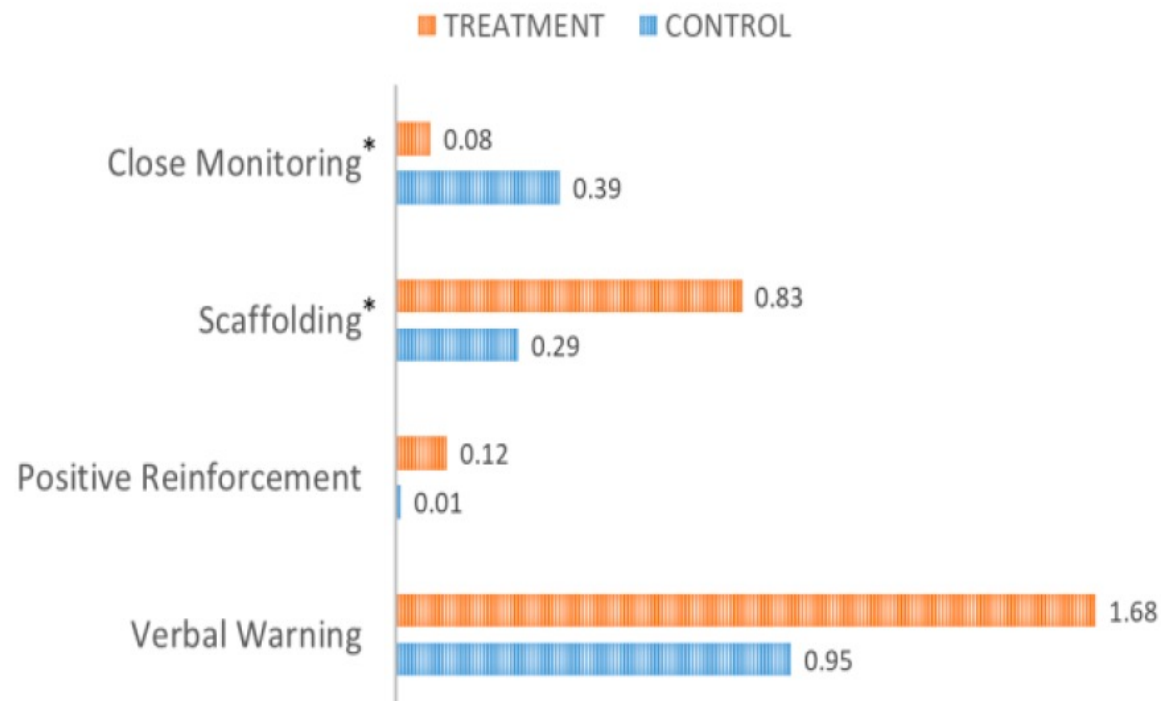
	PWR	PW	W	WR
NB	0.8	0.8	0.6	0.75
SVML	0.6	0.75	0.75	0.8
SVMR	0.75	0.75	0.75	0.75
LR	0.6	0.75	0.5	0.6

Note. NB = naive Bayesian; LR = logistic regression; SVML = support vector machines with linear kernel; SVMR = support vector machines for regression.

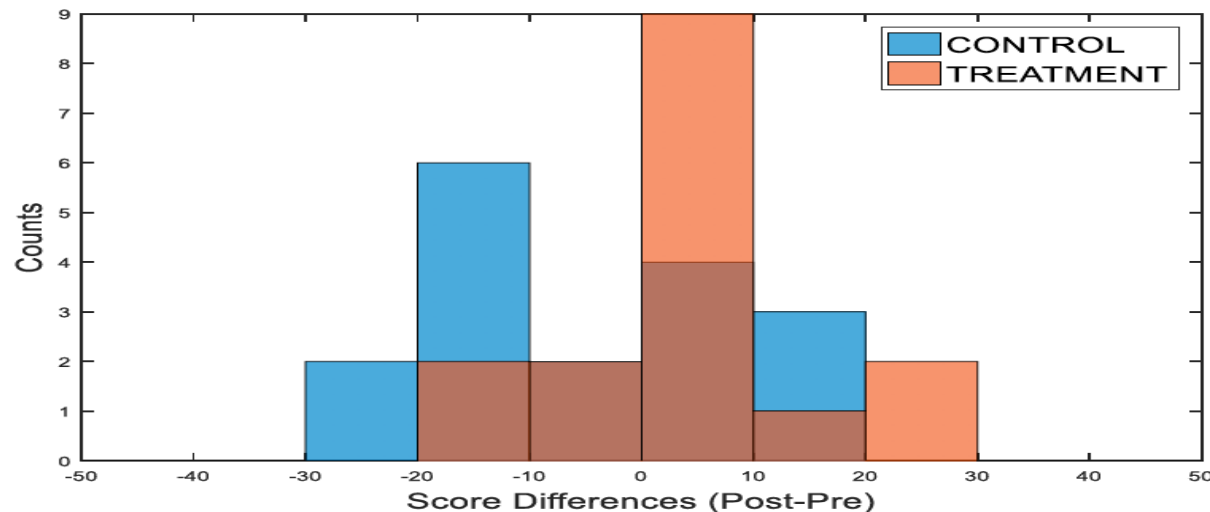
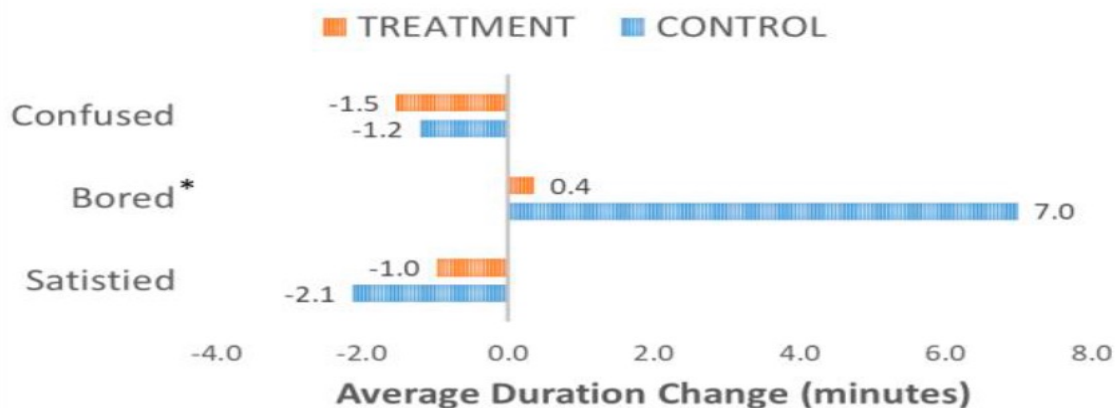
Removed feature	Best result
No features removed	0.129
All faces data	0.21
All Arduino data	0.21
DBF	0.15
DBH	0.21
HMS	0.19
AUD	0.18
Hand pos	0.21
Arduino comp	0.19



TEACHER INTERVENTION TYPES

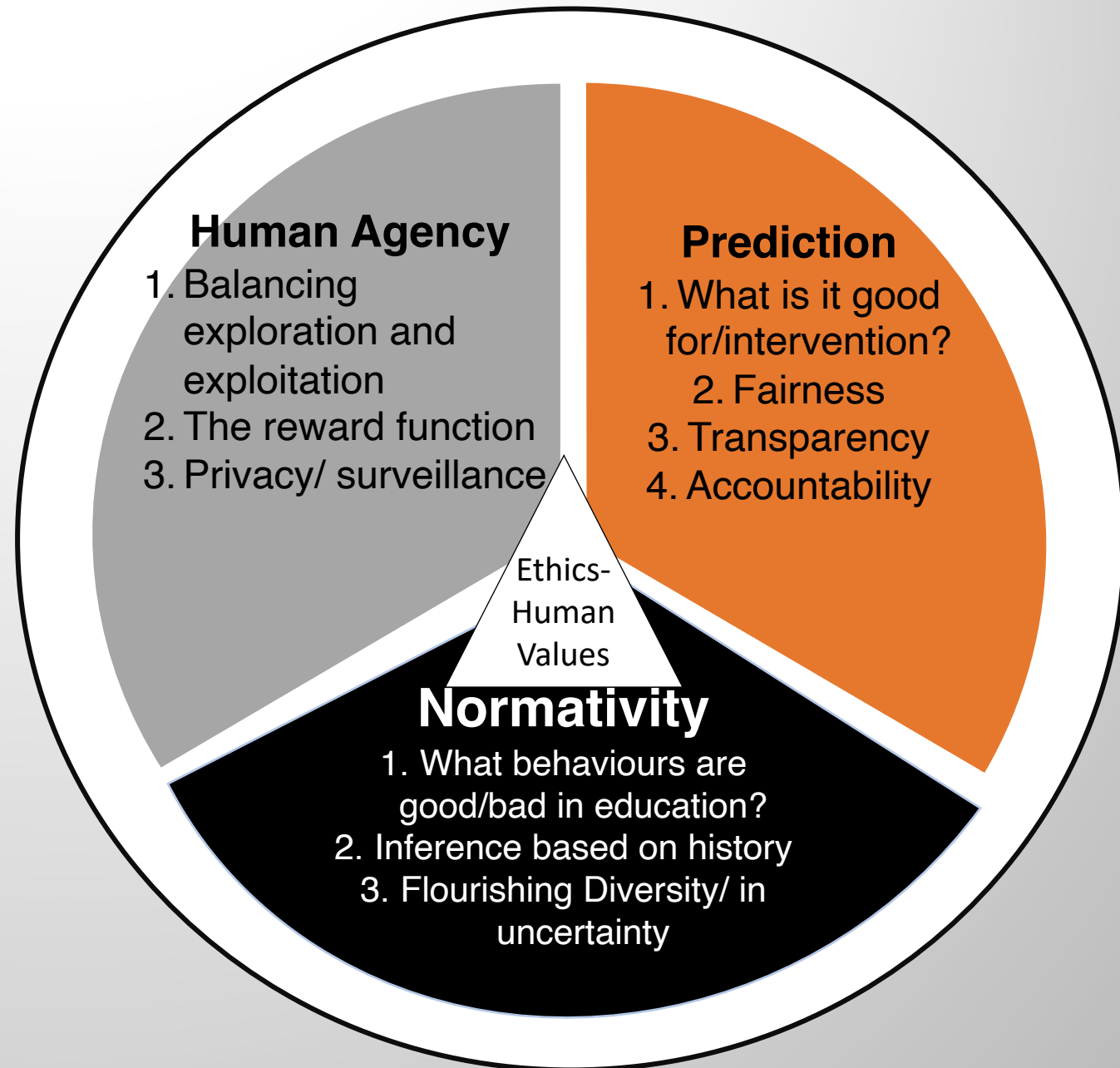


EMOTIONAL DURATION CHANGES

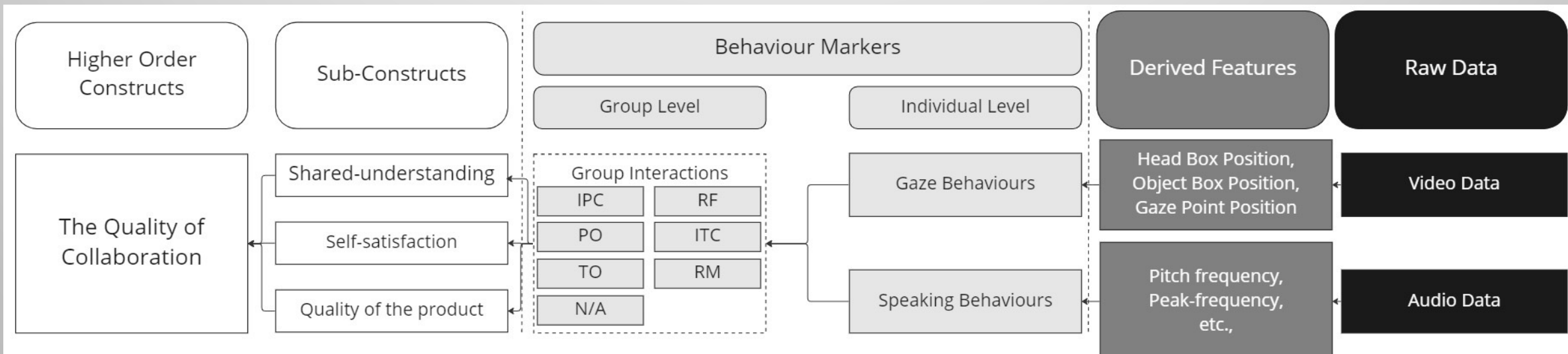
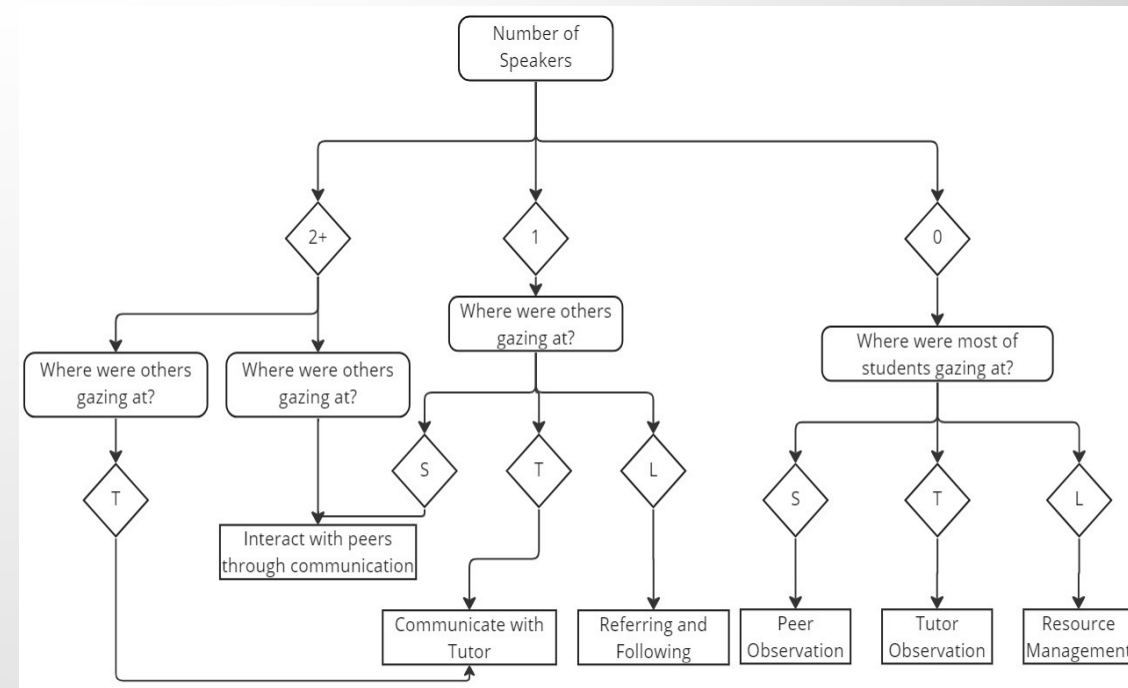
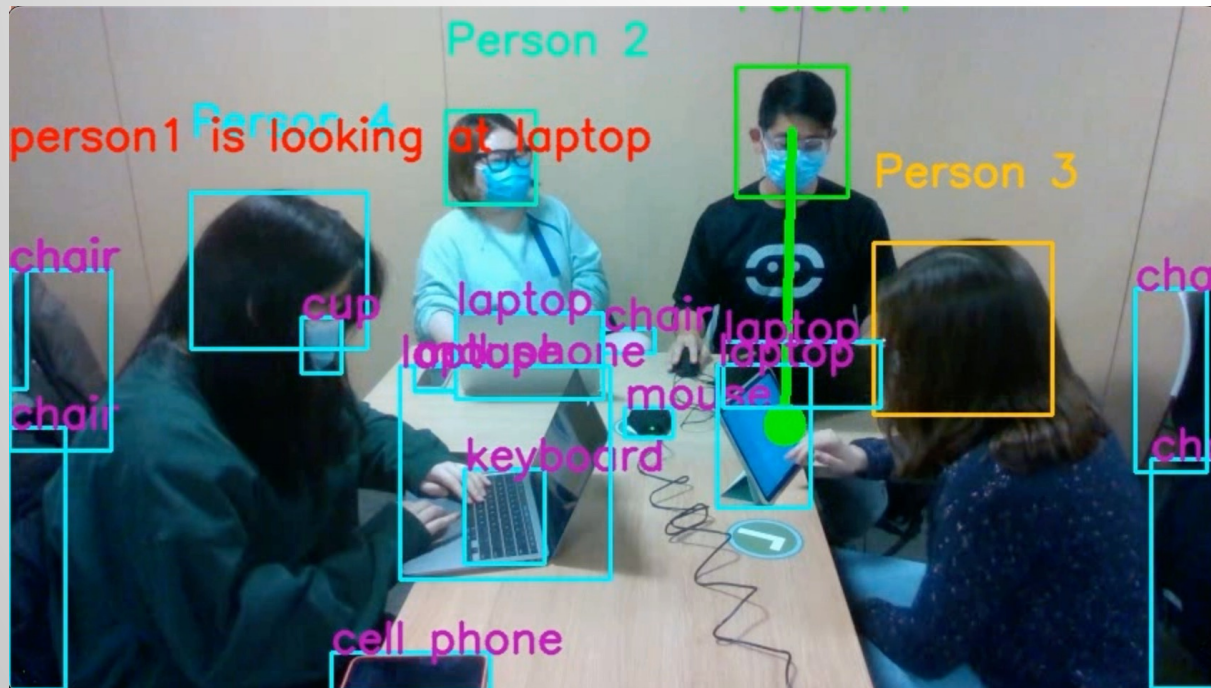


The use of AI as tools to directly intervene on the practice of teaching and learning has significant challenges.

Maybe some aspects of learning just come through the slow experience of living those learning experiences, in the sense that we can't just jump ahead to get the answer!



AI Models as Objects to Think about Learning

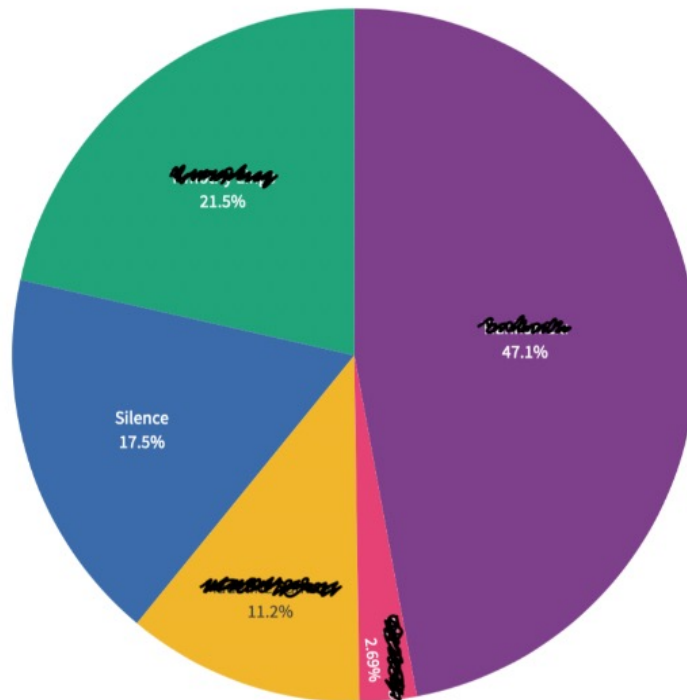


Review your session:

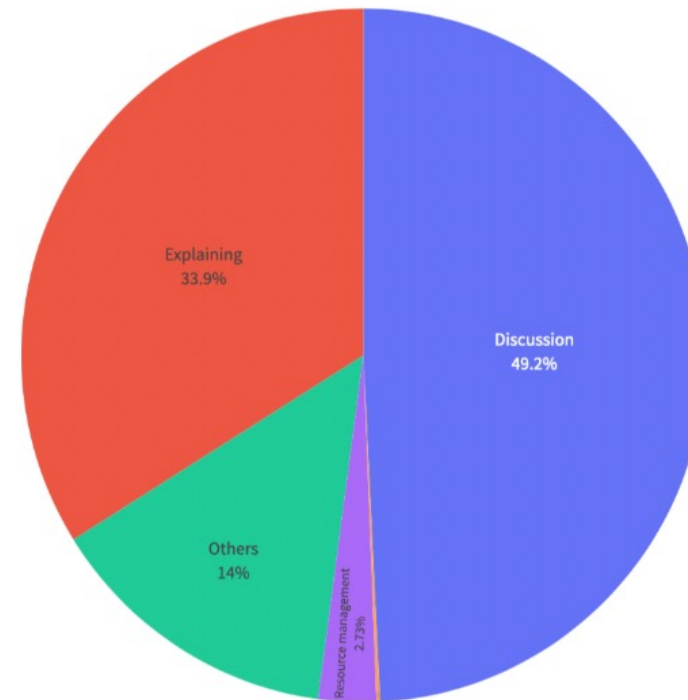
week:

2

Who spoke the most?



Which collaborative processes were your group in?

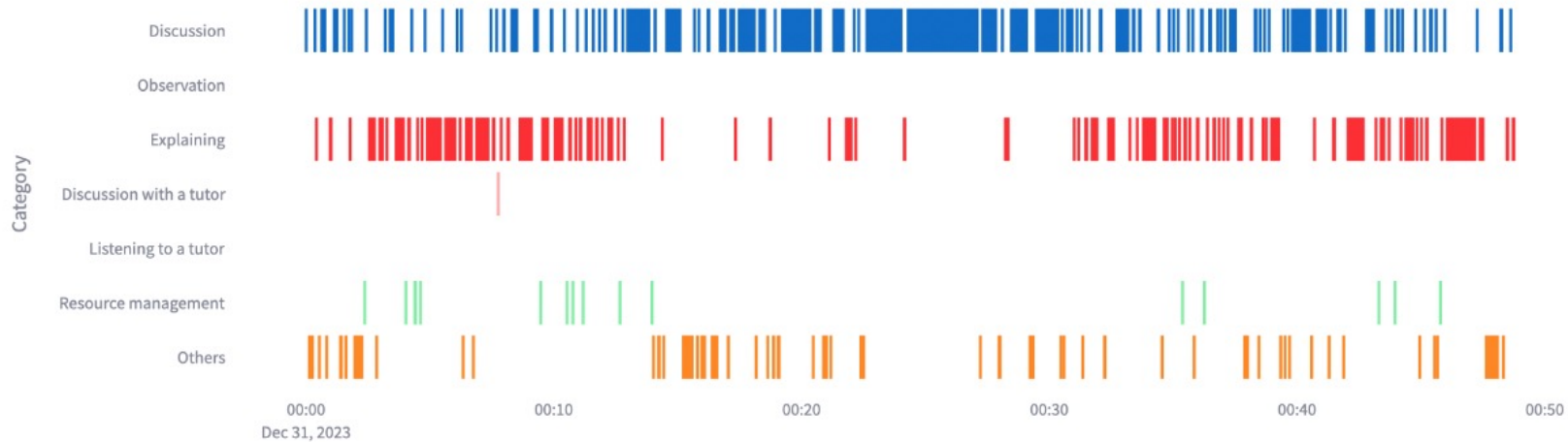


Speaking time

Group Interactions

Is this helpful?

Timeline for Group Interactions



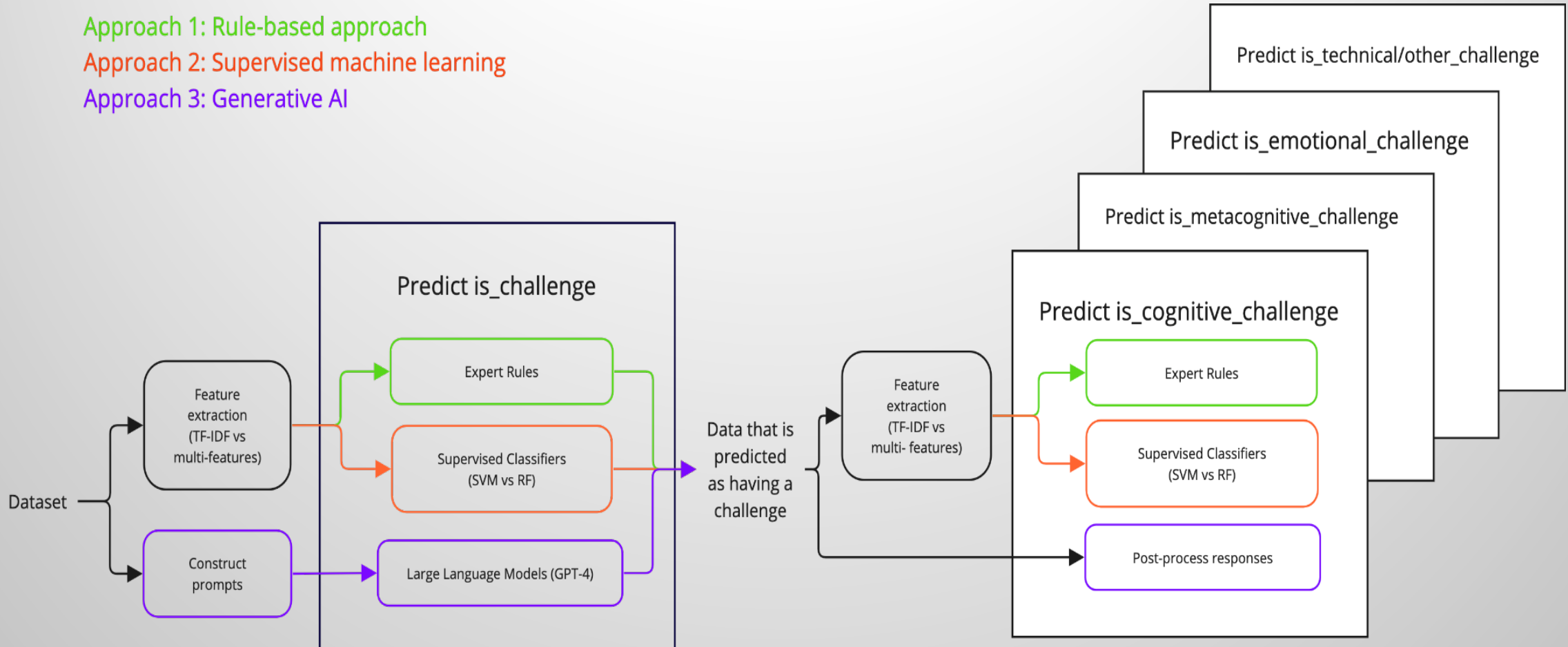
Textual feedback

This graph illustrates the collaboration process of your group in W2. Each bar represents the time spent on various types of group interactions. According to the process chart, your group invested a significant amount of time in both listening to each other explain relevant concepts based on the learning materials and engaging in discussions. It's great to observe that these discussions and explanations occurred in turns, indicating that group members were actively contributing to each other's points of view to negotiate meaning and work towards building a shared understanding. Also, it's great to note that your group engaged in communication with tutors after your internal group discussions, which is an important step for preconditioning your learning during the task activities and it can further facilitate the provision of more insightful support from tutors on your group work. Lastly, the graph indicates that your group experienced some periods with other types of interactions. This is perfectly fine and may be attributed to many reasons including separate discussions occurring during the process. According to previous literature, successful collaboration is built on the consensus of each member through inclusive discussions. It would be helpful to aim for including all members of your group in the discussion. We hope this graph can assist you in reflecting on the collaboration process. Please feel free to reach out if you have any questions about the feedback.

Approach 1: Rule-based approach

Approach 2: Supervised machine learning

Approach 3: Generative AI



<https://duteapp-feedback-test.streamlit.app/>

Review your W4 session:

Select a model

Gpt4

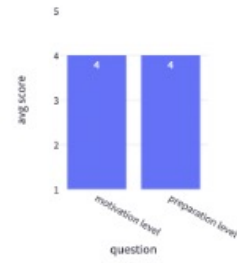
Supervised Machine Learning

pre-survey responses

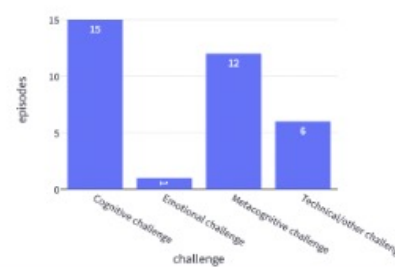
100%

This section shows a sample session of your weekly discussion. The actual transcription, the detected challenges and regulation obtained from the model are presented here. Please select a model on the right to see the predicted results.

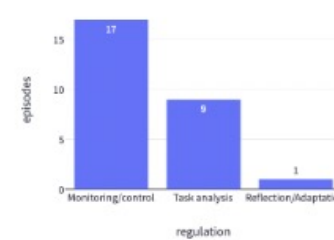
Reported motivation and preparation level



Detected Challenges



Detected Regulation

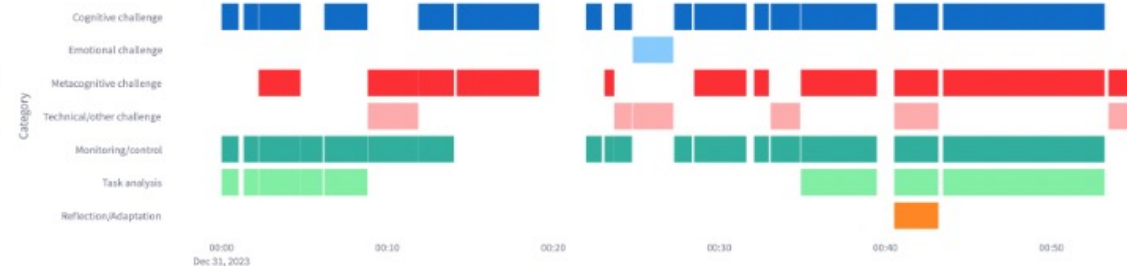


Self-report

Detected challenges/
regulation

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.

Transcription

Select a starting time to view the transcription:

00:00:00 00:57:02

00:00:00 00:57:02

Error

You can double-click on the cell to edit the transcription and press ENTER to submit your edits.

#	start	end	speaker	content
0	00:00:00	00:00:03	[REDACTED]	website and leaving that email so there's always data that's not been
1	00:00:03	00:00:12	[REDACTED]	So let's have a look at section two, I guess.
2	00:00:16	00:00:22	[REDACTED]	And Miro. Yes. No, no. It's so good.
3	00:00:34	00:00:36	[REDACTED]	perfect.
4	00:00:38	00:00:40	[REDACTED]	Yeah, no worries.
5	00:00:40	00:01:02	[REDACTED]	So, in Miro, so we will, but first, in order to see what you need to do in Deepnote, now you need to look
6	00:01:20	00:01:32	[REDACTED]	So based on should we also consider all of these? So we should also be considering for week 3 as well.
7	00:01:36	00:01:44	[REDACTED]	So in the sample in deep notes, they correlate the resource clicks with the final grades.
8	00:01:44	00:01:45	[REDACTED]	Yeah.
9	00:01:45	00:01:46	[REDACTED]	Do you want to try something else?

Episode

Inspect an episode

This section represents a fine-grained information on how the model make judgement over the discussed topics. You can browse through an episode, a single topic of discussion, to explore how the model provides judgement on detecting challenges or regulation. To change the model, use the model selection button on the top of the page.

Summary:

Total episode: 22

Select an episode: 1

Transcription of the selected episode with detected challenges and regulations highlighted in green.

start	end	speaker	content	C	M	E	T	TA	MC	RA
0	00:00:00	00:00:03	[REDACTED] website and leaving that email so there's always data that's not been	Y	N	N	N			
1	00:00:03	00:00:12	[REDACTED] So let's have a look at section two, I guess.	Y	N	N	N			
2	00:00:16	00:00:22	[REDACTED] And Miro. Yes. No, no. It's so good.	Y	N	N	N	Y	Y	N
3	00:00:34	00:00:36	[REDACTED] perfect.	Y	N	N	N	Y	Y	N
4	00:00:38	00:00:40	[REDACTED] Yeah, no worries.	Y	N	N	N	Y	Y	N
5	00:00:40	00:01:02	[REDACTED] So, in Miro, so we will, but first, in order to see what you need to do in Dee	Y	N	N	N	Y	Y	N

C = Cognitive challenges, M = Metacognitive challenges, E = Emotional challenges, T = Technical or other challenges, TA = Task analysis, MC = Monitoring/Control, RA = Reflection/Adaptation Y = Yes, N = No.

Are there any AI-detected challenges?
There is a cognitive challenge (C1) in the discussion. [REDACTED] expresses confusion about where to find the final score, indicating that she is struggling to comprehend the task or content.

Are there any AI-detected regulations?
The students are engaged in task analysis and monitoring & control. They are interpreting the task, setting personal goals, and making plans on how to approach the task (TA1 and TA2). For instance, [REDACTED] and [REDACTED] are discussing the variables and resources to consider for their task. They are also monitoring and controlling their task by engaging in a joint task and responding to each other's questions (MC1 and MC2).

Do you agree/disagree with the prediction results? and why?
Type here
Submit

Summary

Predicted challenges/regulation

Transcription

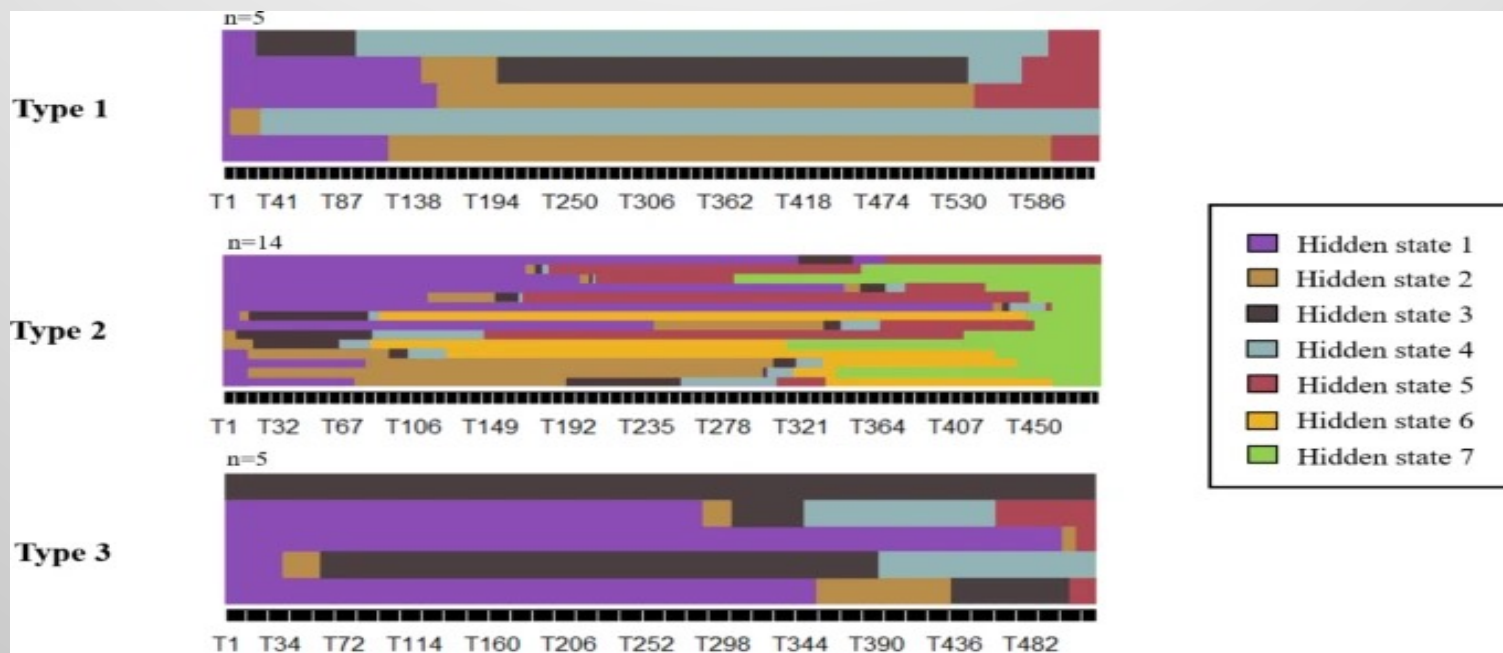
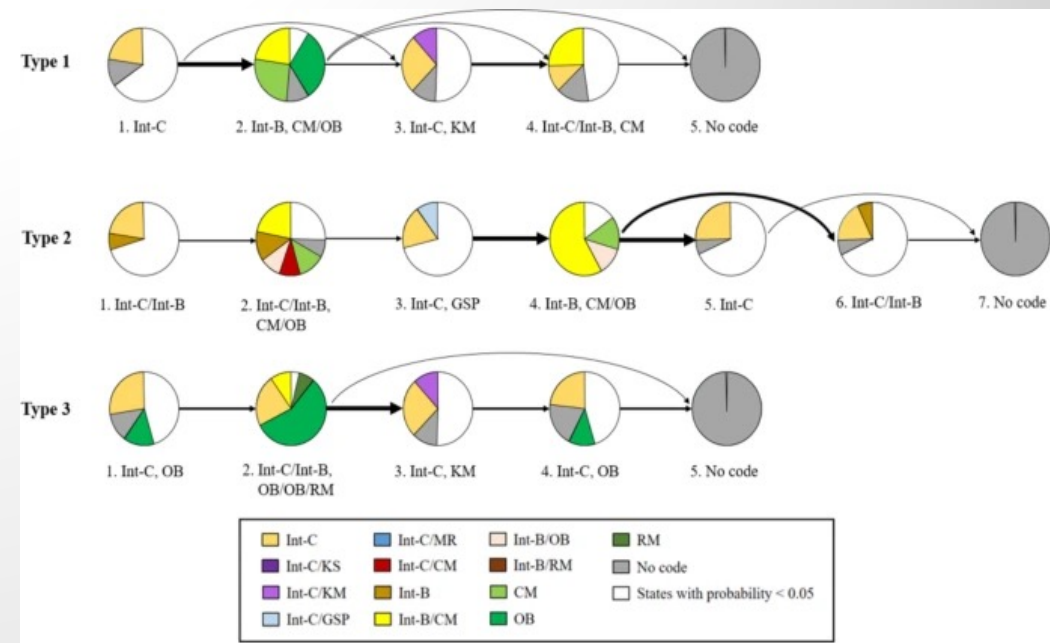
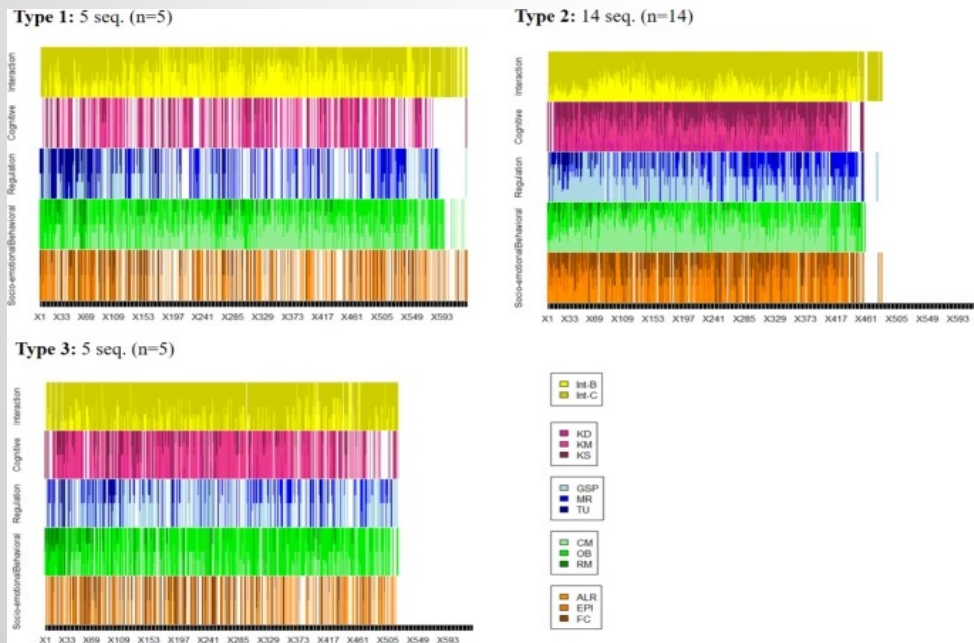
Decision explanation

Value of Making Lived Experiences Visible to End Users

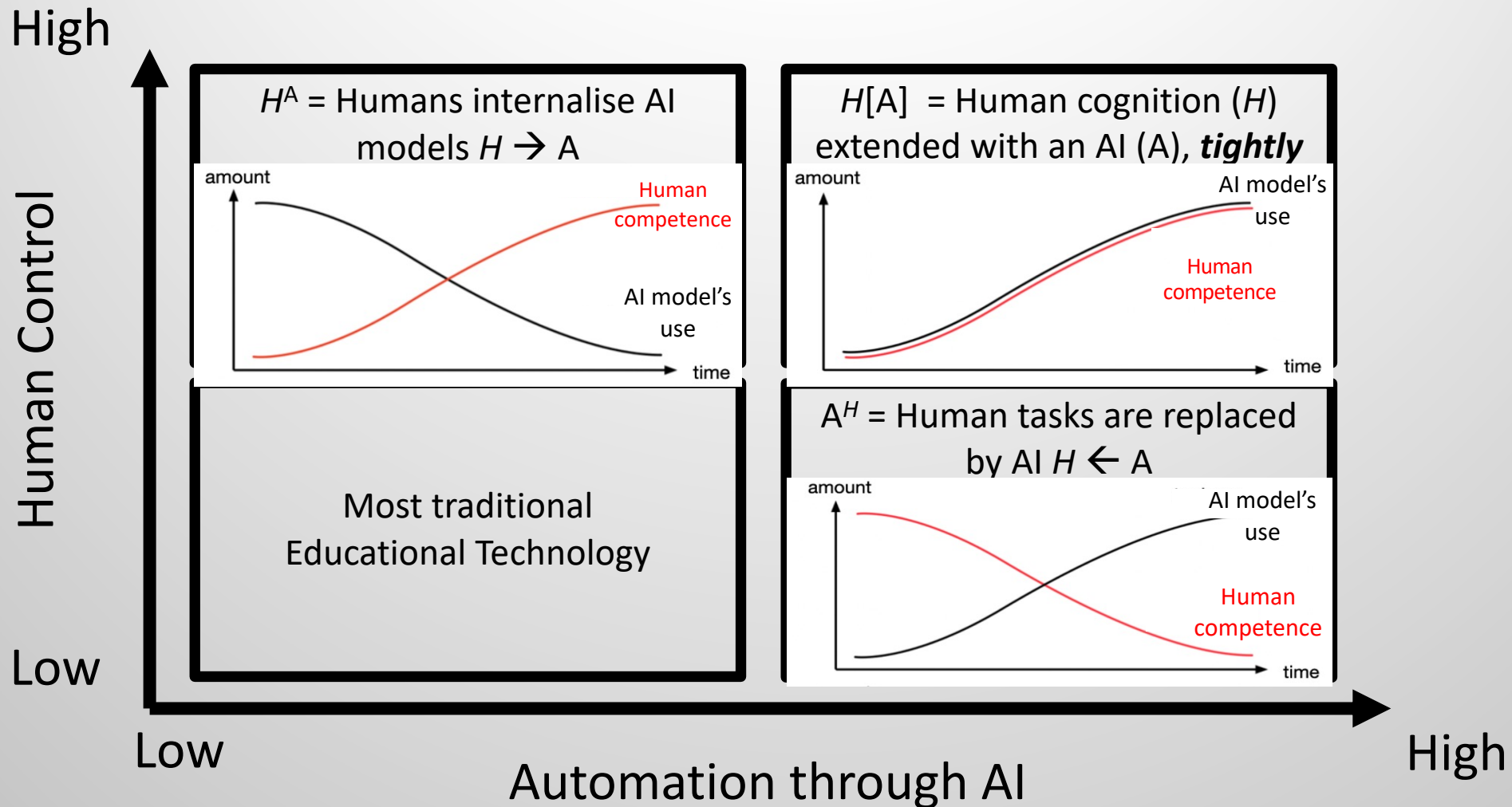
Visibility	<ul style="list-style-type: none">• Comprehensibility of the collaboration analytics (<i>easy to understand/interpret</i>)• Accuracy/Inaccuracy of the analytics information (<i>‘Similar to their findings’, different from lived experiences</i>)• Lack of qualitative feedback and partially represented contribution (<i>contribution is more than observed, speak more doesn’t mean more contribution</i>)
Awareness	<ul style="list-style-type: none">• The value of seeing one’s own performance (<i>as external reflective tool that cannot be distorted by observers/post-experienced effects</i>)• The value of seeing others’ performance (<i>determine who’s struggling</i>)
Accountability	<ul style="list-style-type: none">• Collaboration analytics to foster group discussions (<i>discuss why contribute less</i>)• Self-regulation (<i>adjust level/prepare more/seek for help</i>) and socially shared regulation of behaviours (<i>encourage the least speaker, offer helps, develop group strategies e.g. host</i>)• Gaming the system (<i>particularly for speech time data – is it bad?</i>)• Swinging back to “normal” behaviours (<i>lack of monitoring/assessment</i>)
Privacy	<ul style="list-style-type: none">• Concerns over being monitored<ul style="list-style-type: none">- <i>Were not concerned or faded due to: the module domain, invisibility, not parts of sum assessment</i>- <i>More concerning for low contributors</i>- <i>Positive motives to show for the tutors for high contributors</i>

Zhou, Q., Suraworachet, W., Pozdniakov, S., Martinez-Maldonado, R., Bartindale, T., Chen, P., Richardson, D., & Cukurova, M. (2021). Investigating Students’ Experiences with Collaboration Analytics for Remote Group Meetings. *International Conference of Artificial Intelligence in Education*, Springer, Cham.

Pozdniakov, S., Martinez-Maldonado, R., Shan-Tsai, Y., Cukurova, M., Bartindale, T., Chen, P., Harrison, M., Richardson, D., & Gasevic, D. (2022). The Question-driven Dashboard: How Can We Design Analytics Interfaces Aligned to Teachers’ Inquiry?. *Learning Analytics & Knowledge*, ACM.



AI in Education: A vision for the future



Thank you

Mutlu Cukurova
m.cukurova@ucl.ac.uk
[@mutlucukurova](https://twitter.com/mutlucukurova)

