Supporting Educators with Generative AI for Personalized Education

Meng Xia, Jan. 30, 2025

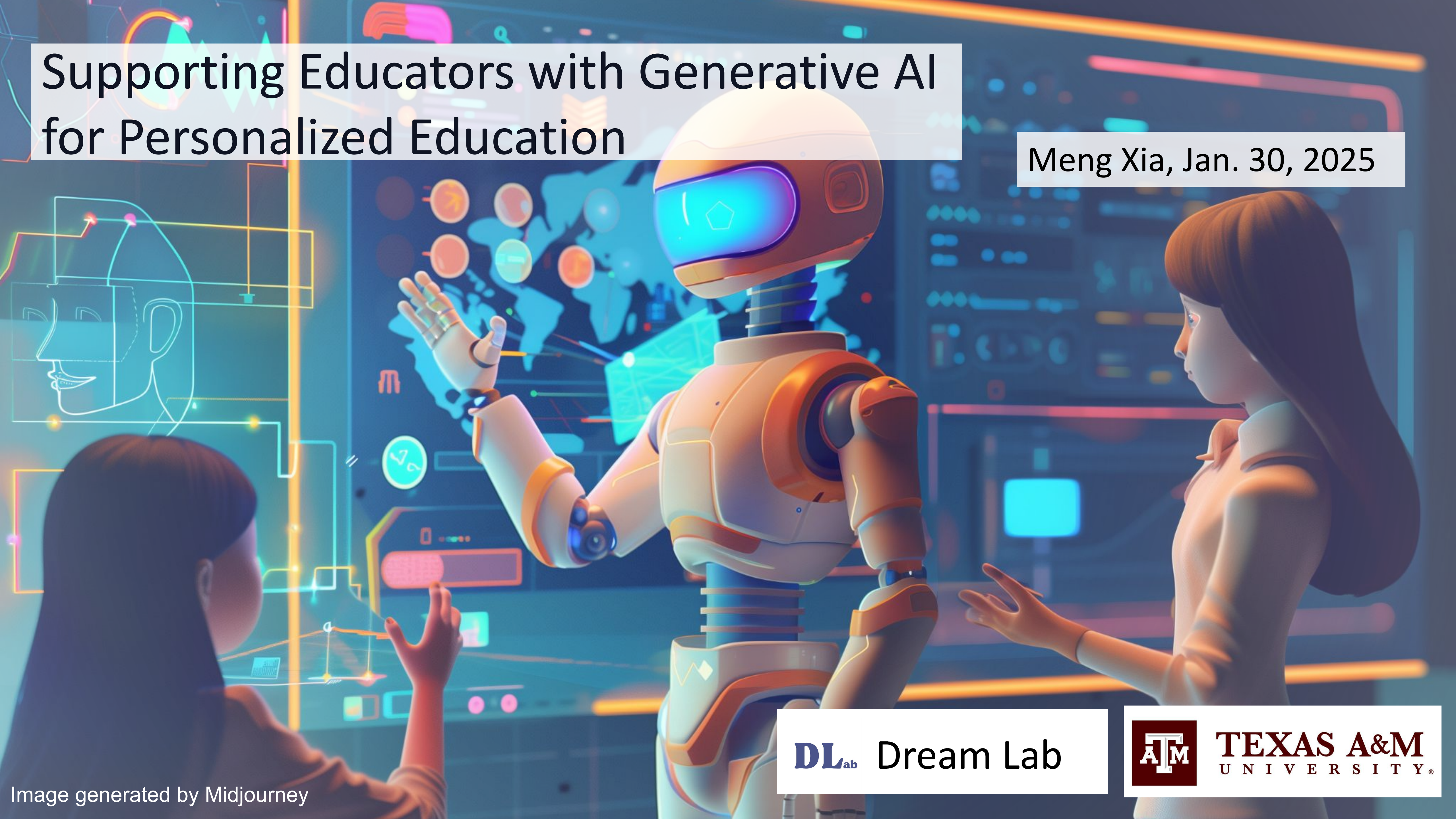


Image generated by Midjourney

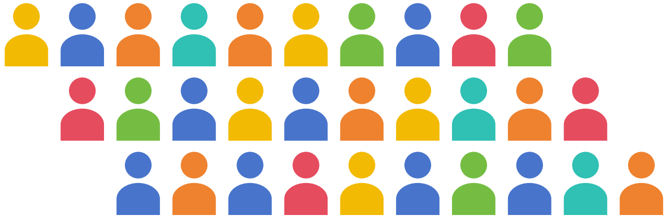
DLab Dream Lab

TEXAS A&M UNIVERSITY

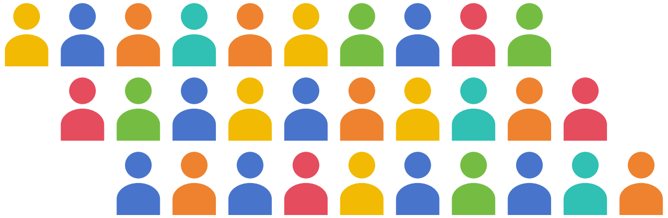
Why personalization?

Non-cognitive	Cognitive
Motivation (D'Mello, Lehman, Pekrun, & Graesser, 2014)	Knowledge (Koedinger,. Stamper, McLaughlin, & Nixon, 2013;)
Self-regulation skills (Aleven & Koedinger, 2002)	Problem-solving strategies, errors (Adams et al., 2014)

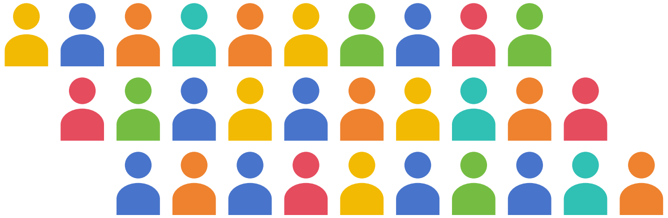
Personalization is a Foundational Education Challenge

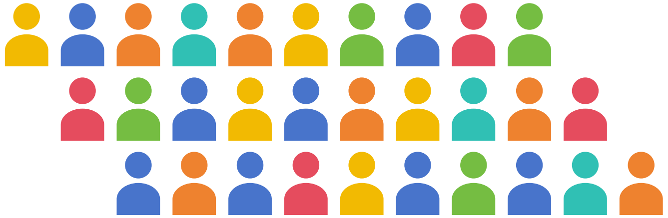


- Large amount of students
- No enough qualified teachers



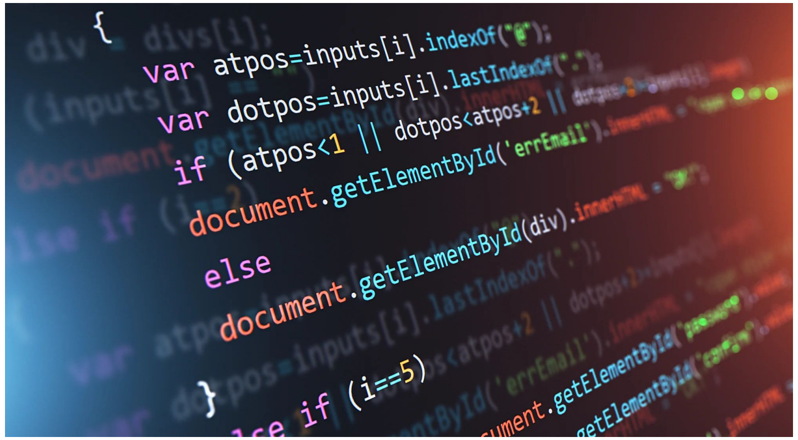
Personalization is a Foundational Education Challenge





- Large amount of students
- No enough qualified teachers
- Hard to analyze students’ multimodality unstructured data
- ……

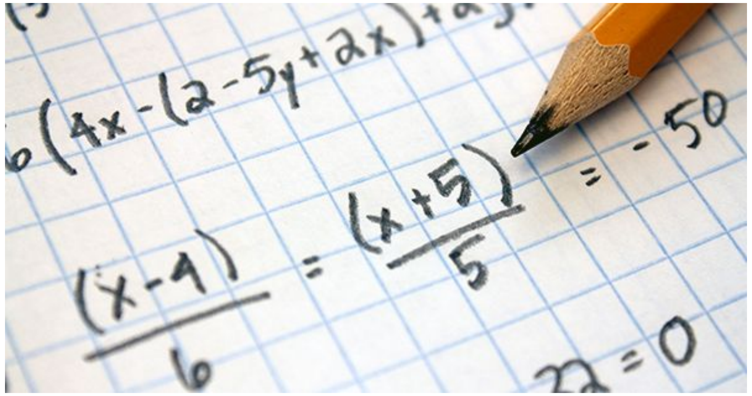
Generative AI is popular



ChatGPT

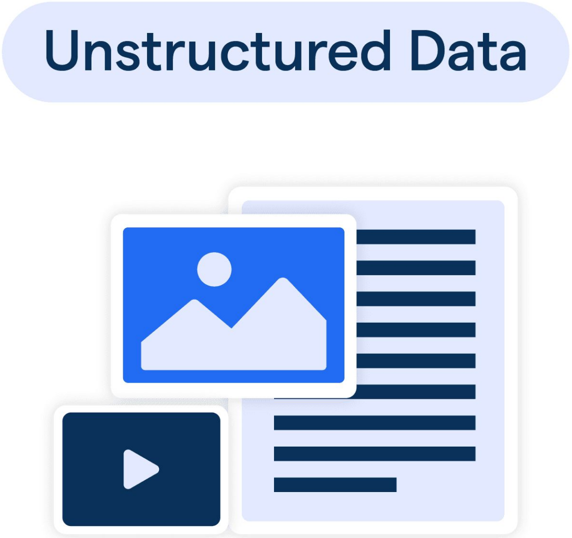
Midjourney

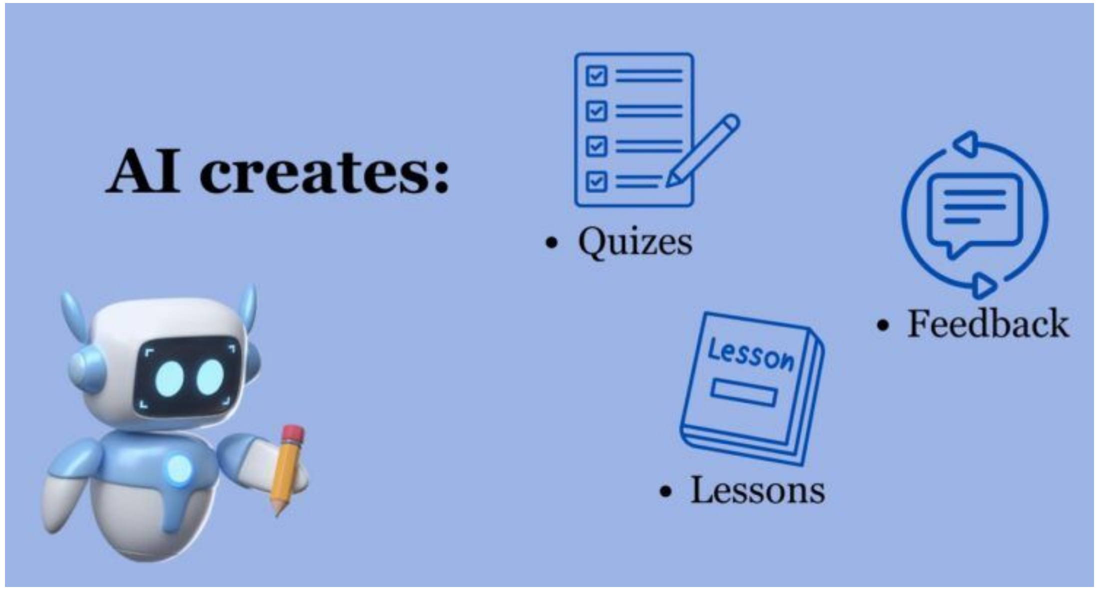
deepseek



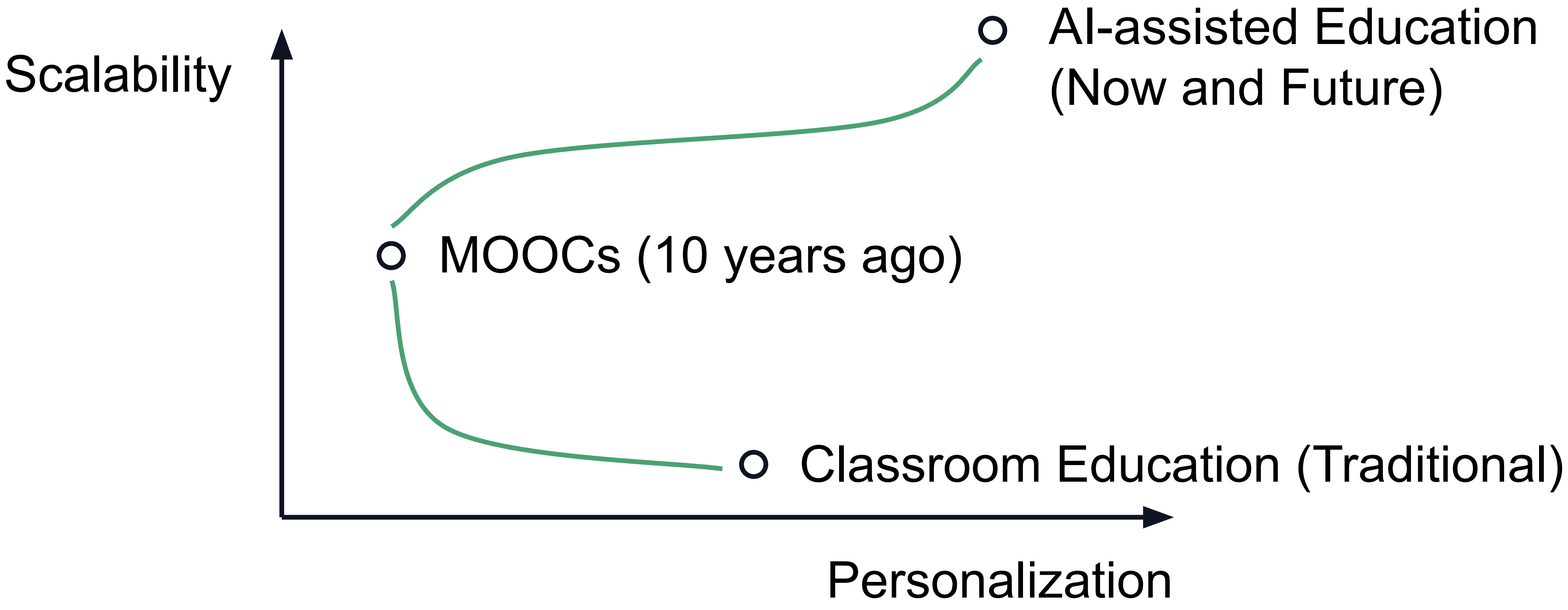
Generative AI’s Characteristics

- Understand unstructured data (e.g., text, image)
- Generate context-aware content

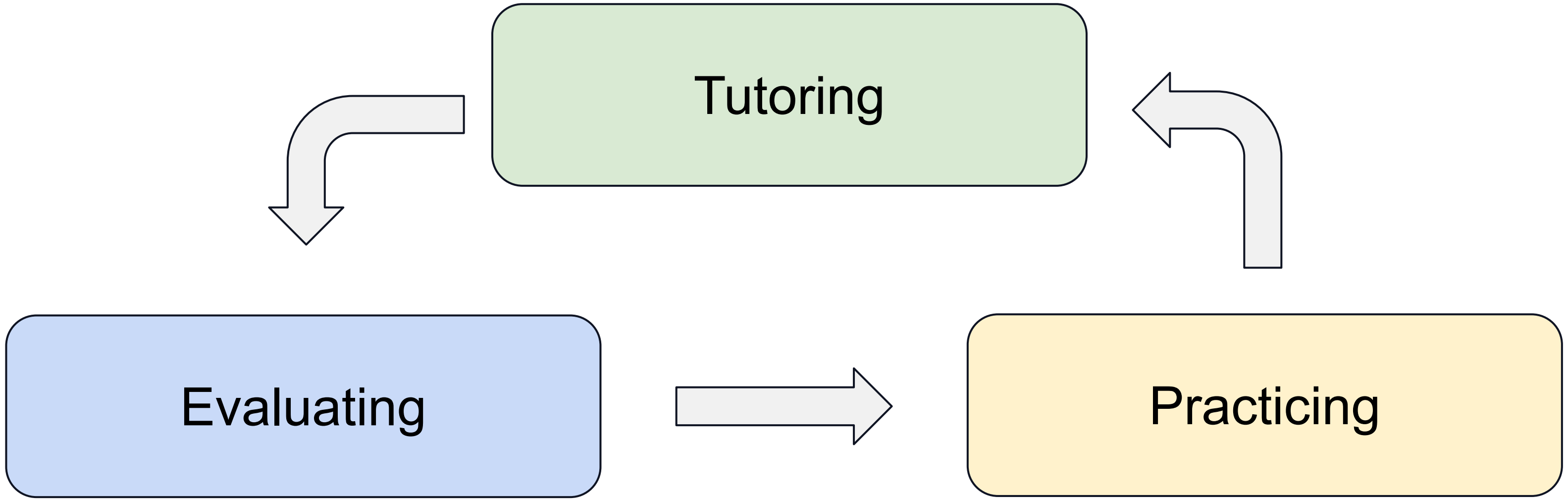




Vision for education: **Personalization @ Scale**

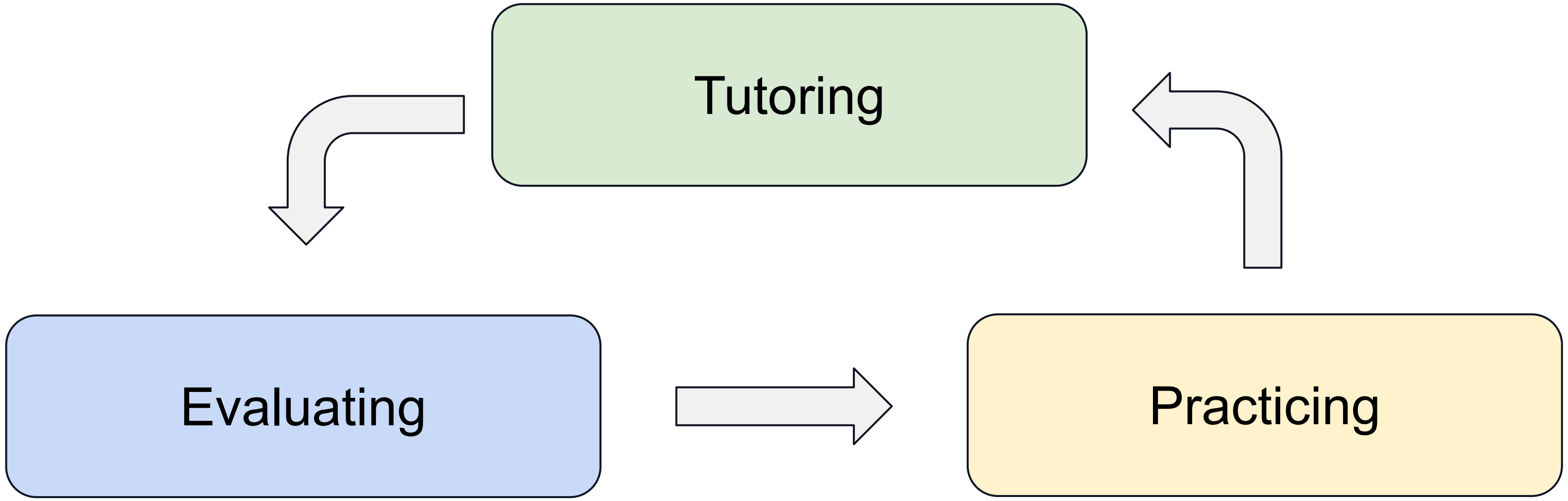


Educators’ Tasks



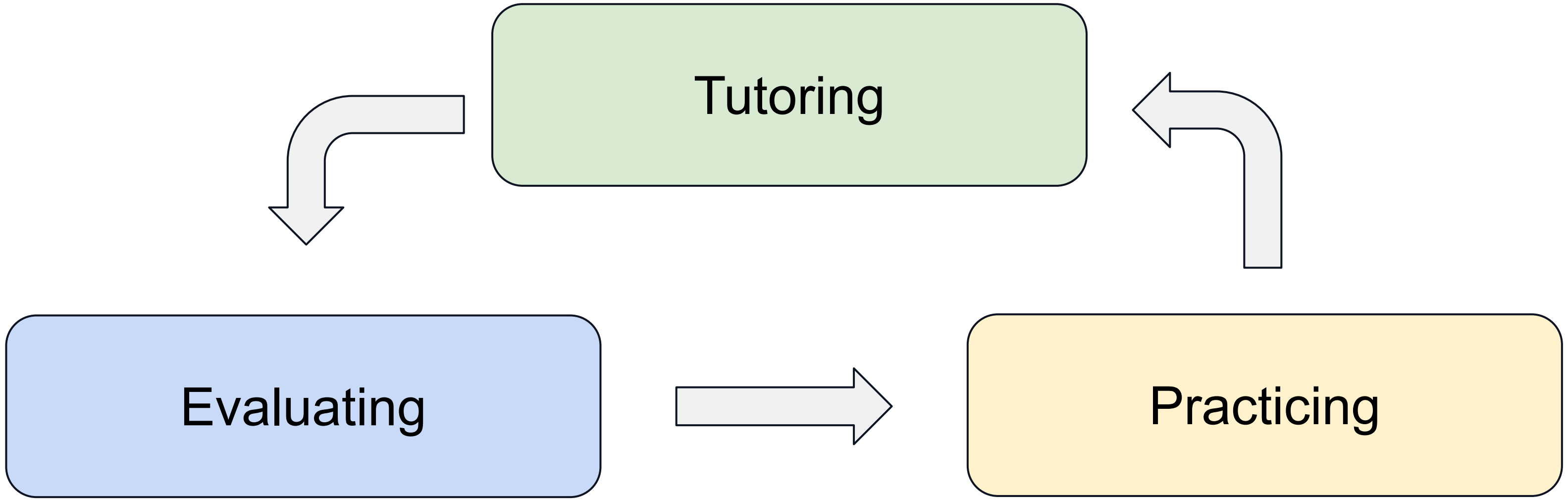
Generative AI’s opportunities for Personalization

- Authoring personalized learning materials



Generative AI’s opportunities for Personalization

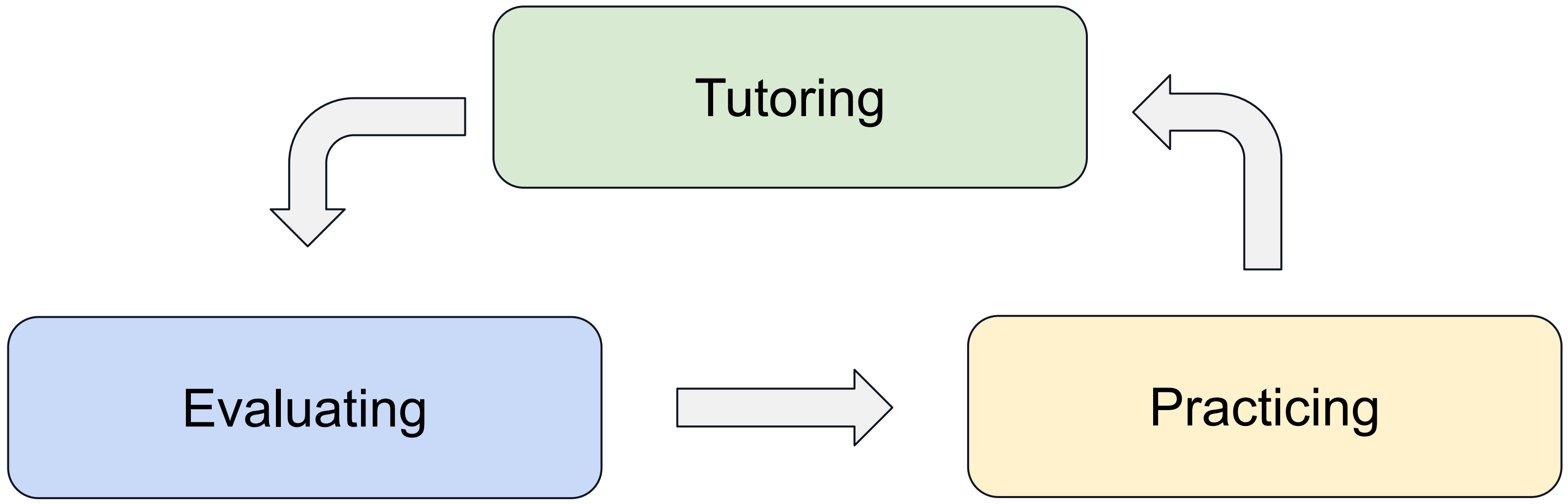
- Authoring personalized learning materials



- Evaluating student learning process

Generative AI’s opportunities for Personalization

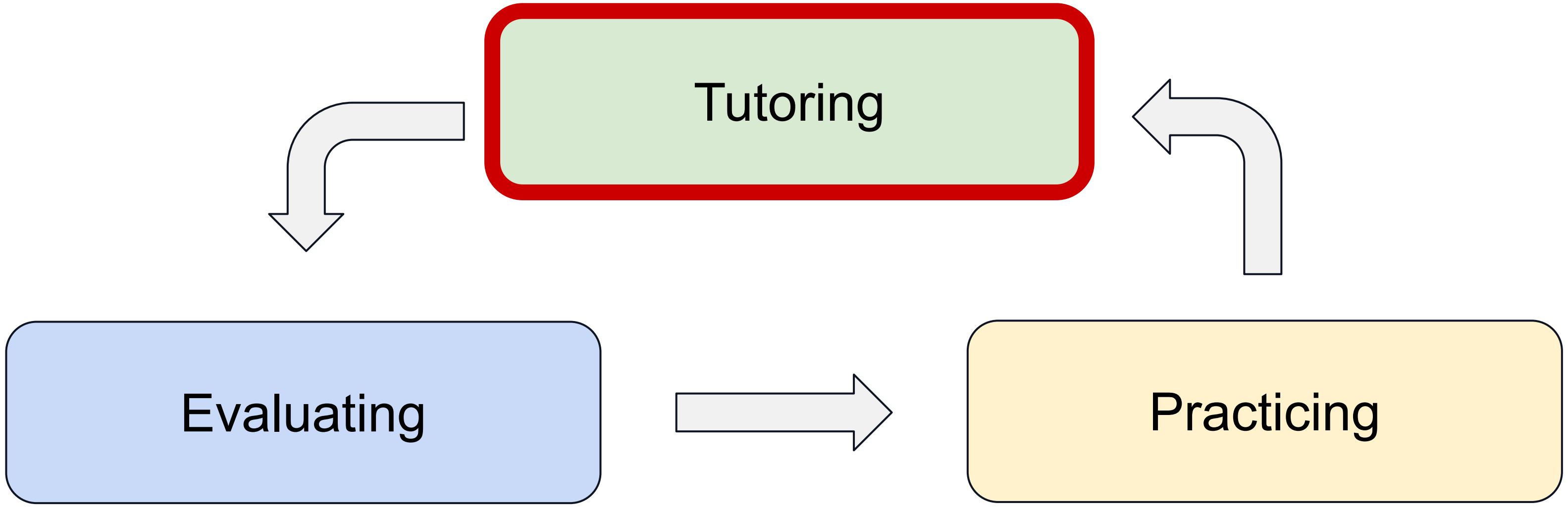
- Authoring personalized learning materials



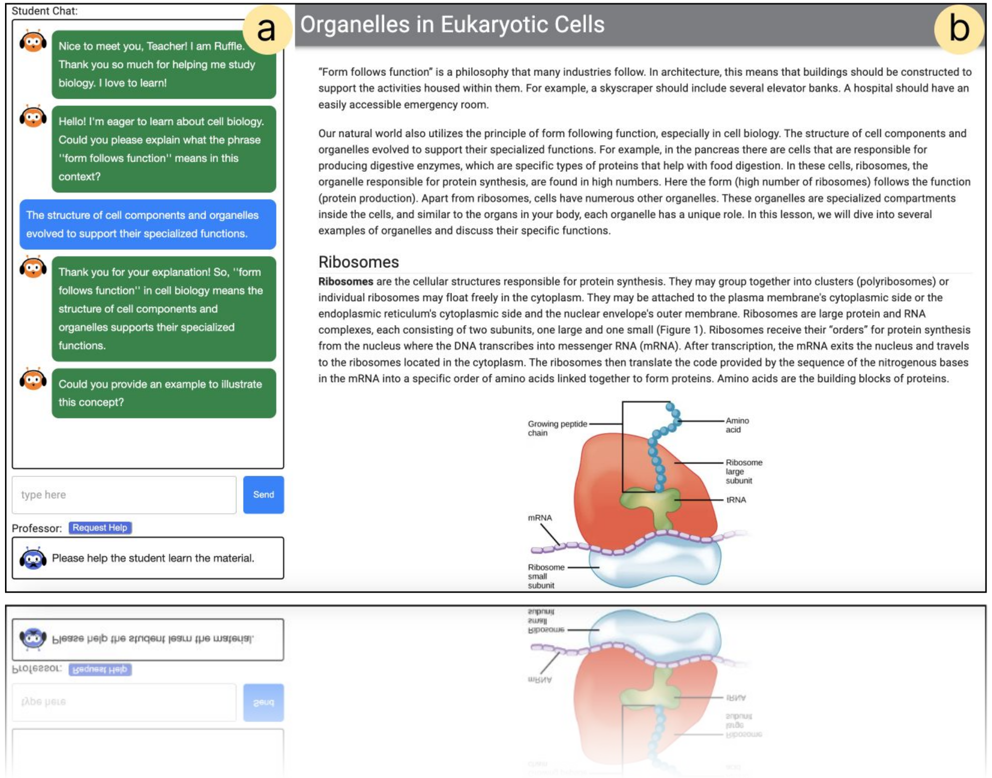
- Evaluating student learning process
- Simulating real-world environment

Generative AI’s opportunities for Personalization

- Authoring personalized learning materials



- Evaluating student learning process
- Simulating real-world environment

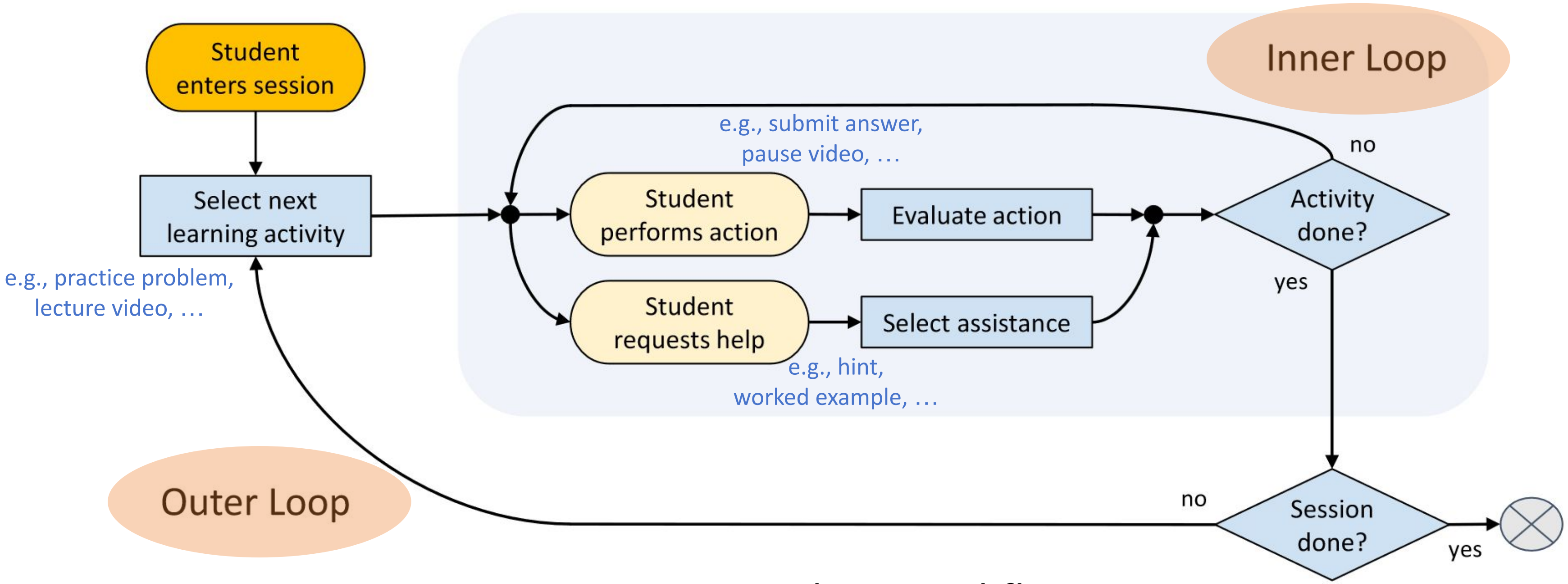


Ruffle&Riley: Insights From Designing and Evaluating a LLM-Based Conversational Tutoring System

Robin Schmucker, **Meng Xia**, Amos Azaria, Tom Mitchell

AIED 2024

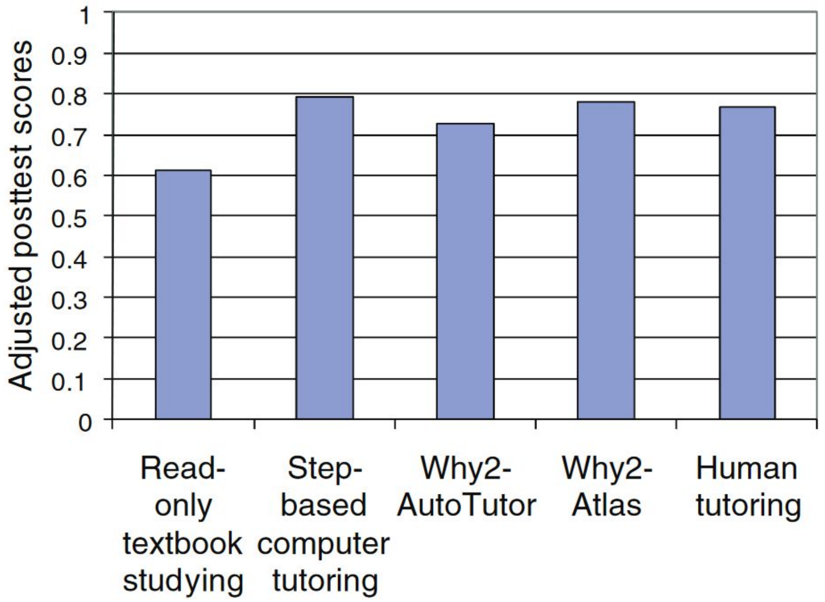
Provide **personalized** and **adaptive** instruction



Prototypical ITS workflow

Benefits of ITSs:

- Can be as effective as human tutoring
- Can be scaled to millions of learners
- Provide a formal framework for thinking about tutoring processes

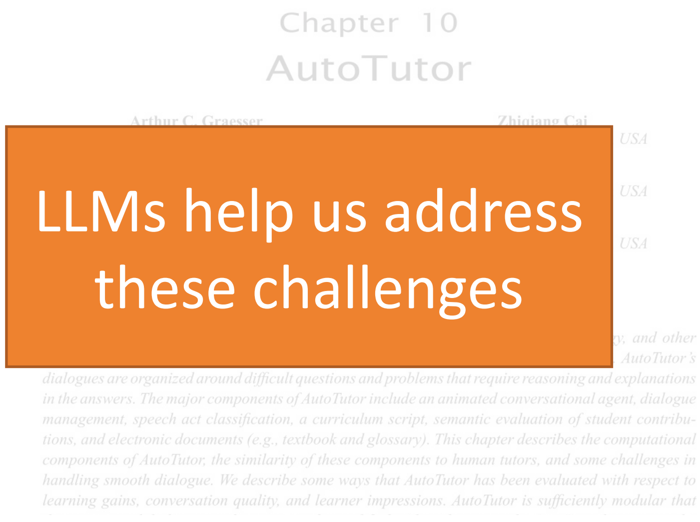


Learning outcomes of different activities [1]

[1] VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational psychologist*, *46*(4), 197-221.

Limitations of ITSs:

- High cost of content authoring
- Limited language understanding
- Limited conversational facilities
- Limited question answering facilities



AutoTutor Recap [1]

[1] Graesser et al. (2012). AutoTutor. *Applied Natural Language Processing: Identification, Investigation and Resolution*, 169-187.

LLMs can provide components to ITS designers:

- By generating lesson texts
- By generating question and hints
- By adding automatic grading capabilities
- By adding question answering capabilities
- …

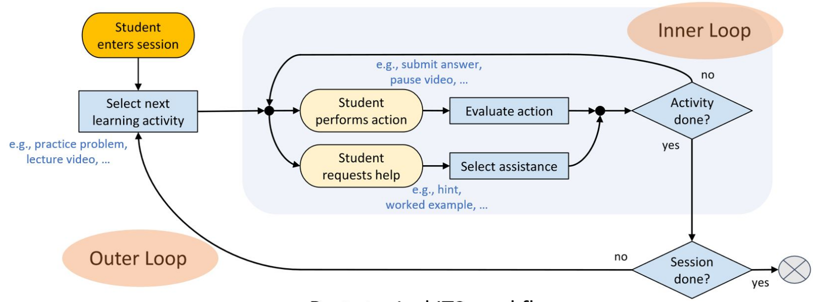
What would it take to generate an **entire** ITS?

The ITS workflow has two components:

- *Outer loop*: Learning activity sequencing
- *Inner loop:* Guidance during learning activity

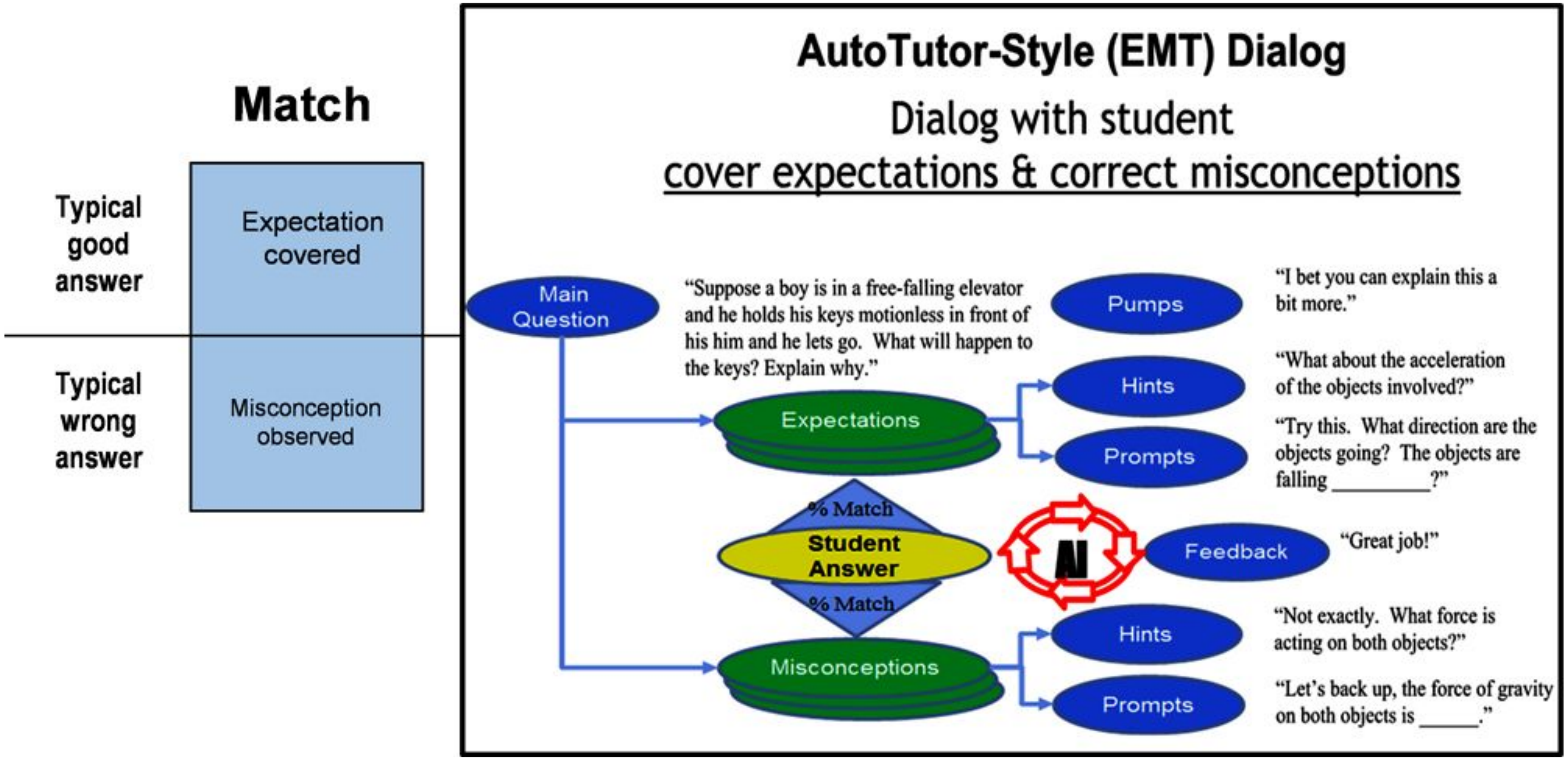
<u>Playing to our strengths</u>:

- LLMs excel at **conversational** activities



Prototypical ITS workflow

How to structure Conversational Tutoring?



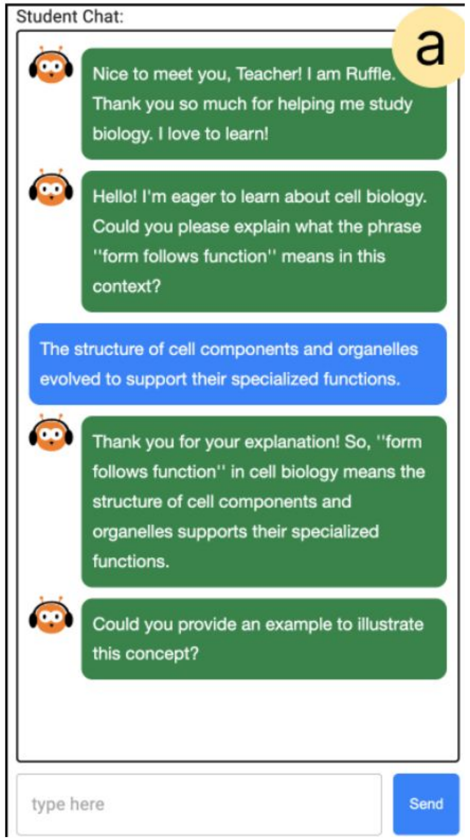
Expectation Misconception Tailoring (EMT) [1]

[1] Xiangen Hu (2017). Workshop on Educational and Social Science Technologies (WESST).

Ruffle&Riley: User Interface

Pre-existing textbook content

Free-form dialog



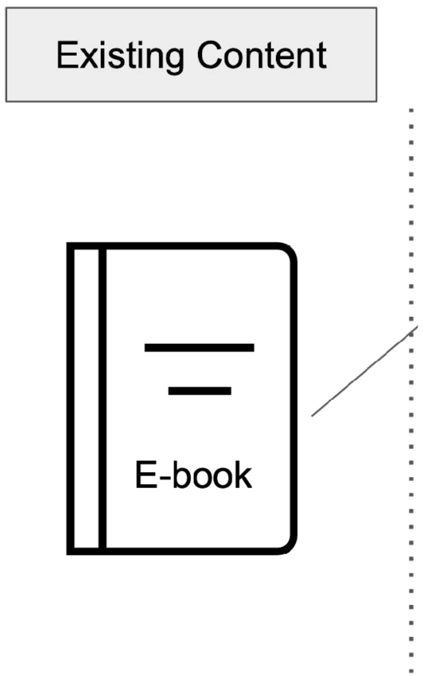
Intelligent Feedback

On demand assistance

Misconcept. Correction

User Interface

*Facilitate tutoring **script generation** and **orchestration***



System Architecture

Tutoring Script

Specifies **what** we want to teach

Enables our LLM-based agents to conduct structured workflow

Topic 1: What does the phrase "form follows function" mean in the context of cell biology? Provide an example to illustrate your answer.

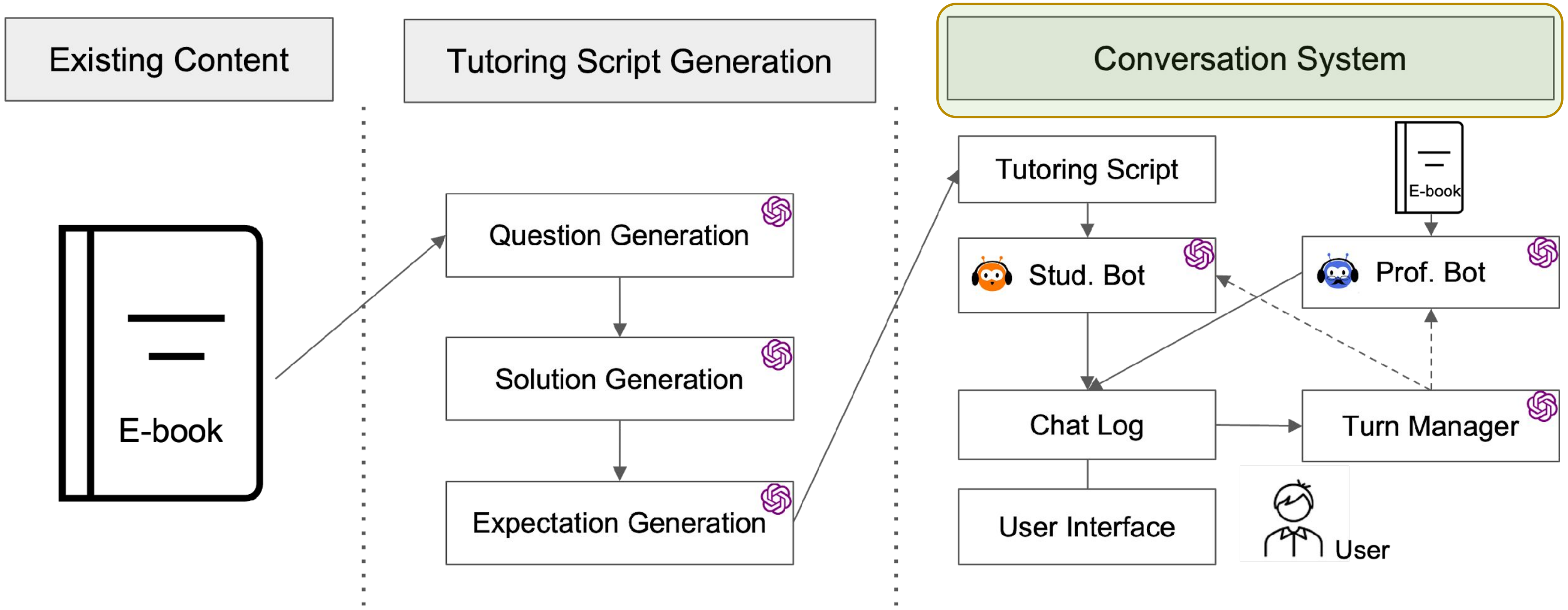
Fact 1.1: "Form follows function" in cell biology means the structure of cell components and organelles supports their specialized functions.

Fact 1.2: An example is the high number of ribosomes in pancreas cells that produce digestive enzymes, supporting their function of protein production.

Topic 2: Describe the structure and function of ribosomes. How do they contribute to protein synthesis?

...

*Facilitate tutoring **script generation** and **orchestration***



System Architecture

Conversation System

Specifies **how** we want to teach

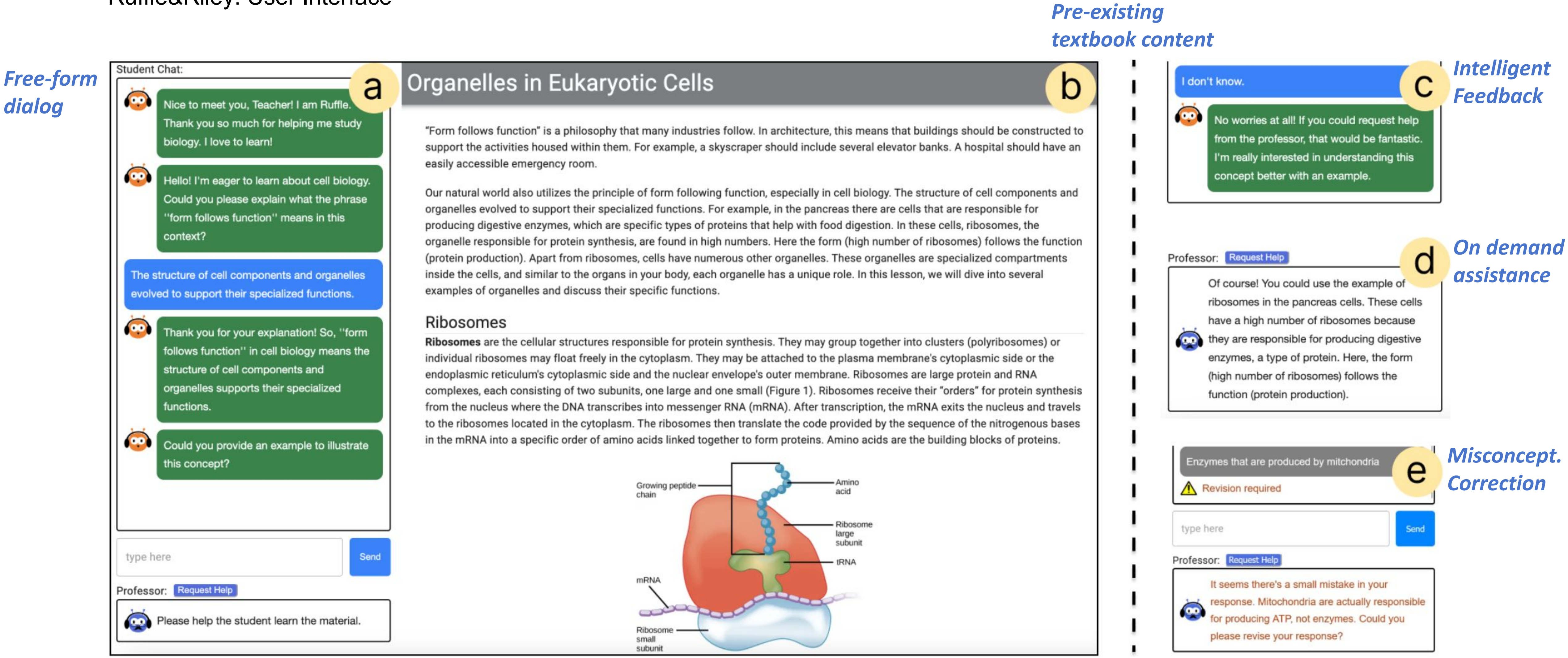
Executes tutoring script as free-form conversation

Student Agent (Ruffle) :

You are an enthusiastic 18-year-old student who is trying to learn. You need the user (who is a teacher) to teach you all topics in the material. You have access to a list of topics and facts that the teacher needs to convey to you. You must learn one topic at a time. This is the list of topics you found on the internet that you need the user to slowly teach you (in order): + {TUTORING-SCRIPT}

Ask the user (who is the teacher) to teach you the material, little by little. If the teacher gives an answer, you must (a) show appreciation and summarize the answer; (b) insert [SMILE]; and then (c) ask a follow-up question that does not give the solution away if the teacher has not touched all facts about the current topic OR ask a question about the next topic. Do not move on to the next question before getting an answer for your current question. If the teacher doesn't know something, tell the teacher you will be thrilled if the teacher can check it and get back to you. When all the topics are covered, thank the teacher, say I've asked all the questions.

Ruffle&Riley: User Interface



User Interface

https://github.com/rschmucker/ruffle-and-riley

Evaluate learning experience/test performance of 4 conditions:

1. Reading only
2. Q/A Chatbot with Human-generated questions/answers
3. Q/A Chatbot with LLM-generated questions/answers
4. Ruffle&Riley (EMT + Learning-By-Teaching)

Participants (N = 200) are free to decide how to approach learning activity

Prolific

Findings of user study

Learning Experience Survey

Conditions	Learning Experience (1-strongly disagree, 7-strongly agree)						
	Engagement	Understanding	Remembering	Interruption	Coherence	Support	Enjoyment
Reading	4.33 ± 0.52	-	-	-	-	-	-
Teacher Q/A	5.0 ± 0.53	4.43 ± 0.65 *	4.43 ± 0.65 *	2.71 ± 0.64	5.43 ± 0.53	4.57 ± 0.57 *	3.71 ± 0.52 *
LLM Q/A	4.8 ± 0.47	4.4 ± 0.4 *	4.33 ± 0.42*	2.67 ± 0.45	4.8 ± 0.43*	4.0 ± 0.44*	4.0 ± 0.44 *
Ruffle & Riley	**5.81 ± 0.3**	**5.81 ± 0.24**	**5.76 ± 0.22**	**2.19 ± 0.34**	**6.1 ± 0.21**	**5.9 ± 0.26**	**5.62 ± 0.31**

Symbol "" marks $p < 0.05$*

Findings of user study

Post-Test Performance

Conditions	Previous Knowledge	Learning Performance
		Post-test Scores (i.e.,Multiple-Choice Questions)
Reading	2.53 ± 0.41	5.07 ± 0.33
Teacher Q/A	**3.0 ± 0.58**	4.14 ± 0.83
LLM Q/A	2.2 ± 0.3	4.67 ± 0.35
Ruffle & Riley	2.67 ± 0.43	**5.19 ± 0.25**

Symbol "" marks $p < 0.05$*

Findings of user study

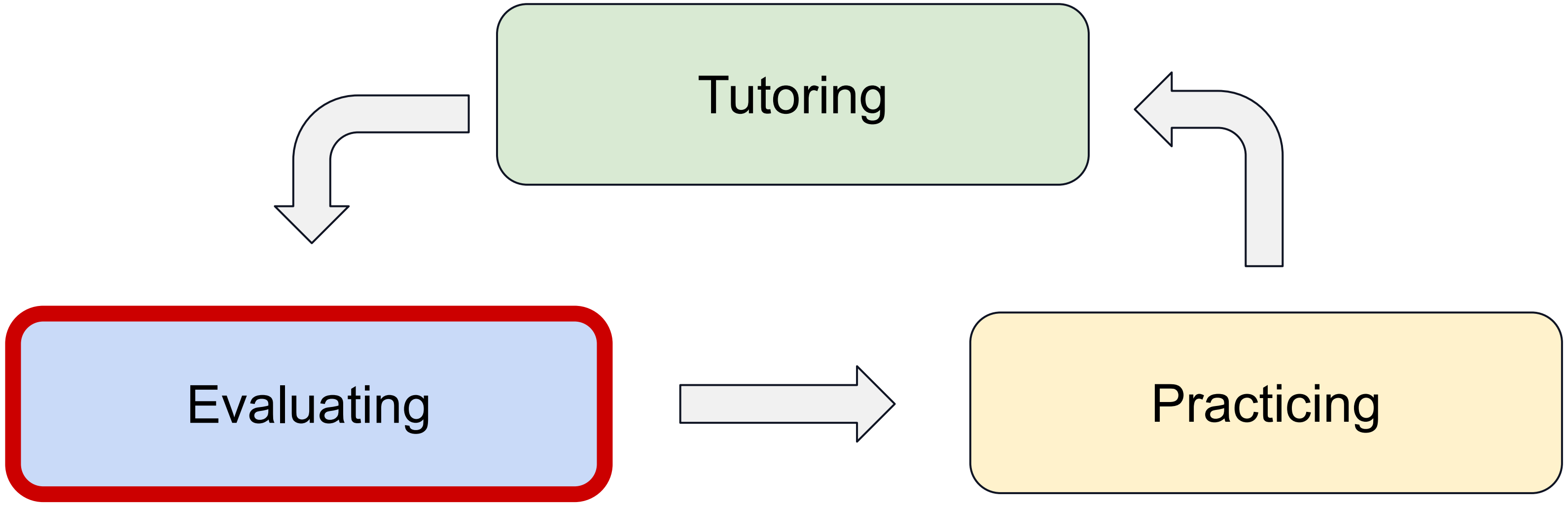
Learning Gains

Conditions	Pre-test Score	Post-test Score	Learning Gain	
			Absolute	Normalized
Reading	1.37 ± 0.17	**3.53 ± 0.25**	**2.16 ± 0.25**	0.44 ± 0.07
Ruffle&Riley	**1.54 ± 0.23**	3.49 ± 0.28	1.94 ± 0.26	**0.47 ± 0.05**

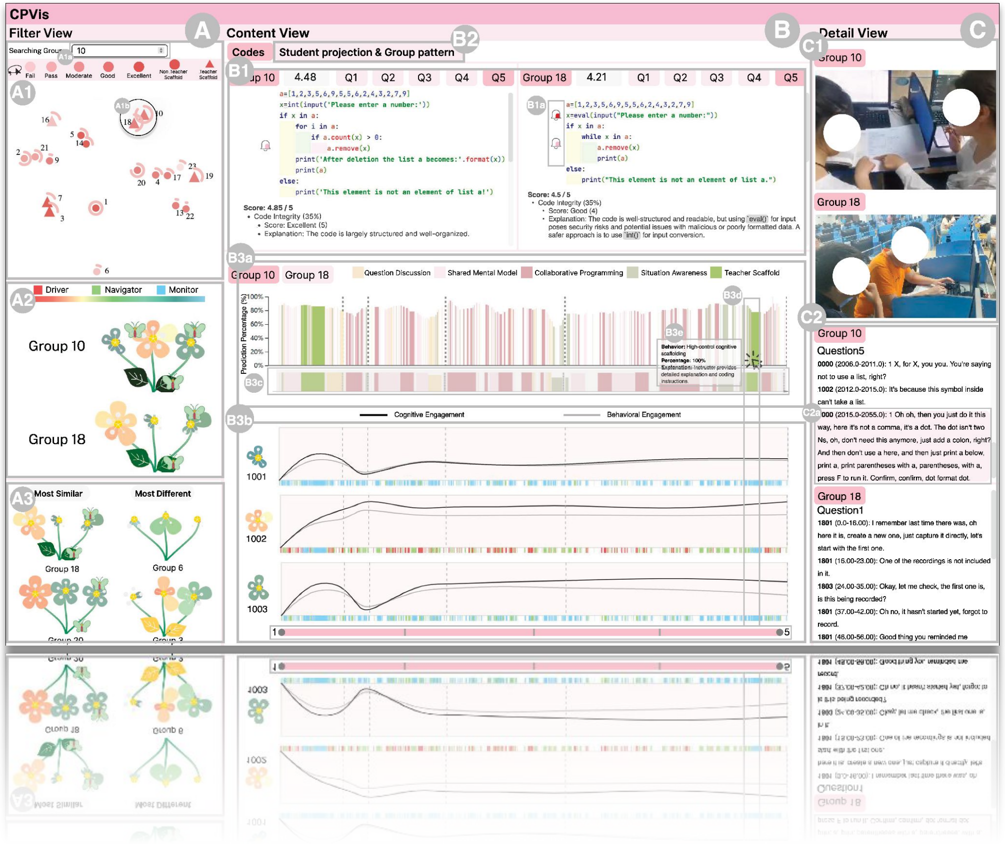
Symbol "" marks $p < 0.05$*

Generative AI’s opportunities for Personalization

- Authoring personalized learning materials



- Evaluating student learning process
- Simulating real-world environment



CPVis: Evidence-based Multimodal Learning Analytics for Evaluation in Collaborative Programming

Gefei Zhang, Shenming Ji, Yicao Li, Jingwei Tang, Jihong Ding, **Meng Xia***, Guodao Sun, Ronghua Liang

CHI 2025 (Conditionally Accepted)

Background

Challenges

- Viewing Students' Code is A Pain
- Student work is often assessed only by the final solution
- Difficulty in understanding students' engagement in problem-solving

```
a = [1, 5, 9, 10, 13]
b = [4, 6, 8, 11, 14, 15]
```

```
a = [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
e = [1, 2, 3, 4, 5, 6, 7, 8, 9]
f = [3, 4, 5, 6, 7, 8, 9, 10]
for n in f:
    for b in e:
        for c in a:
            for d in a:
                if b ** n + c ** n + d ** n == 100 * b + 10 * c + d:
```

```
for i in range(100, 1000):
    j = i // 100
    k = i // 10 % 10
    l = i % 10
    if i == j ** 3 + k ** 3 + l ** 3:
        print(i)
```

```
):
 != k and
100 + j * 10
```

```
a = [49, 38, 65, 97, 76, 13, 27, 55, 4]
a.sort()
print(a)
```

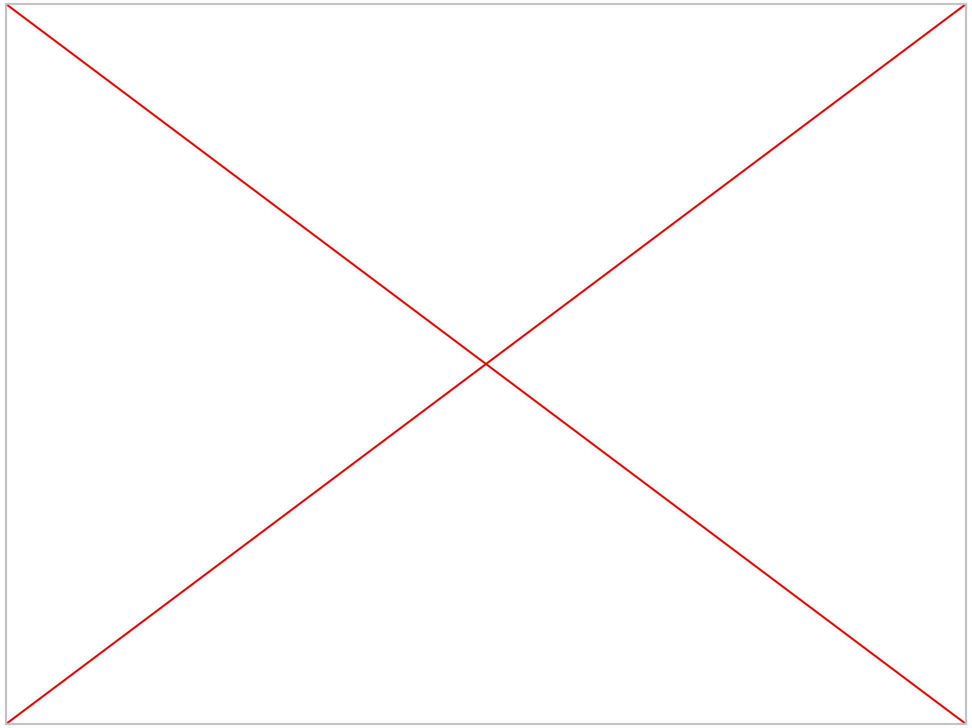
```
a = [1, 5, 9, 10, 13]
b = [4, 6, 8, 11, 14, 15]
c = a + b
c.sort()
```

Dataset

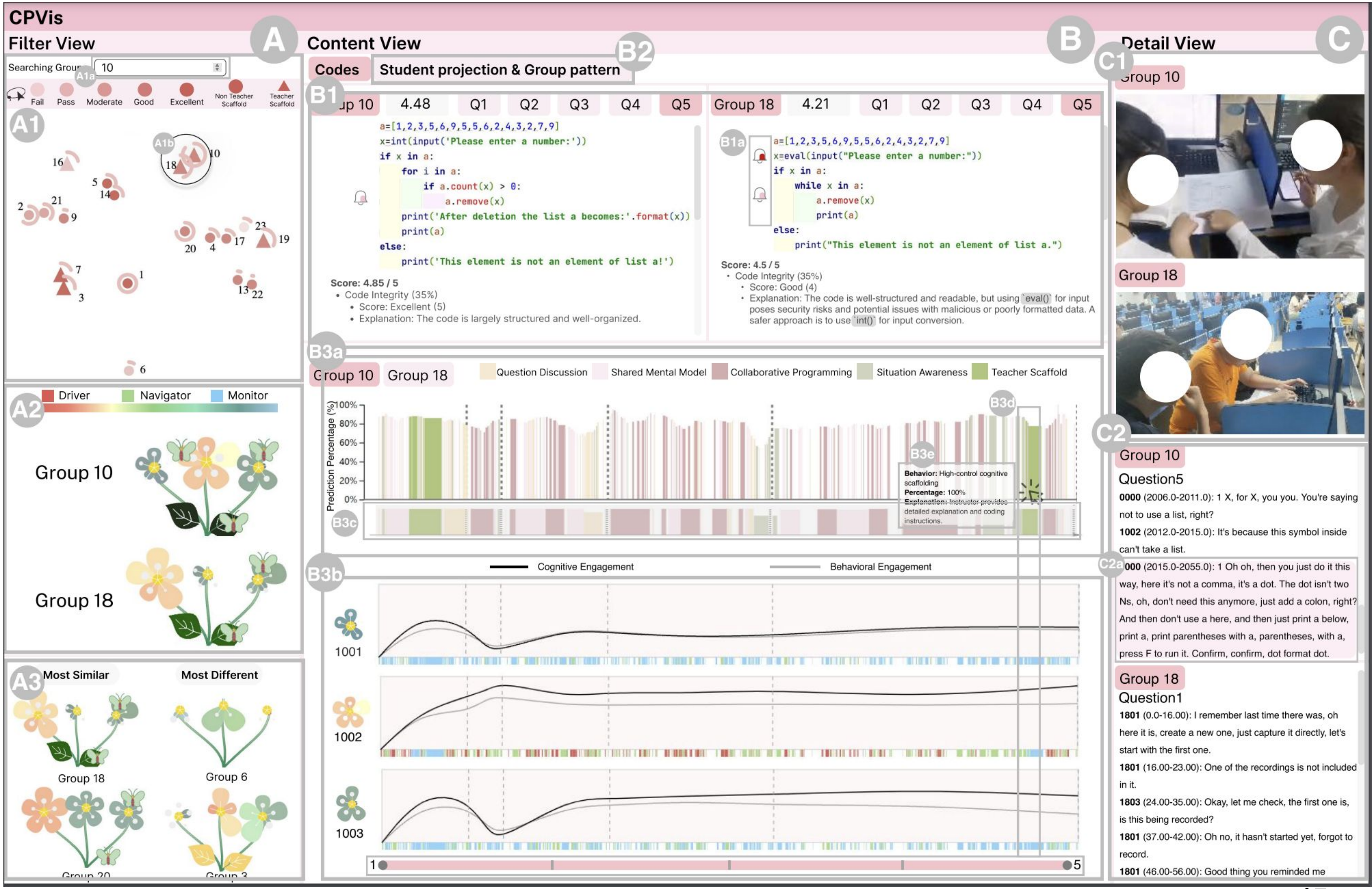
21 groups, 3 students per group in one class session (five coding problems)

- audio
- video
- screen sharing
- codes

System demo



Generative AI's Roles



Categories	Communication behaviors	Definitions	Examples
Question Discussion	Material reading [68]	Students read the distributed material together.	"Let's go over the handout the teacher gave us."
	Question allocation [78]	Students explicitly assign a question to others or proactively self-allocates a task.	"You debug the code, I'll write the test cases."
	Question planning [78]	Students list several questions remaining to be done to provoke subsequent question allocation.	"We still need to write the test cases debug the code."
	Question understanding [58]	Students explore programming with peers without providing detailed descriptions of Python coding.	"There's a problem. This one hasn't been modified."
Shared Mental Model	Information sharing [78]	Students proactively share information that no one asked.	"I found a better algorithm that improves efficiency."
	Information request [78]	Students ask someone else a question to obtain information.	"How should this function work?"
	Responding to request [78]	Students provide information in response to a asked question.	"This function takes two arguments……"
	Acknowledgement [78]	Students acknowledge receipt of information from others.	"Okay", "I agree", "Got it"
Collaborative Programming	Debugging [68]	Students are debugging the final code.	"There's a bug here, I need to double-check the values"
	Python coding [58]	Students provide detailed explanations of programming.	"You switch to the function remove"
	Print and evaluate code[58]	Students write and test code in a cyclical process, continuously writing and testing.	"Let me run the code to see the results and then tweak it."
Situation Awareness	Escalation [78]	Students ask for assistance from the instructor either verbally.	"I think we need to ask the teacher about this."
	Unrelated chat among students [58]	Students engage in unrelated conversations with peers.	"What are the other groups doing?"
	Difficult-to-reconcile conflicts [68]	Students encounter conflicts that are challenging to resolve.	"We've been debating which way to implement this"

Figure 1: Collaborative programming coding schemes, along with their definitions and examples.

Evaluation

We evaluated LLMs’ performance in code quality by comparing it to human-labeled (two experienced educators, I1, and I2) results.

The results showed that I1 and I2 reached **93.43% agreement**, while ChatGPT-4o’s annotations matched I1 and I2’s annotations with **85.62% and 89.32%** consistency, respectively.

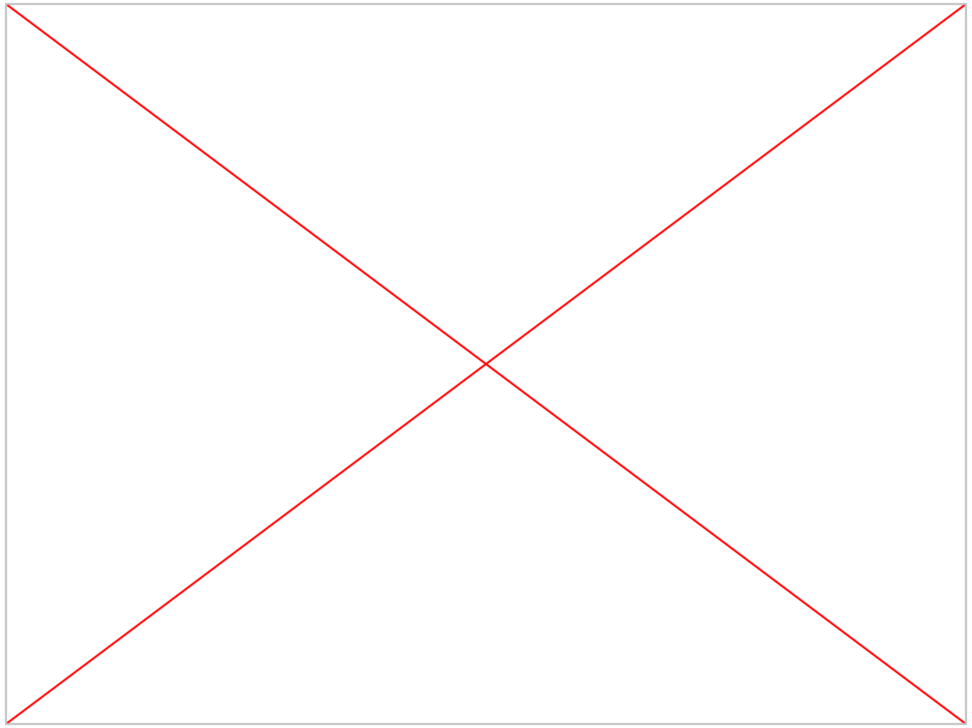
ChatGPT-4o’s accuracy was relatively lower in classifying collaborative programming behaviors (90.32%) and code quality (93.43%) but higher in identifying student roles (96.54%) and teacher scaffolding (97.42%).

Mitigation

To mitigate the impact of annotation errors, we added prediction percentage and explanations to ChatGPT-4o's annotations of collaborative programming behaviors, indicating the uncertainty of classification.

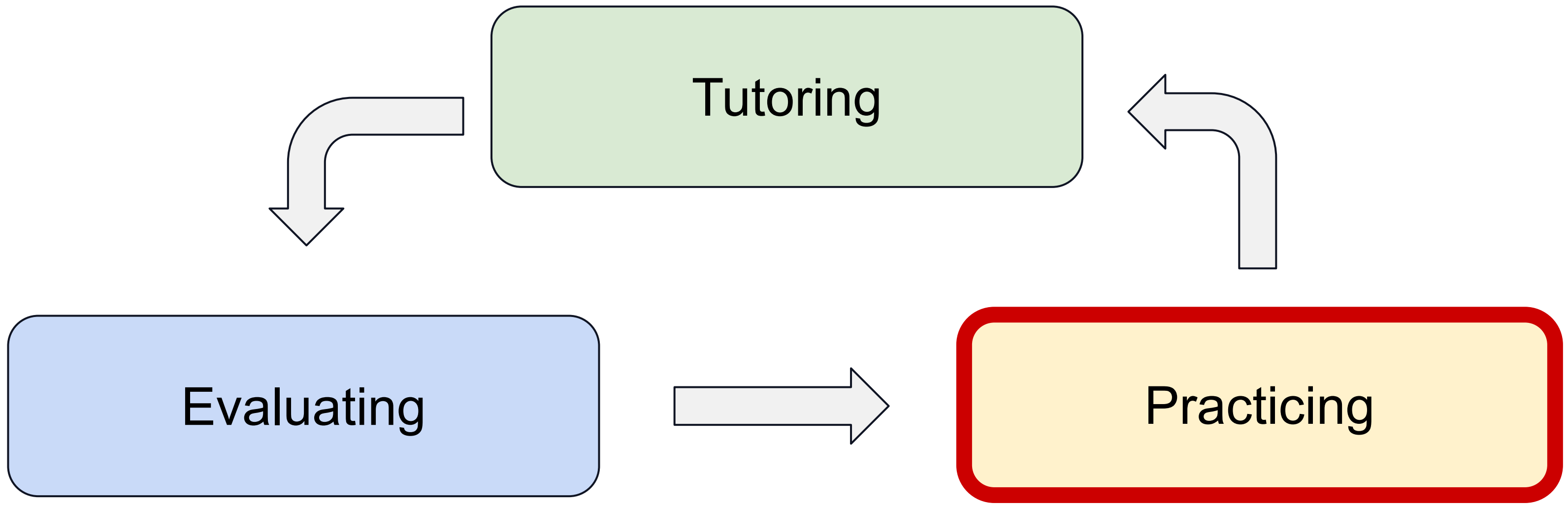
Similarly, we added explanations for code quality, providing more evidence for instructors during analysis.

Case Study and Evaluation

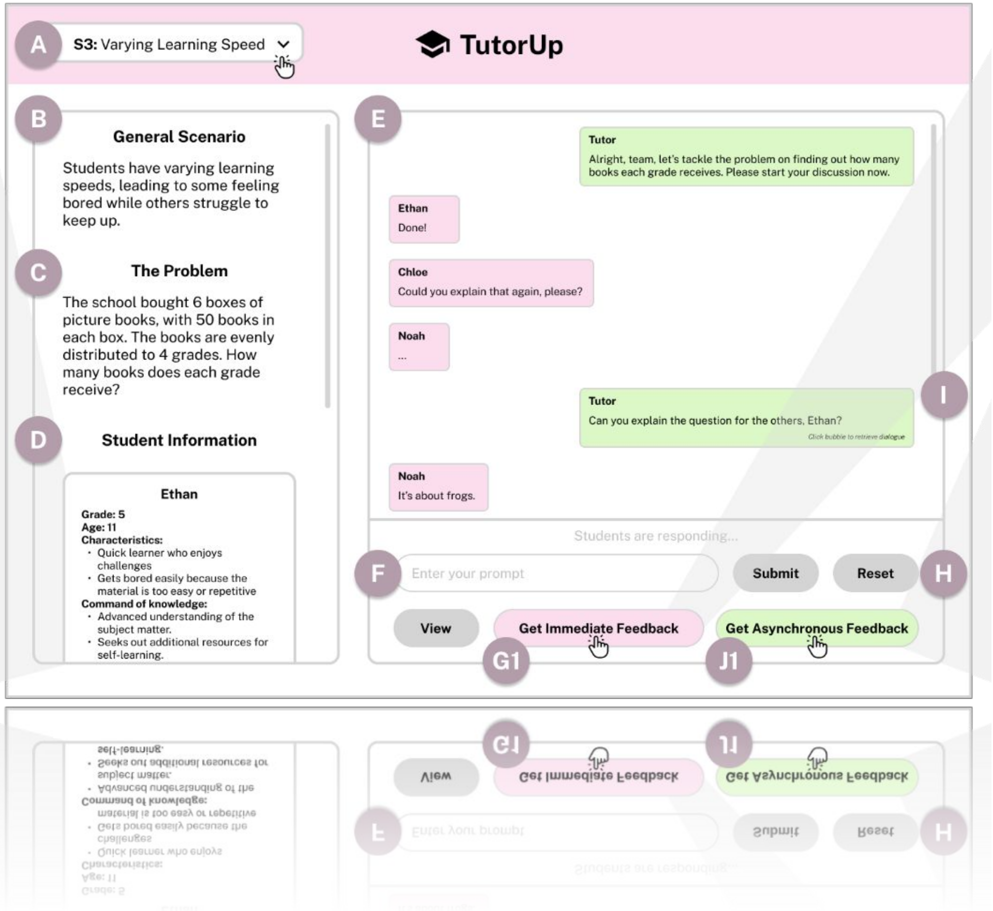


Generative AI's opportunities for Personalization

- Authoring personalized learning materials



- Evaluating student learning process
- Simulating real-world environment



TutorUp: What If Your Students Were Simulated? Training Tutors to Address Engagement Challenges in Online Learning

Sitong Pan, Robin Schmucker, Bernardo Garcia Bulle Bueno, Salome Aguilar Llanes, Fernanda Albo Alarcón, Hangxiao Zhu, Adam Teo, **Meng Xia***

CHI 2025 (Conditionally Accepted)

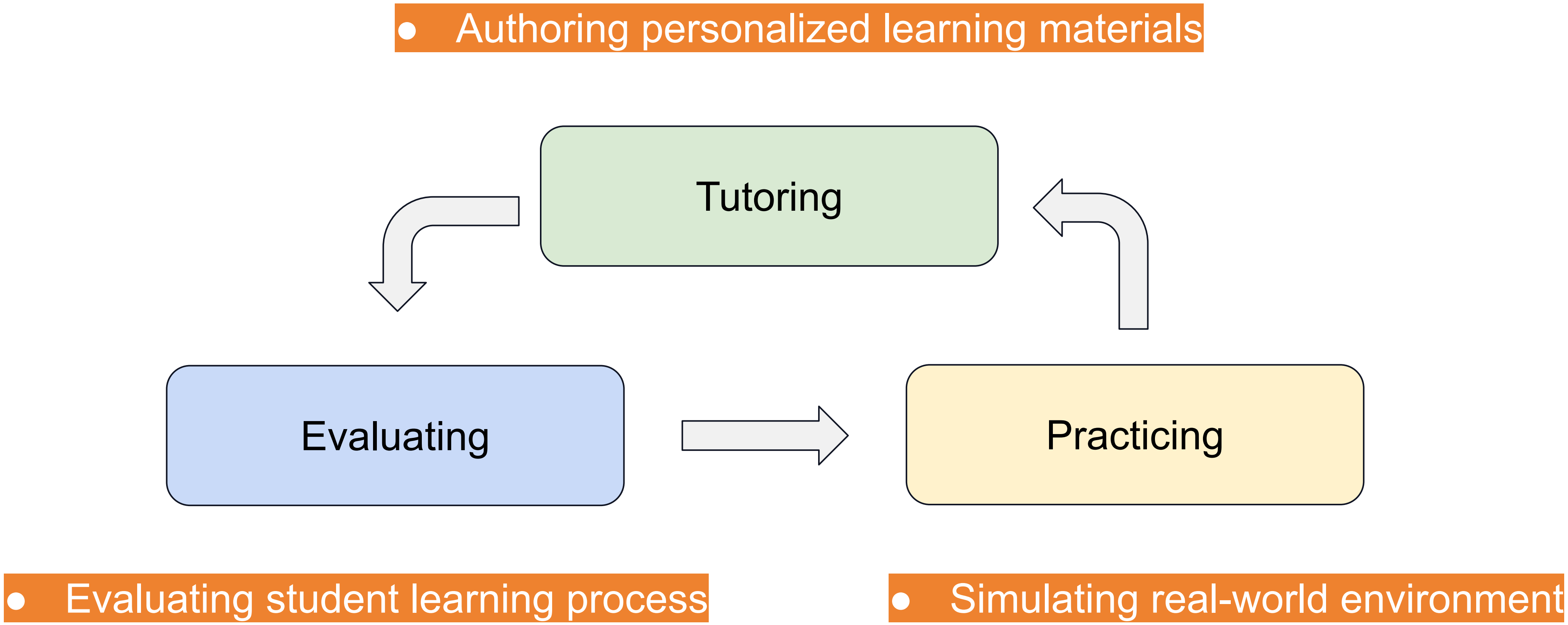
Engaging students is challenging in online learning

Identity Challenging Scenarios

Based on a formative study involving two surveys ($N1 = 86$, $N2 = 102$) on student engagement challenges, we summarize scenarios that mimic real teaching situations:

- Lack of Interest and Engagement
- Lack of Confidence
- Varying Learning Speeds
- Fatigue and Focus Issues

Generative AI's opportunities for Personalization



What are the Generative AI’s challenges?

- Improper use of AI (e.g., overreliance)
- Huluciation, content inaccuracy
- Lack of pedagogical guidance

StuGPTViz: A Visual Analytics Approach to Understand Student-ChatGPT Interactions

Zixin Chen, Jiachen Wang, **Meng Xia***, Kento Shigyo, Dingdong Liu, Rong Zhang, Huamin Qu

VIS 2024

Background: An inevitable trend in using LLMs

ChatGPT

Jennifer L. Steele, To GPT or not GPT? Empowering our students to learn with AI, Computers and Education: Artificial Intelligence, Volume 5, 2023, 100160, ISSN 2666-920X, https://doi.org/10.1016/j.caeai.2023.100160.

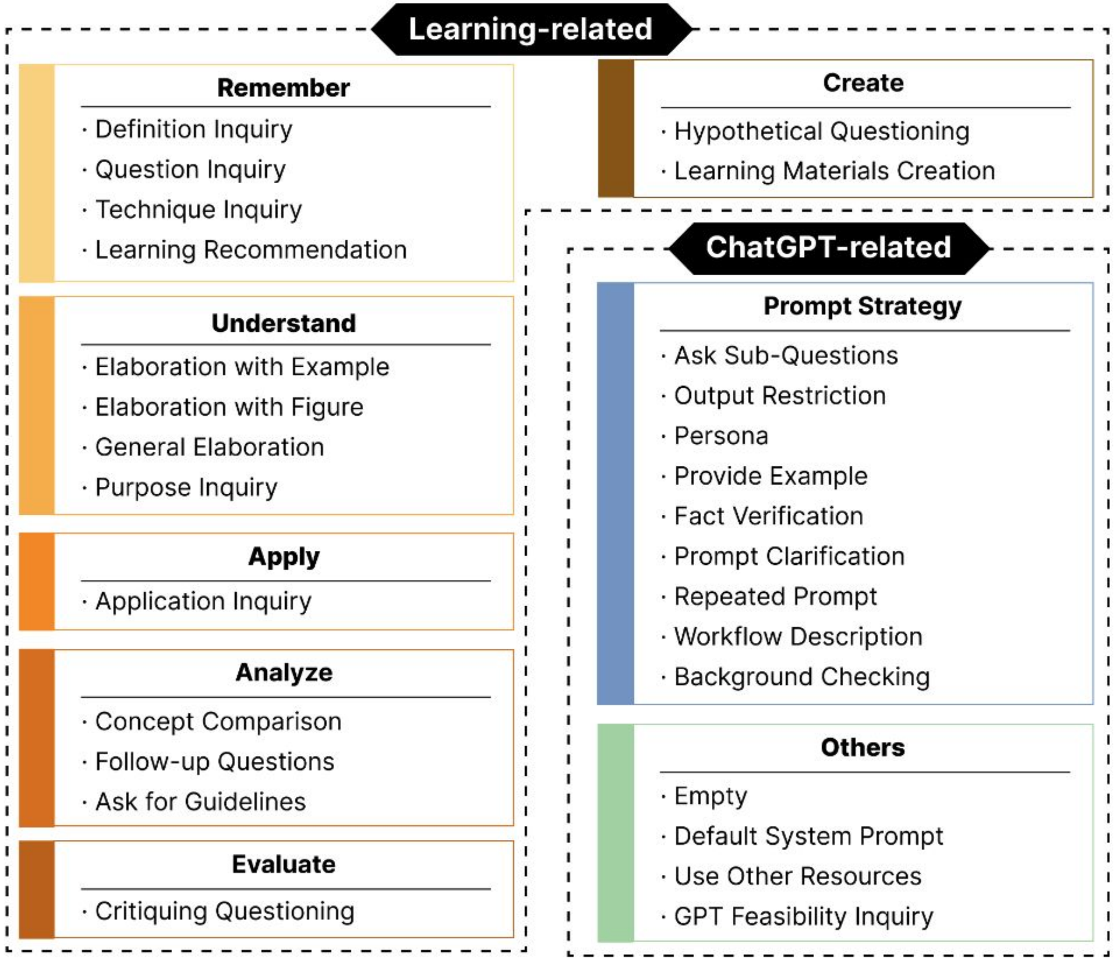
Concerns from instructors:

- How about the performance of these advanced AI tools?
- Using these advanced AI tools, can students practice higher-order thinking (e.g., independent thinking)?
- How can we better design tasks and guide students to use these advanced AI tools?

Integration of ChatGPT in Education

- We integrated ChatGPT into the curriculum of a postgraduate data visualization course for computer science majors in the first semester of 2024.
- Each in-class exercise session, we conducted the experiment during the last 40 minutes of the lecture, included a 10-minute self-learning segment with ChatGPT, a 25-minute task completion segment, and a 5-minute conversation log upload phase.
- 744 unique conversations with 2507 turns after filtering out the empty conversations and those unrelated to the learning tasks

Dataset Creation with Pedagogical Insights



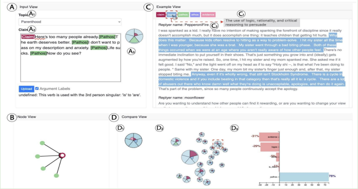

StuGPTViz
Filter View
A
Task Overview
a1
Search for Task ID
Set
Difficulty Distribution
Question Type Distribution
self_learning
apply
understand
analyze
create
evaluation
remember
Average Score Distribution
Student Overview
a2
Search for Student ID
Visualization Background
No
20
28
Yes
CS Background
No
23
25
Yes
ChatGPT Background
No
13
35
Yes
Average Score Distribution
Pattern View
B
StudentG
TaskG
Pattern Summary
Remember
Understand
Apply
Analyze
Evaluate
Create
Prompt strategies
Others
b1
b2
b3
ALL
analyze
evaluation
create
apply
IG
RL
Score
Task3
Task6
Task2
Task11
Task9
Task19
Task20
Task23
Task27
Task12
Task15
Task17
Task22
Task18
Task21
Task24
Pattern Nuance
b4
Select Pattern
L...
Pattern
C...
Avg.
1
Follow up Questions
17
0.806...
1
Question Inquiry
16
0.5615...
1
Definition Inquiry
12
0.326...
2
Follow up Questions
Follow up Questions
8
0.589...
b6
2
Follow up Questions
Definition Inquiry
7
0.870...
b7
1
Output Restriction
7
0.828...
First
Prev
Next
Last
b5
b8
QI
DI
FuQ
Detail View
C
Task Description
c1
This diagram illustrates the updated Visual Analytics Pipeline (named the knowledge generation model), which involves a series of operations, representations, and analytical stages. Using GPT's assistance, please identify and fill in the blank nodes which associated Analytical Stages (1 to 5) in the process.
Analytical Stages
Stage 1 is [Stage_A]
Stage 2 is [Stage_B]
Stage 3 is [Stage_C]
Stage 4 is [Stage_D]
Stage 5 is [Stage_E]
| Student Answer
Finding,Action,Insight,Hypothesis,Knowledge
Raw Conversation
c2
Do you know the diagram of updated visual analytics pipeline
Yes, the updated visual analytics pipeline typically consists of the following stages: 1. Data acquisition and integration 2. Data cleaning and preprocessing 3. Data transformation and feature extraction 4. Data mining and analysis 5. Visualization and exploration 6. Knowledge representation and reasoning 7. Decision making and action This pipeline is iterative and interactive, allowing analysts to refine their understanding of the data and adjust their analysis and visualization approaches as needed.

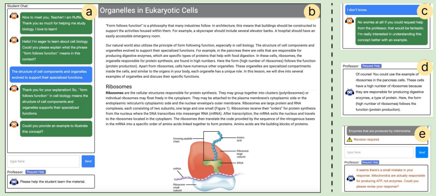
Evaluation and Result

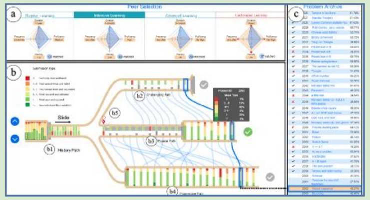
- Students’ learning perspective:
  More than 90% students enjoy using ChatGPT in their learning process
- ChatGPT performance:
  Strong positive correlation between the IG (information gain) metric and experts’ judgment of ChatGPT’s response quality
- Expert interviews:
  “The ability to discern students’ overall cognitive level at a glance is highly appreciated.”

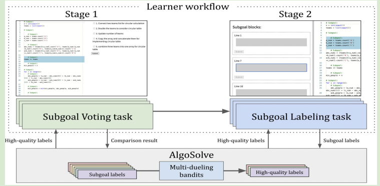
  “The workflow’s logical progression and the interconnection of each view were particularly impressive, enabling a diverse analytical focus through a unified procedure.”

Tutoring

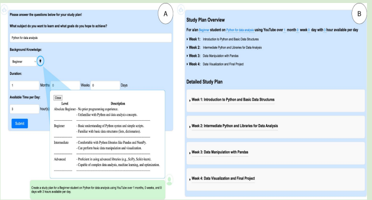

Persua (CSCW 2022)


Ruffle&Riley (AIED 2024)


Peerlens (CHI 2019)

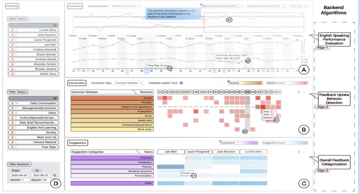


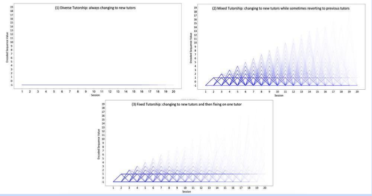
AlgoSolve (CHI 2022)

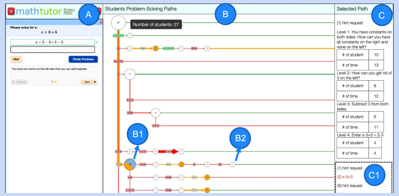

PlanGlow (Ongoing)

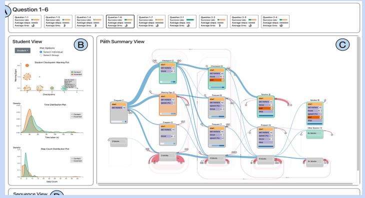
"Gaming the system" (L@S 2020)

Evaluating


RLens (L@S 2022)

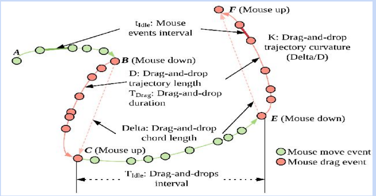

Distributed Tutorship (LAK 2022)


SolutionVis (AIED 2023)

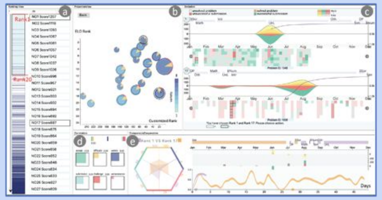

BlockLens (L@S 2022)

Mobile MOOCs (CHI 2022, **Best Paper Award**)

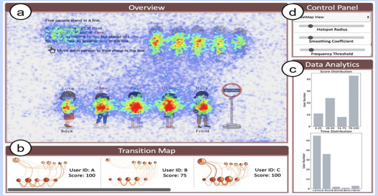

QLens (TVCG 2021)



Predication (LAK 2020)

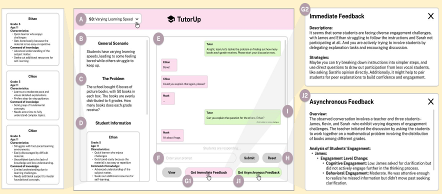

SeqDynamics (EuroVIS 2020)

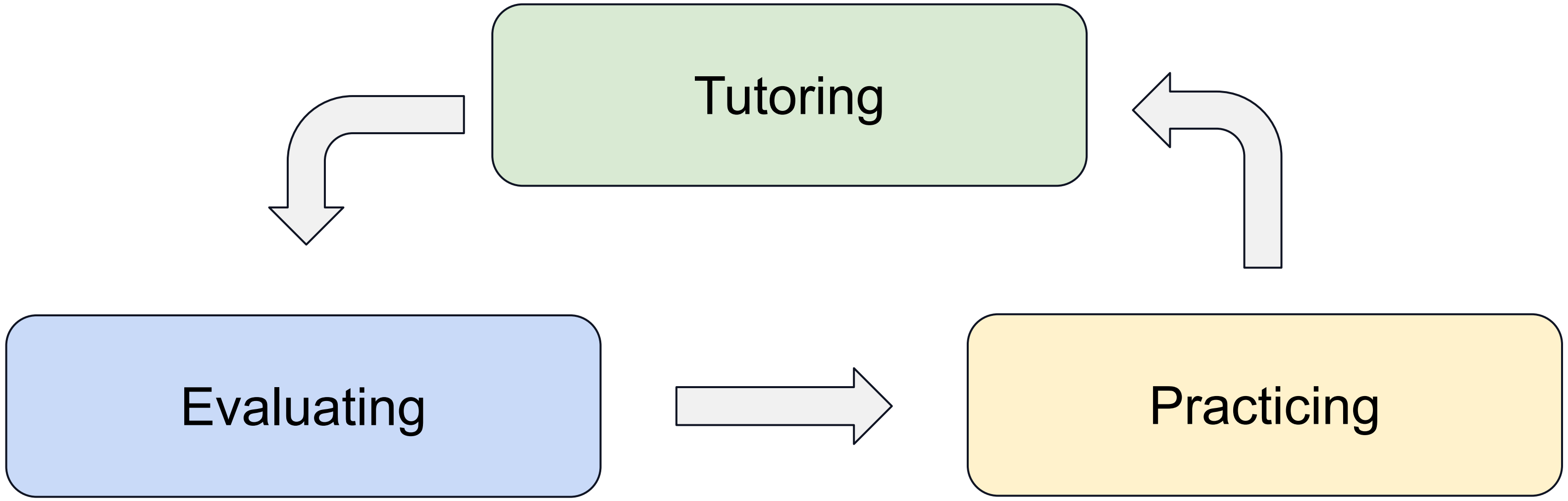

StuGPTVis (TVCG 2024)


Visual Analytics K-12 (VIS 2019, **Best Poster Award**)


CPVis (CHI 2024)

Practicing


TutorUp (CHI 2025)



Meng Xia
mengxia@tamu.edu
www.mengxia.org
DLab Dream Lab
TEXAS A&M UNIVERSITY

Use **visual analytics**, **AI**, and other **human-AI interaction techniques** and research metaphors to **upskill** educators and learners to better utilize data and AI for **Personalization@Scale**!

DL Dream Lab

News People Publications Contact

Graduate Students

Jiwon Chun

Master Student (2024 Spring)

Gefei Zhang

PhD Student (2024 Summer)

Hangxiao Zhu

PhD Student (2024 Fall)

Yi Wen

PhD Student (2024 Fall)

Jing Cao

Master Student (2024 Fall)

Fatemeh Mirhosseini

PhD Student (2025 Spring)

Undergrads

Shenming Ji

Undergraduate Student (2024 Summer)

Sitong Pan

Undergraduate Student (2024 Summer)

Adam Teo

Undergraduate Student (2024 Fall)

https://www.xiameng.org/DreamLab/