# Supporting Educators with Generative AI for Personalized Education

Meng Xia, Sep. 19, 2025

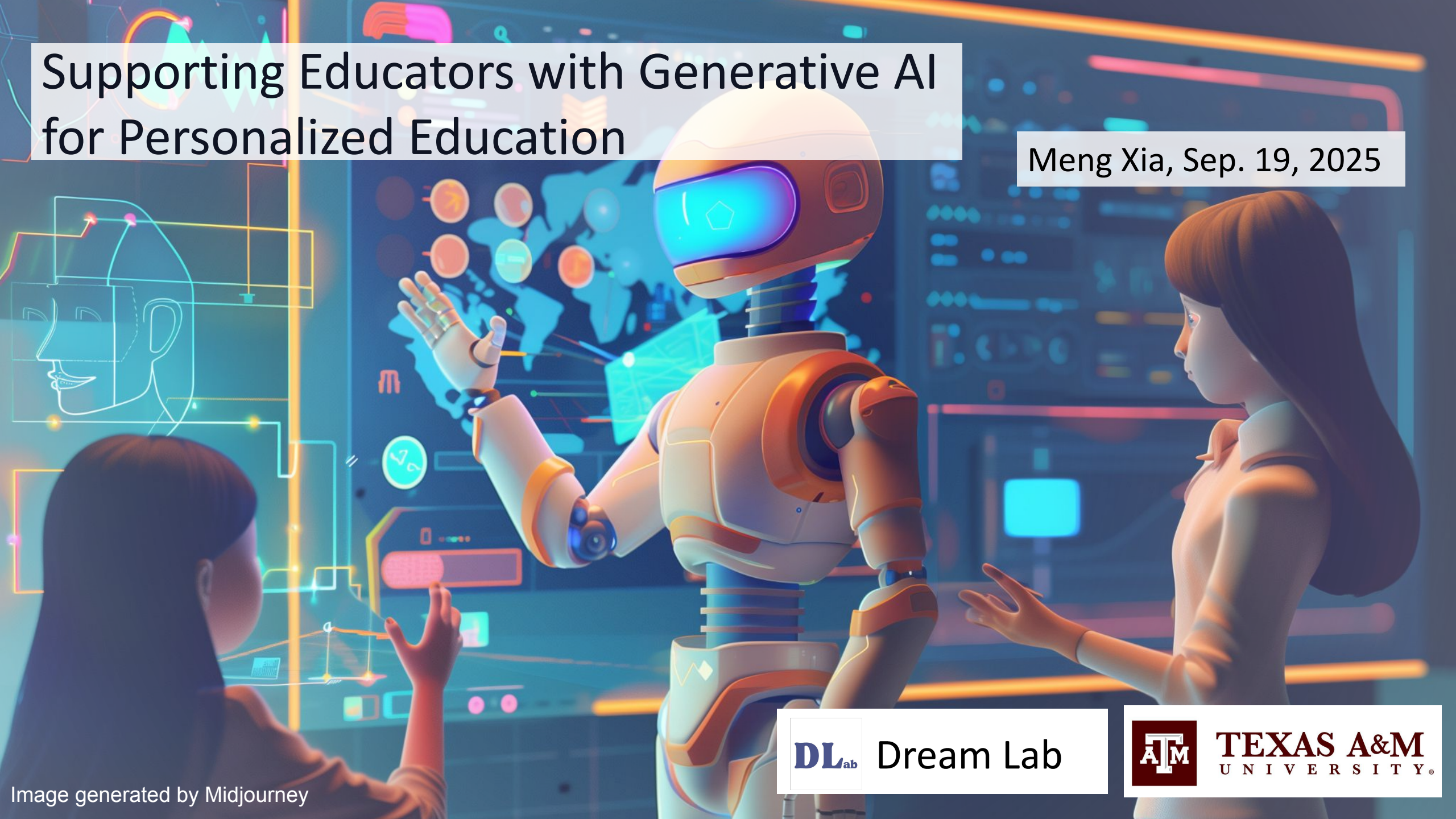


Image generated by Midjourney

Dream Lab

TEXAS A&M UNIVERSITY

# Why personalization?

Non-cognitive

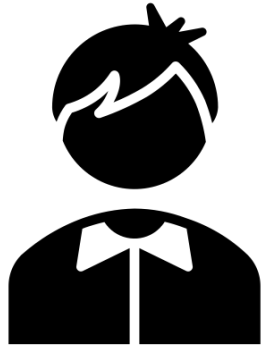
Motivation
(D'Mello, Lehman, Pekrun, & Graesser, 2014)

Self-regulation skills
(Aleven & Koedinger, 2002)

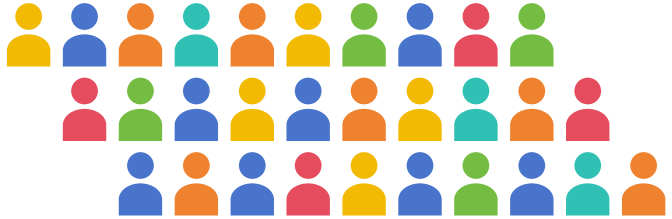
Cognitive

Knowledge
( Koedinger,. Stamper, McLaughlin, & Nixon, 2013;)

Problem-solving strategies, errors
(Adams et al., 2014)

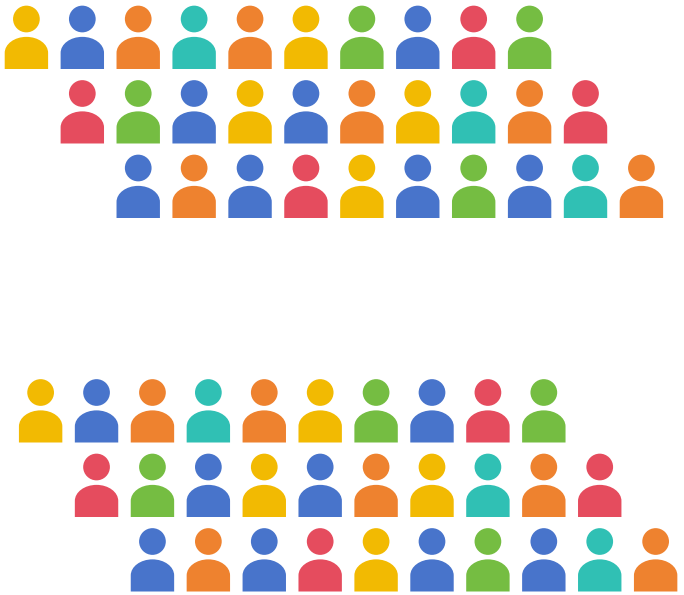


# **Personalization** is a Foundational Education Challenge



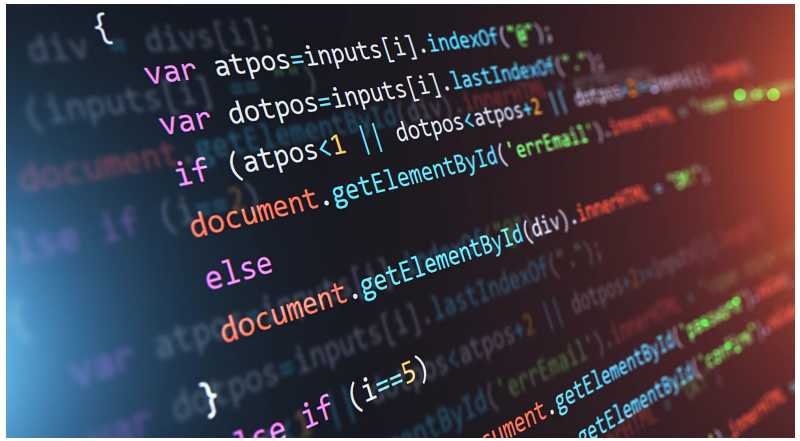
- Large amount of students
- No enough qualified teachers

# **Personalization** is a Foundational Education Challenge



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

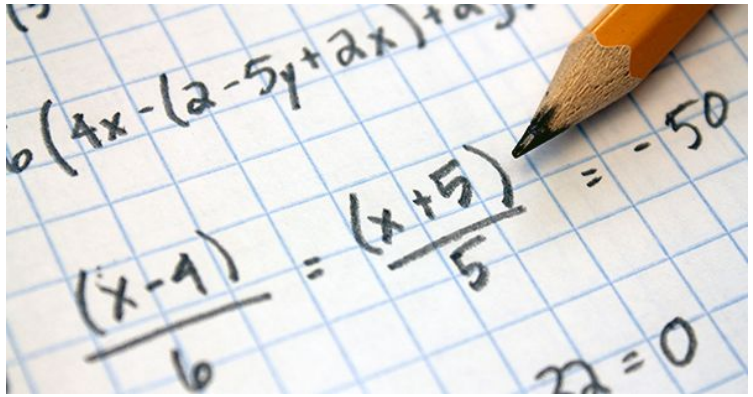
# **Generative AI** is popular




ChatGPT

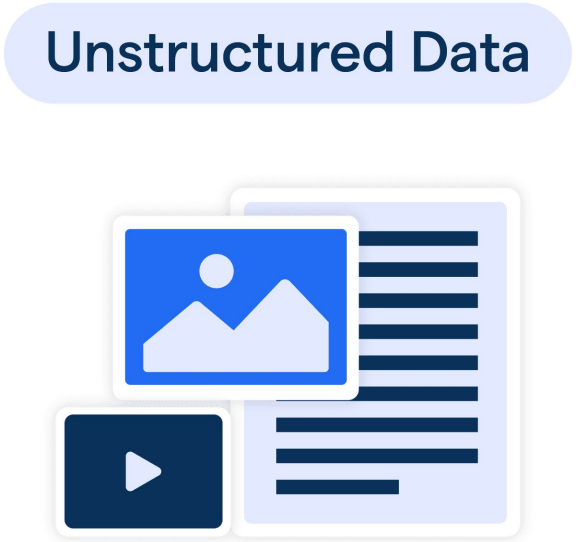
Midjourney

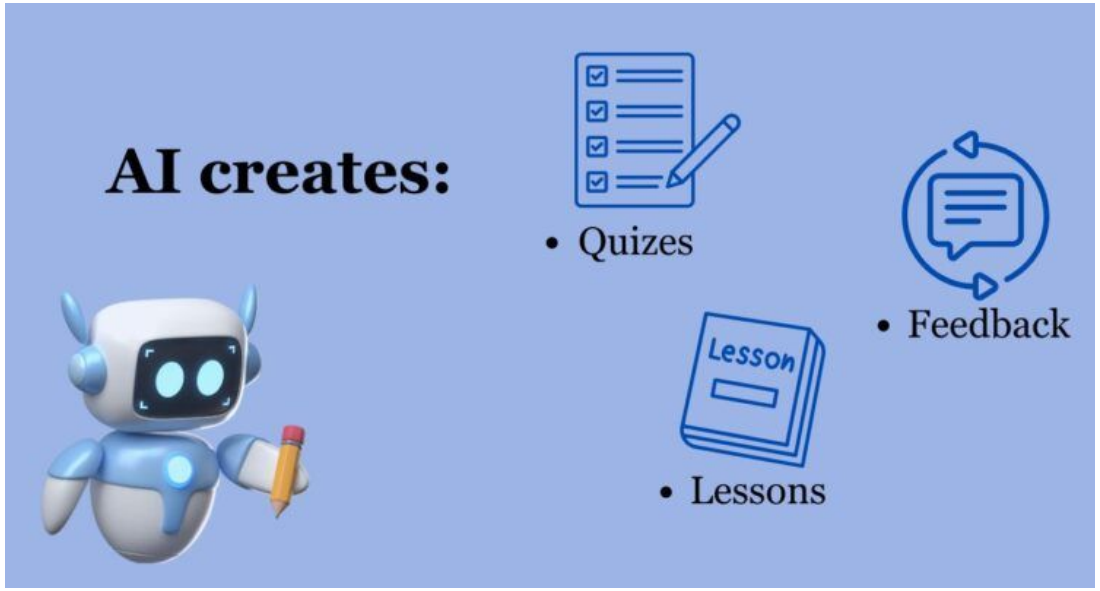
deepseek



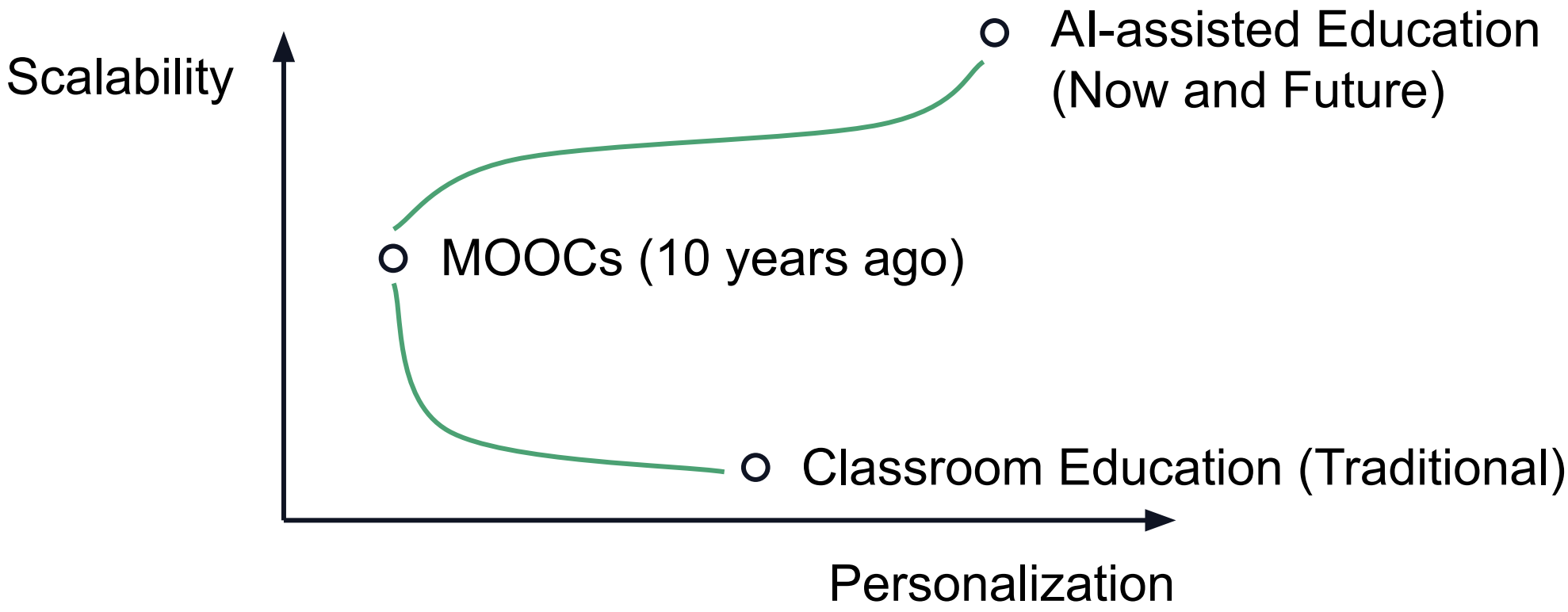

# Generative AI’s Characteristics

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

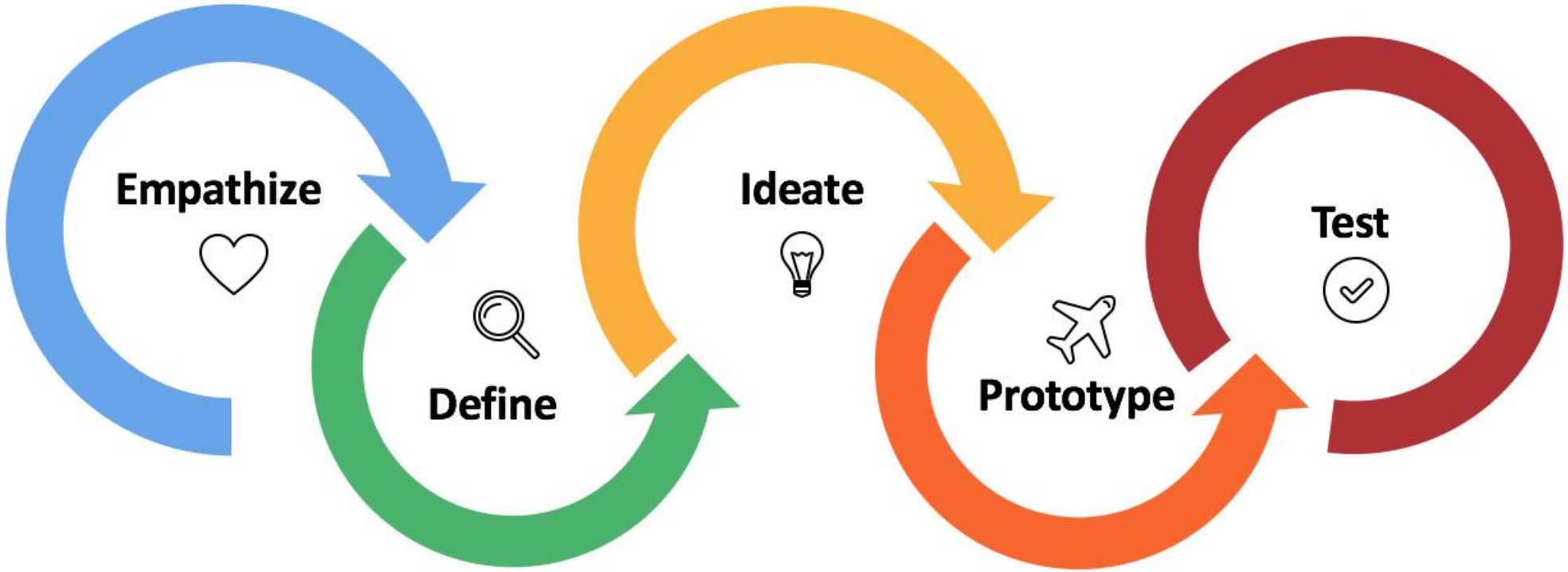


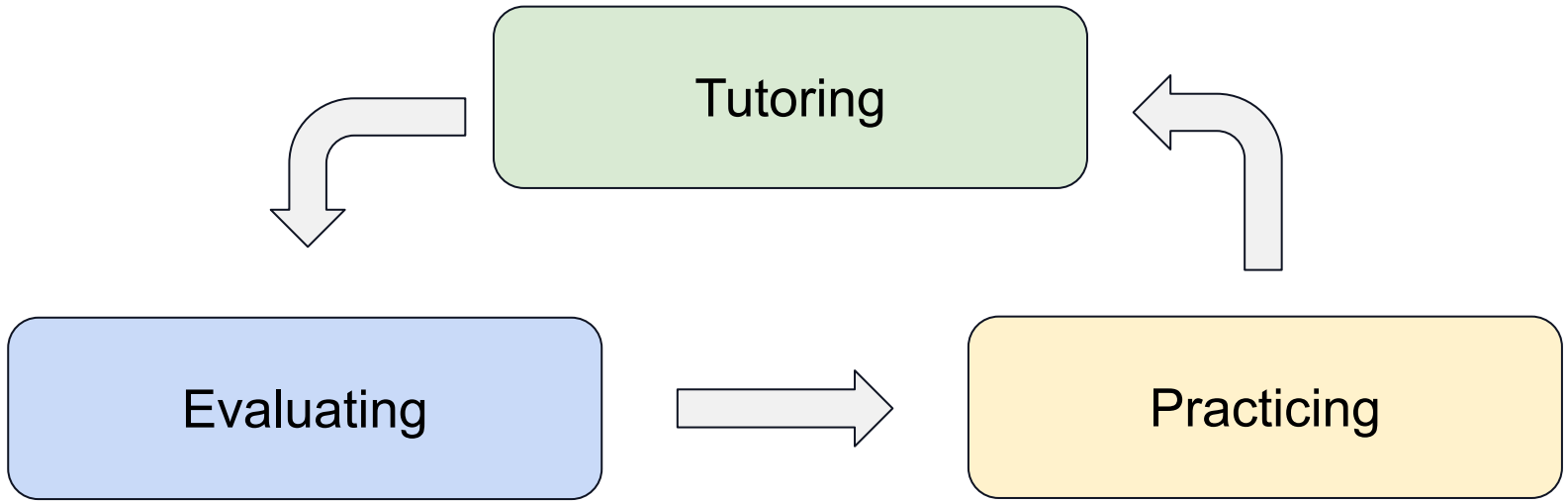
# Vision for education: **Personalization @ Scale**



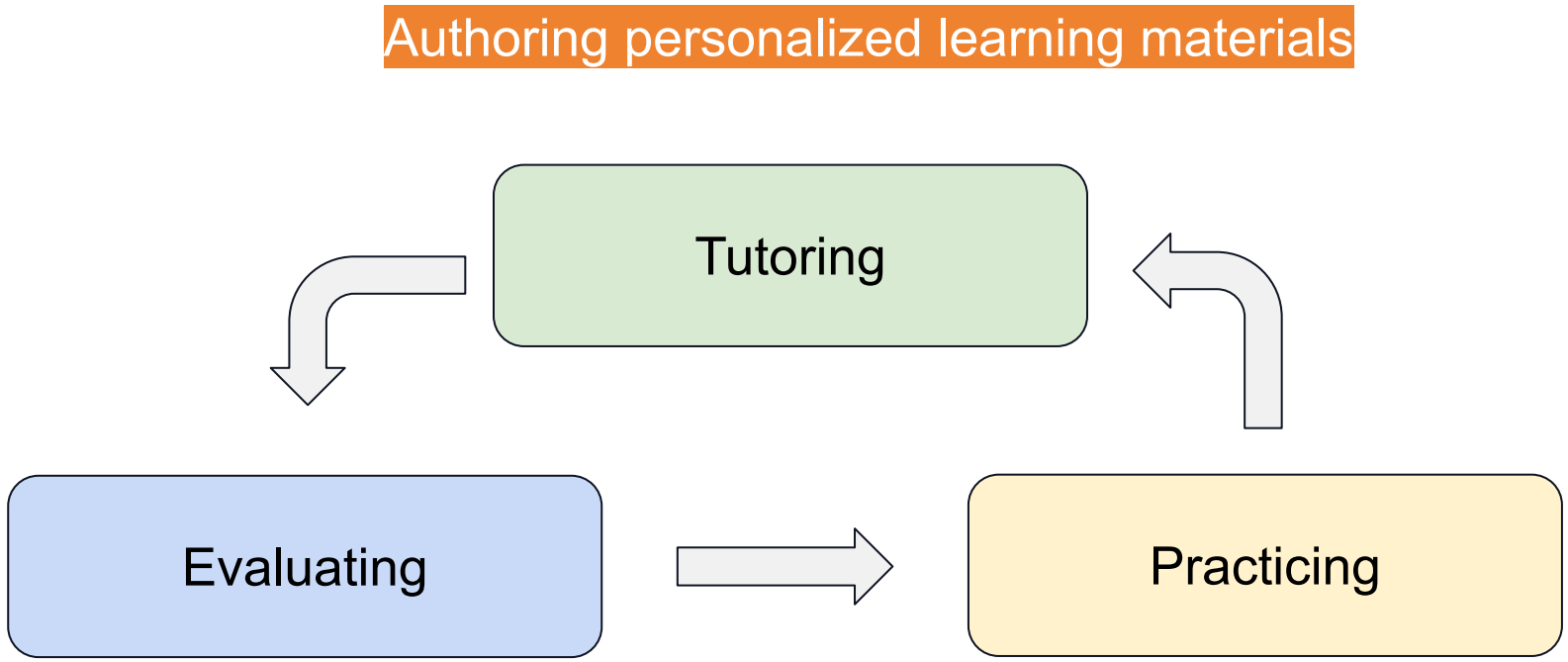
# User-Centered Design Process



# Educators’ Tasks

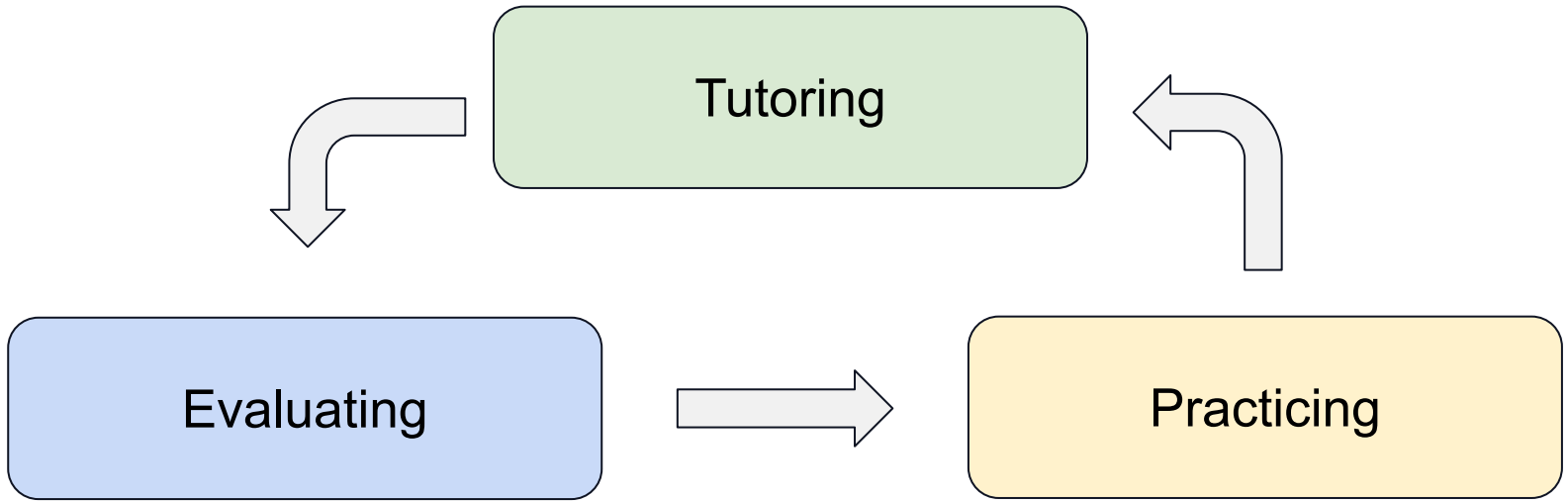


# Generative AI’s opportunities for Personalization



# Generative AI’s opportunities for Personalization

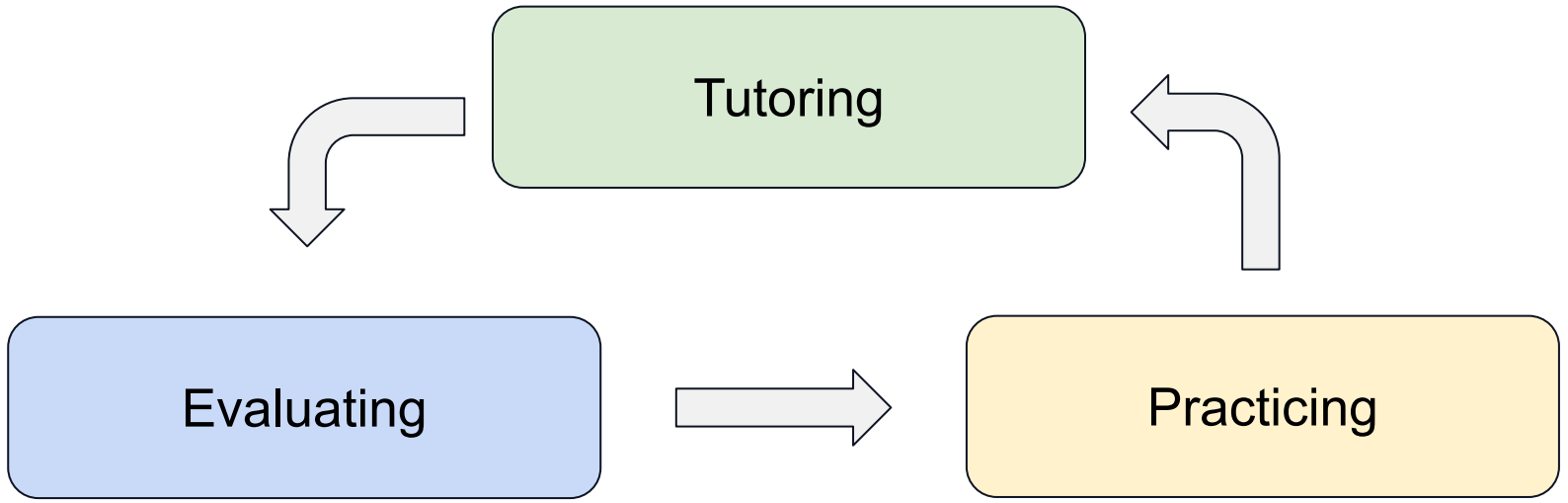
Authoring personalized learning materials




Evaluating student learning process

# Generative AI’s opportunities for Personalization

Authoring personalized learning materials

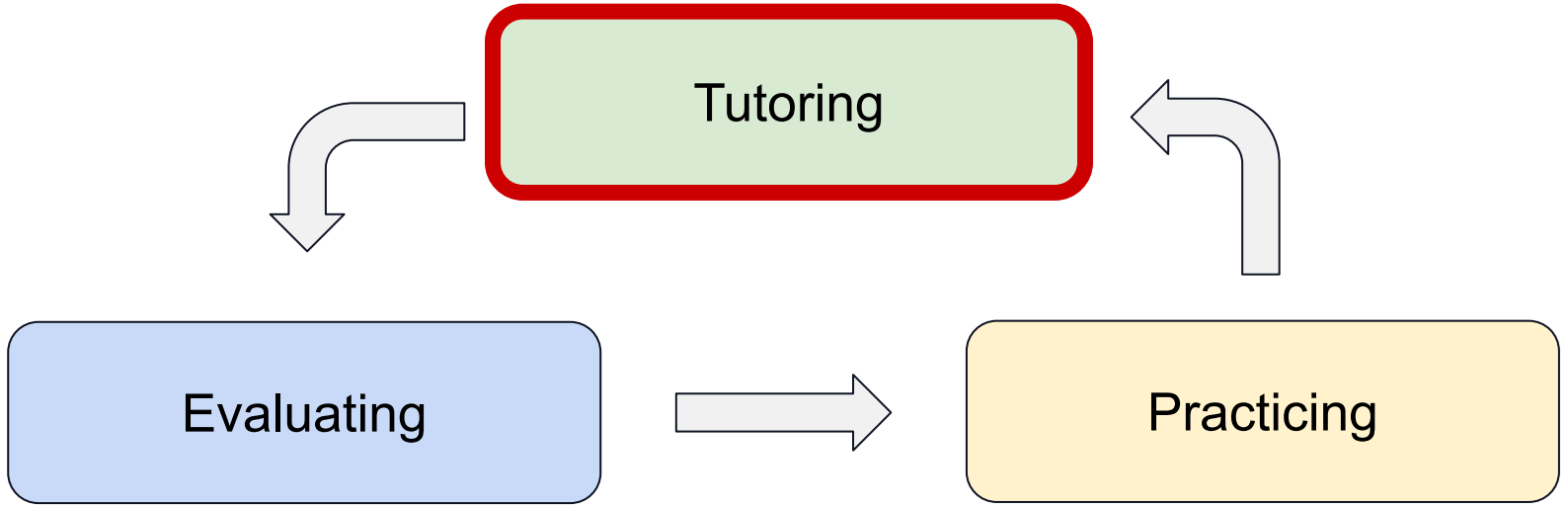



Evaluating student learning process

Simulating real-world environment

# Generative AI’s opportunities for Personalization

Authoring personalized learning materials




Evaluating student learning process

Simulating real-world environment

# VizQStudio

## A — Sample View

Model

Sample Question

Figure — Upload

## B — Feature View

### Question Features

History

- Context Complexity — High
  - Length of Context — low / medium / high
  - Domain-specific — Yes
- Cognitive Complexity — Analyze
- Hints/ Guidance — No

### Chart Features

History

- Chart Type — Bar Chart
- Data Complexity — medium
  - Num. Data — low / medium / high
- Visual Encoding — Sequential
  - Color — Rectangle
  - Shape

### Distractor Features

History

- Num. Distractors — 3
- Plausibility — low
- Similarity w. Answers — medium

### Knowledge Features

History

- Task Type — Retrieve Value, Find Extreme
- Misleader Type — Data-Visual Disproportion
- Design Principles — NA

Generate/Update

## C — Simulation View

V1 V2 V3 V4

Group 1 Group 2 Group 3 Group 4

### Student Groups

**C1**

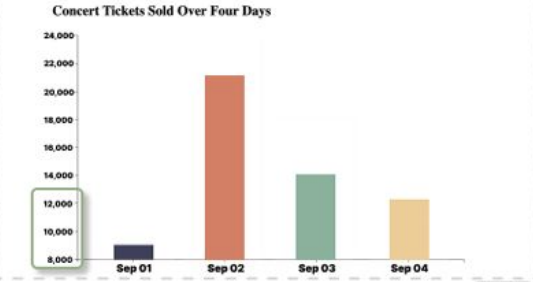
- 25.0% Group 1 — Major Dist.: CS: 5 — Grade Dist.: Master : 5 — Age: 24-26
- 35.0% Group 2 — Major Dist.: CS: 7 — Grade Dist.: Junior : 5, Sophomore: 2 — Age: 17-18
- 20.0% Group 3 — Major Dist.: Design: 2, CS: 2 — Grade Dist.: Senior : 4 — Age: 22-24
- 20.0% Group 4 — Major Dist.: Business: 2, CS: 2 — Grade Dist.: Senior : 2, Master : 2 — Age: 23-27

Learning Traits / Knowledge Points

### Generated Question

**C2**

**The chart displays the number of tickets sold for a concert during the first week of September. What is the number of concert tickets sold on Sep 03 as a proportion of that on Sep 02?**




▲ Editor

A: About 37.5%

B: About 66.7% ✓

C: About 50%

D: About 200%

**Hint:** *Pay attention to the Y-axis scale.* **C7**

**Explanation:** *The inappropriate Y-axis scale distorts the visual representation. Although the tickets sold on September 3 (around 14,000) are approximately 66.7% of those sold on September 2 (around 21,000), the bar height misleadingly suggests a value closer to 50%.*

Checked

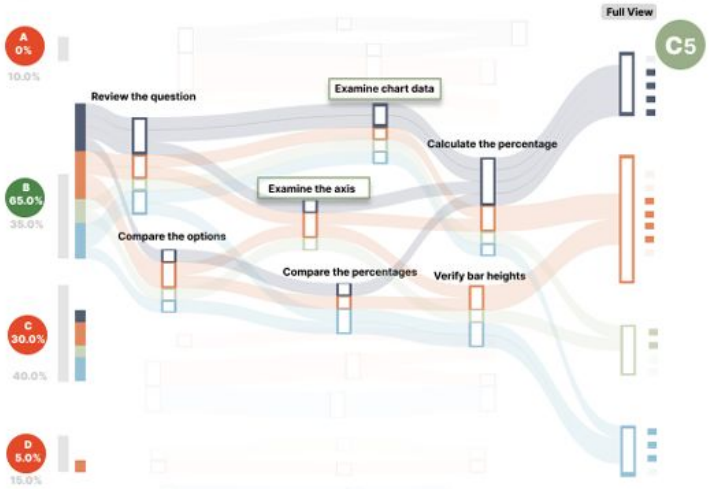
**C3** Question Revision — Update

### Simulation Results

**C4**

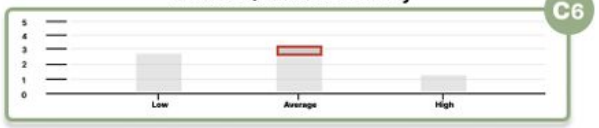

Full View

**C5**




### Overall Question Difficulty

**C6**




## D — Agent Profile View

*e.g., students mostly are a computer science background, with some design and business. They have limited foundational knowledge, and average learning traits ....*

Agents Profiles — Update

## E — Agent Setting View

### Demographic Settings

*e.g., a class of 20 students, with roughly 50% from a CS background, 25% from Engineering, and 25% from Design, aged 18–30, at both undergraduate and master's levels...*

**Demographic Attributes**

Major — Grade — Age

***Default Setting***: majority are CS undergraduates, with a small number of Business master's students, ages 18–30.

### Learning Traits Settings

*e.g., the Students generally have average logical reasoning and critical thinking skills typical of college students, with relatively lower visual processing and working memory...*

**Learning Traits**

Logical Reasoning — Visual Processing

Critical Thinking — Working Memory

***Default Setting***: normal distribution of each traits among the students

### Knowledge Points Settings

*e.g., the students generally have basic skills and understanding of "Retrieve Value" and "Find Extreme," but show limited understanding of "Characterize Distribution."*

**Knowledge Points**

Retrieve Value — Determine Range

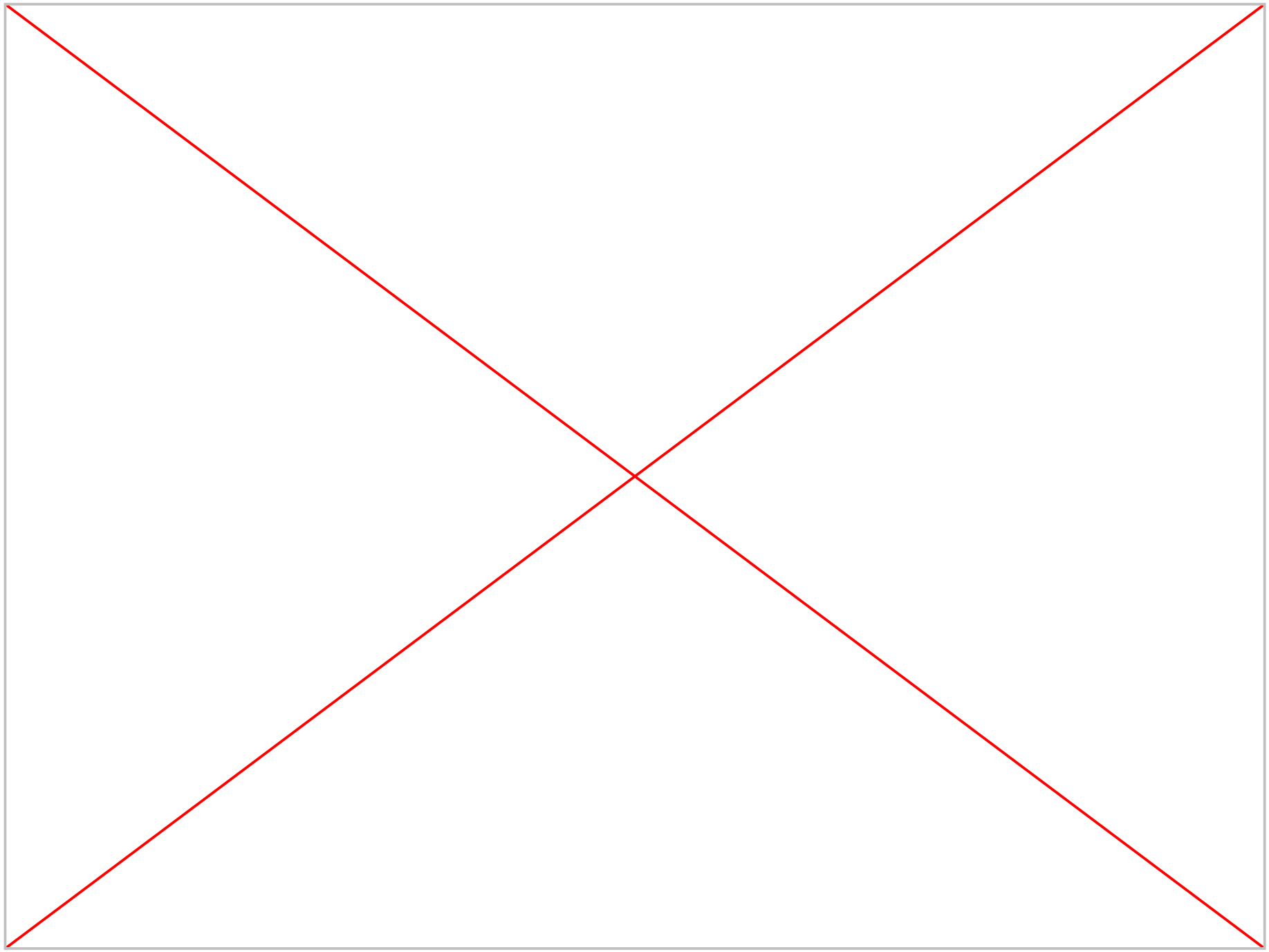
Find Extreme — Characterize Distribution

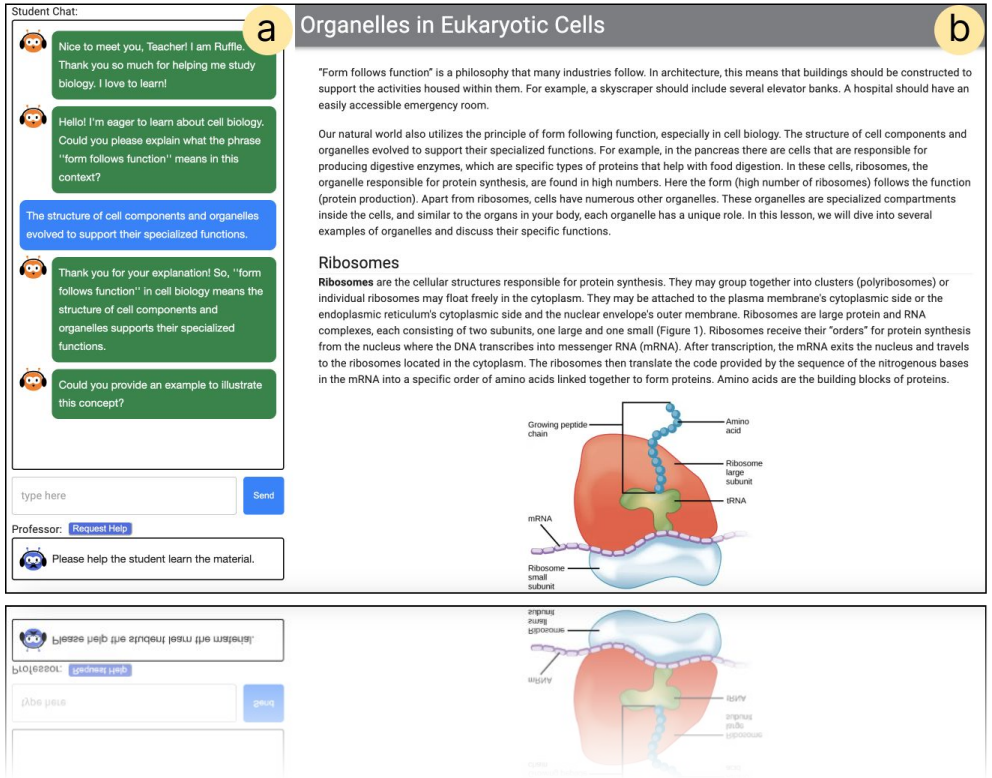
Find Anomalies — Make Comparisons

Find Clusters — Find Correlations/ Trends

***Default Setting***: normal distribution of student understanding for each knowledge point

Count: — Generate






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

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

AIED 2024

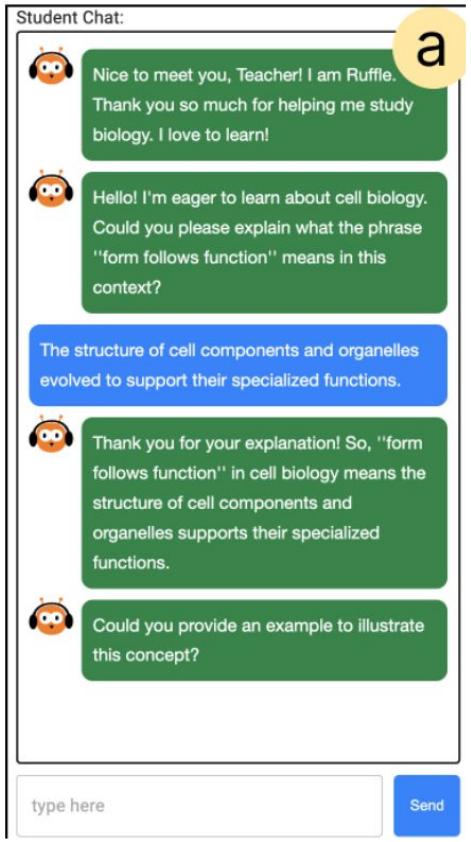
# LLMs can provide components to ITS designers:

- 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?

# Ruffle&Riley: User Interface

*Free-form dialog*




*Pre-existing textbook content*

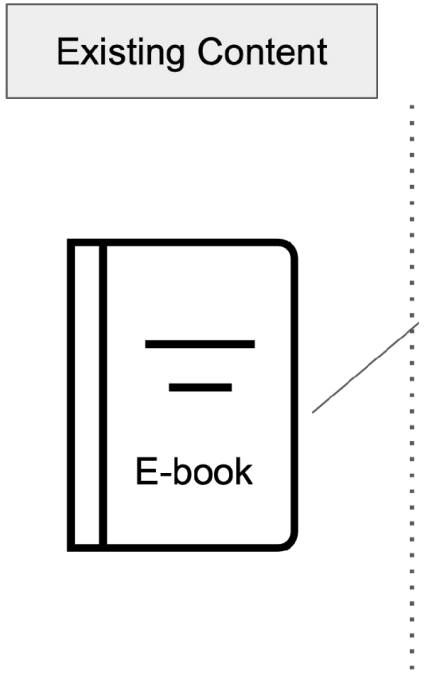
*Intelligent Feedback*

*On demand assistance*

*Misconcept. Correction*

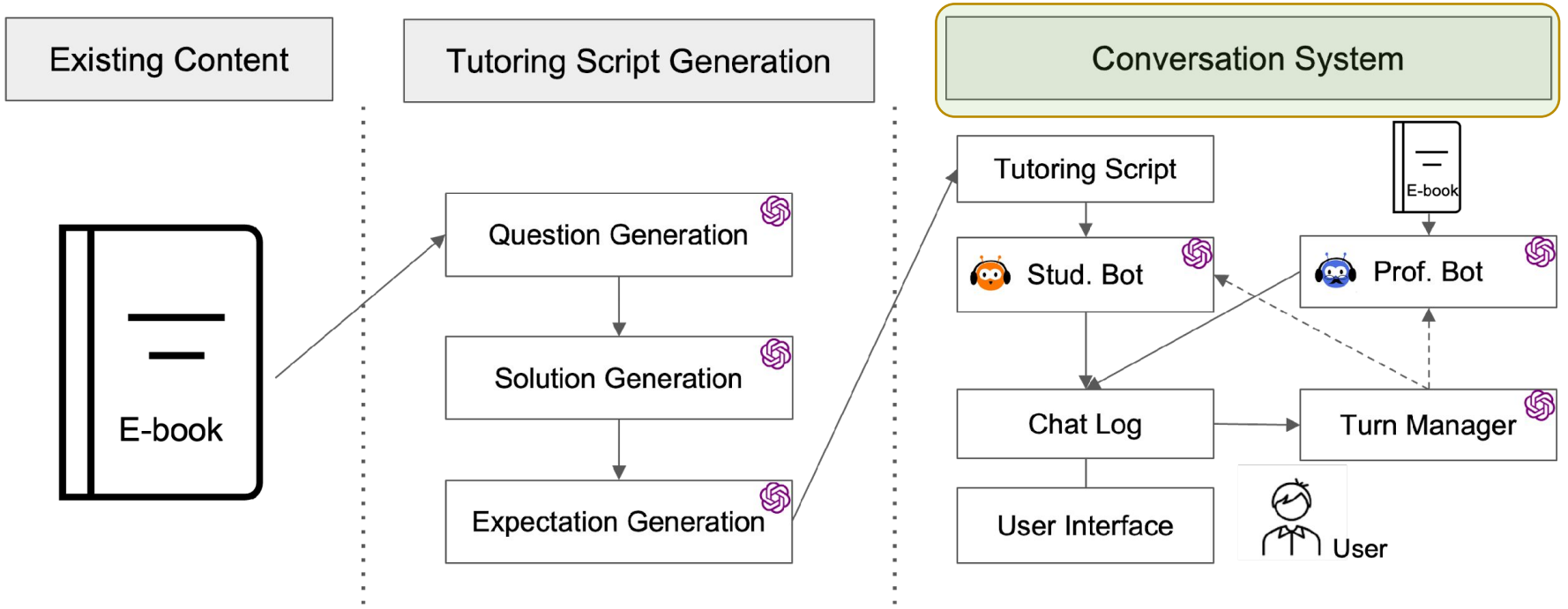
User Interface

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




System Architecture

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

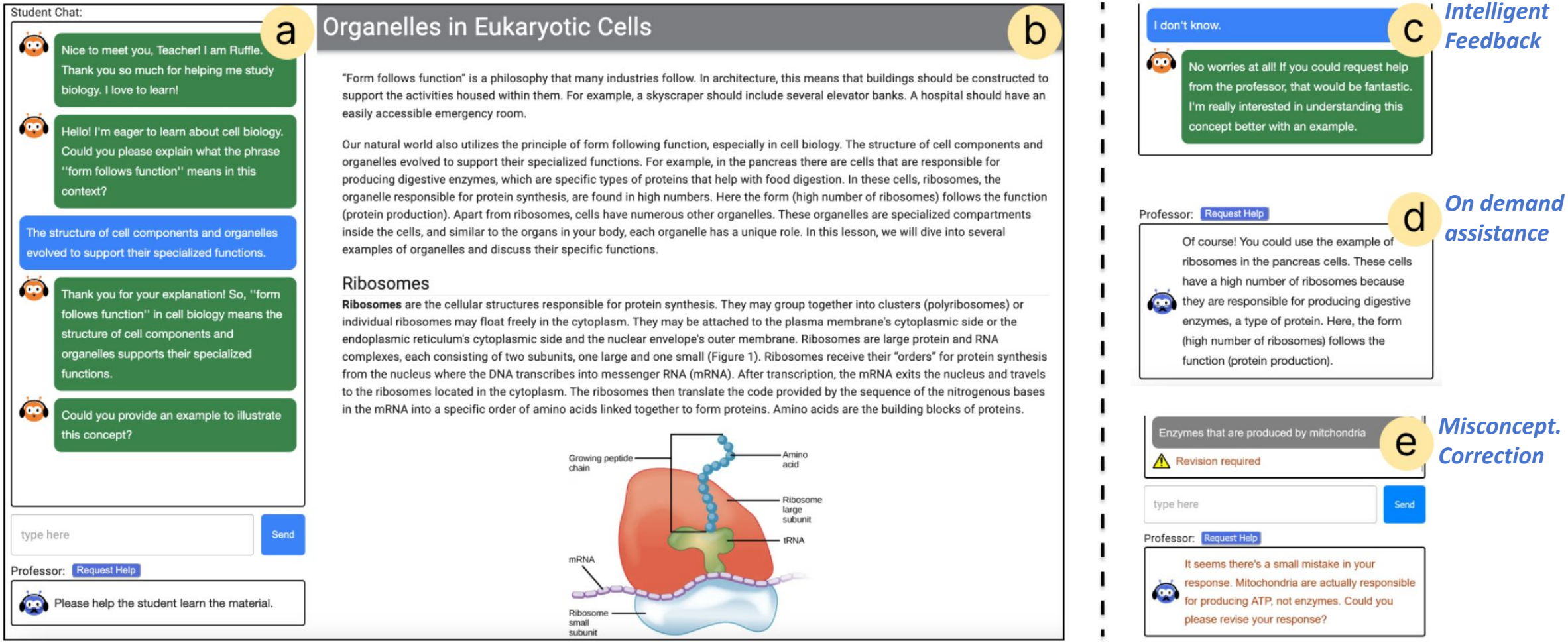



System Architecture

Ruffle&Riley: User Interface

*Free-form dialog*

*Pre-existing textbook content*




*Intelligent Feedback*

*On demand assistance*

*Misconcept. Correction*

User Interface

# *Findings of user study*

Learning Experience Survey

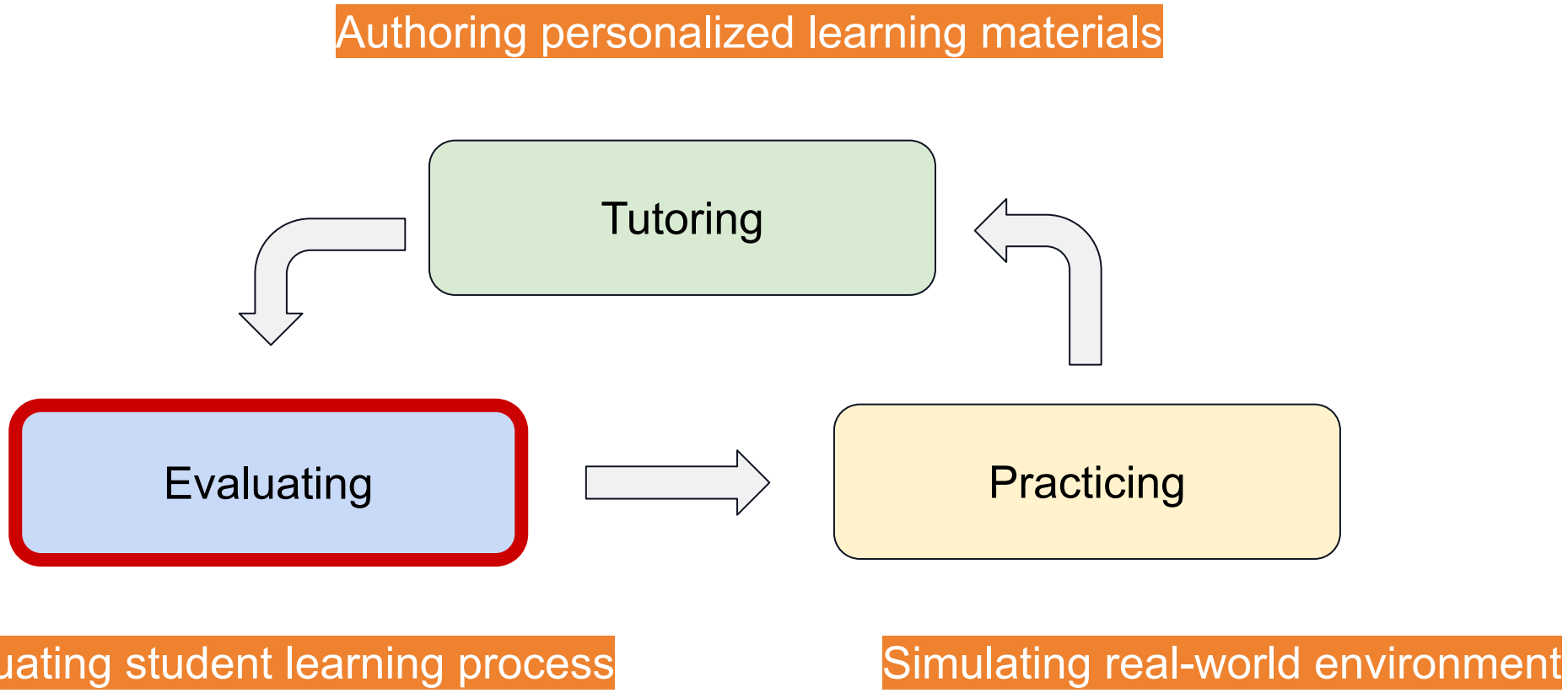
| Conditions | Learning Experience (1-strongly disagree, 7-strongly agree) | | | | | | |
|---|---|---|---|---|---|---|---|
| | Engagement | Understanding | Remembering | Interruption | Coherence | Support | Enjoyment |
| Reading | 4.33 ± 0.52 | - | - | - | - | - | - |
| Teacher Q/A | 5.0 ± 0.53 | 4.43 ± 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** |

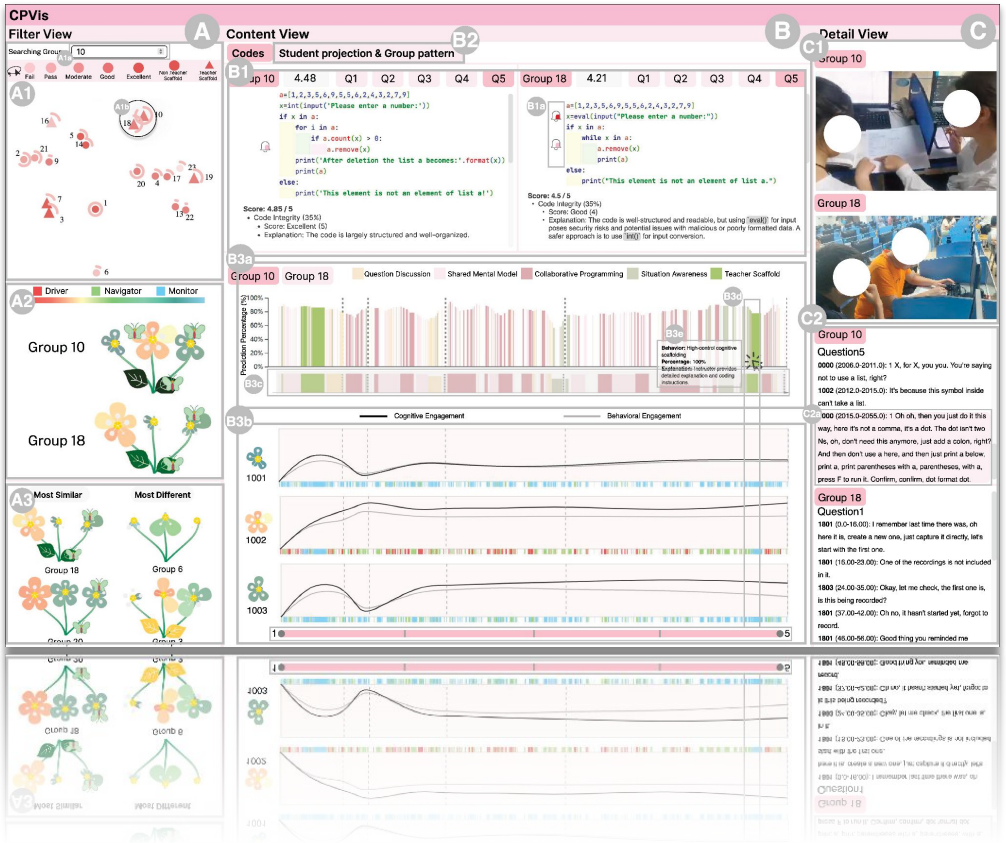
*Symbol "*" marks $p < 0.05$*

Ruffle&Riley: User Interface

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

# Generative AI’s opportunities for Personalization






# 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

# 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)
```

```
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()
```

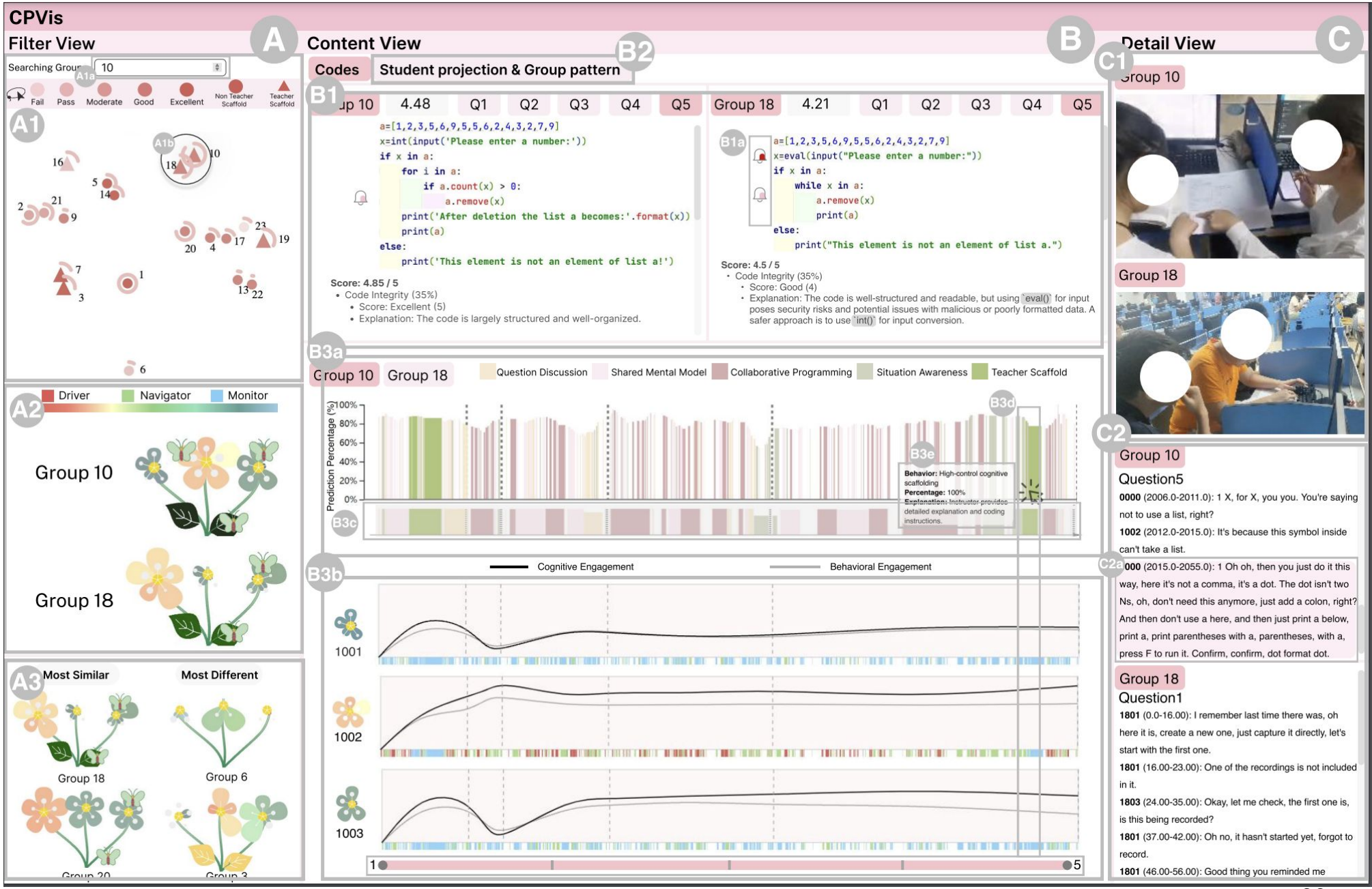
# Dataset

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

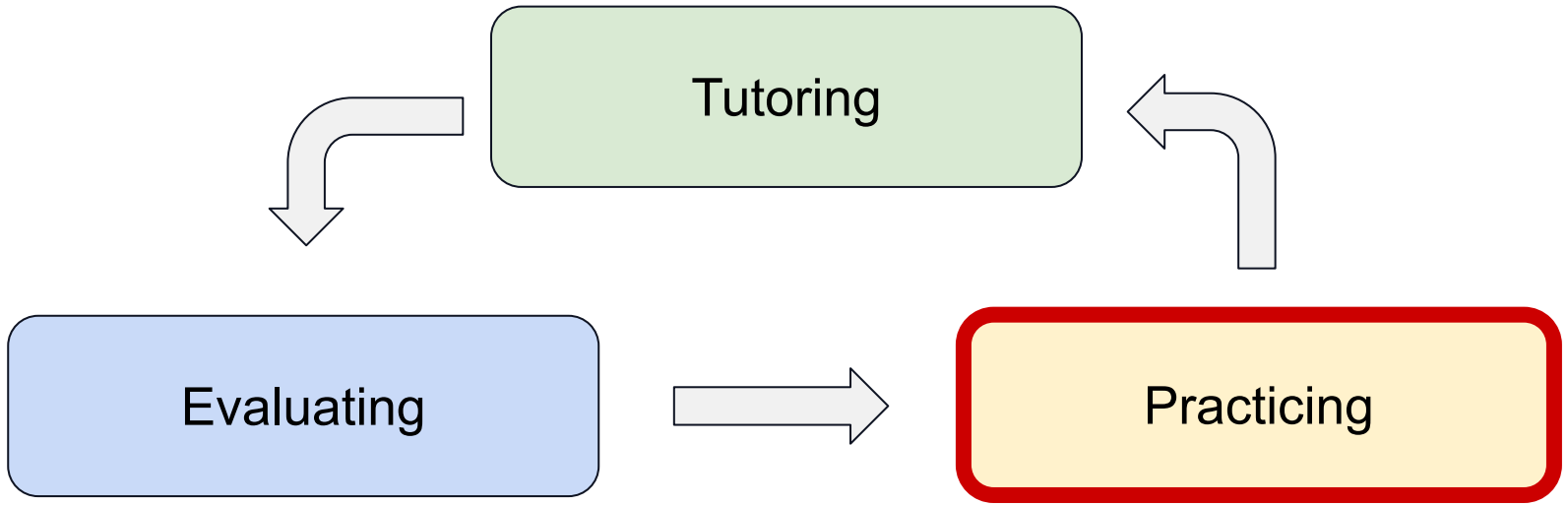
We evaluated LLMs' performance in code quality by comparing it to human-labeled (two experienced educators, I1, and I2) results.

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%).

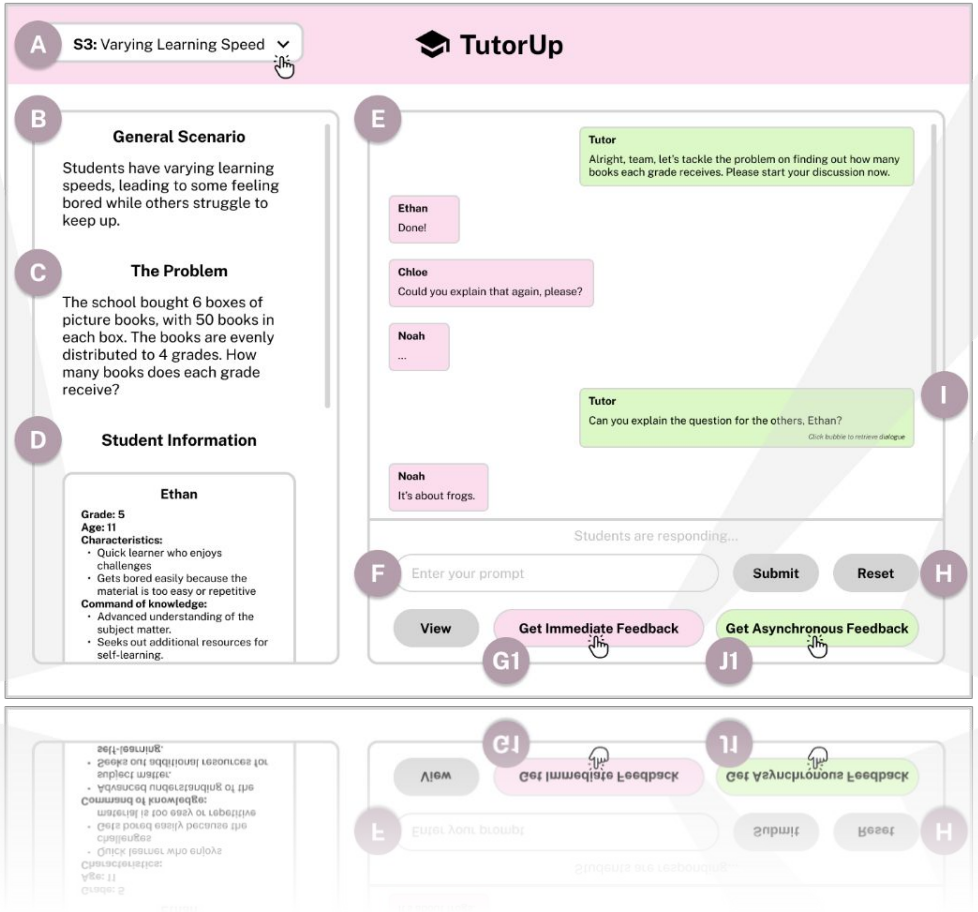
# Generative AI’s opportunities for Personalization

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***

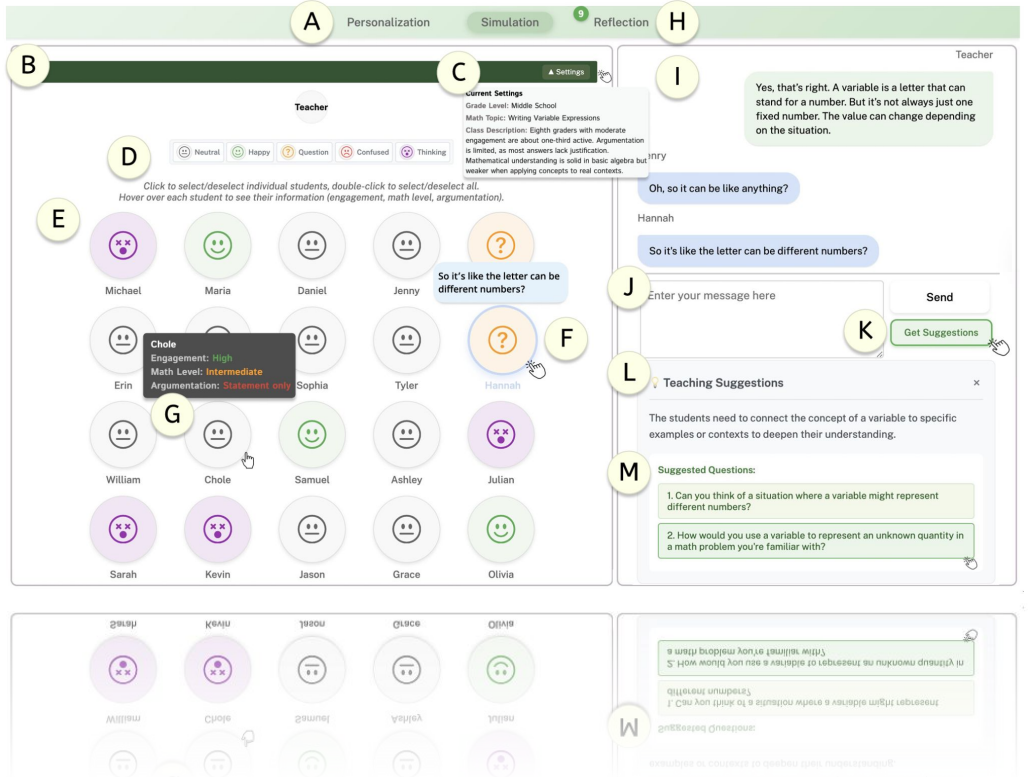
CHI 2025

# Engaging students is  challenging in online learning

# Identity Challenging Scenarios

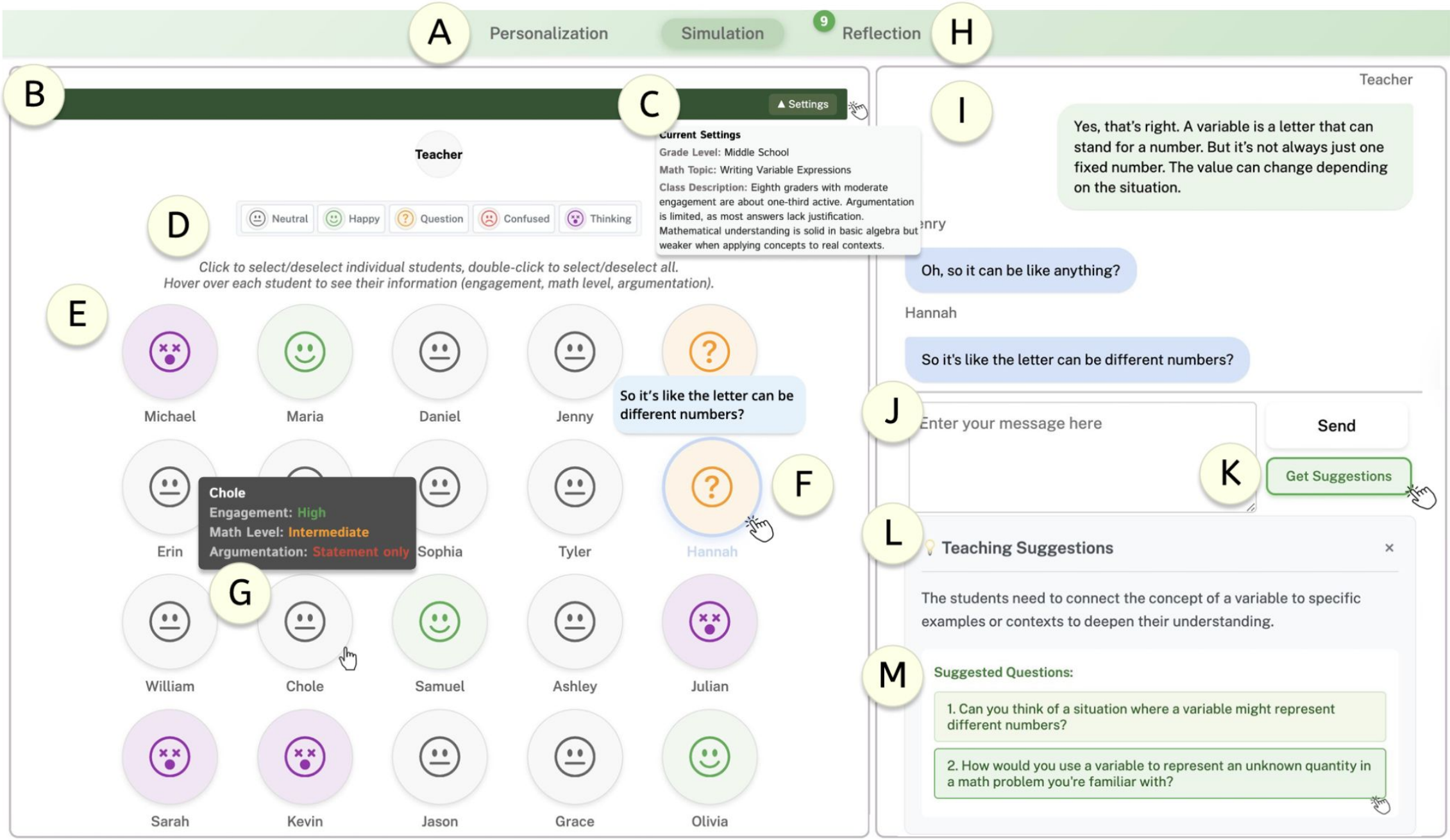
Based on a formative study involving two surveys ($N1$ = 86, $N2$ = 102) on student engagement challenges, we summarize scenarios that mimic real teaching situations:

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



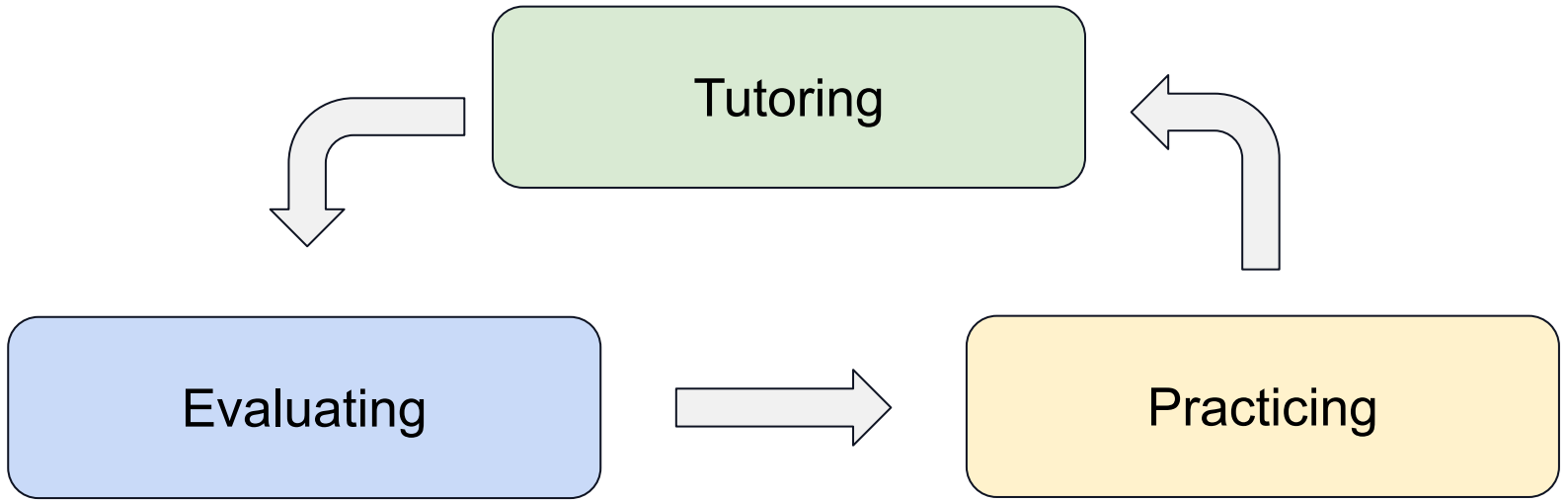

# Pre-service Teaching Training

(under review)


A
Personalization
Simulation
9
Reflection
H
B
C
▲ Settings
Current Settings
Grade Level: Middle School
Math Topic: Writing Variable Expressions
Class Description: Eighth graders with moderate engagement are about one-third active. Argumentation is limited, as most answers lack justification. Mathematical understanding is solid in basic algebra but weaker when applying concepts to real contexts.
Teacher
D
Neutral
Happy
Question
Confused
Thinking
Click to select/deselect individual students, double-click to select/deselect all.
Hover over each student to see their information (engagement, math level, argumentation).
E
Michael
Maria
Daniel
Jenny
So it's like the letter can be different numbers?
Erin
Sophia
Tyler
Hannah
F
Chole
Engagement: High
Math Level: Intermediate
Argumentation: Statement only
G
William
Chole
Samuel
Ashley
Julian
Sarah
Kevin
Jason
Grace
Olivia
I
Teacher
Yes, that's right. A variable is a letter that can stand for a number. But it's not always just one fixed number. The value can change depending on the situation.
Oh, so it can be like anything?
Hannah
So it's like the letter can be different numbers?
J
Enter your message here
Send
K
Get Suggestions
L
Teaching Suggestions
The students need to connect the concept of a variable to specific examples or contexts to deepen their understanding.
M
Suggested Questions:
1. Can you think of a situation where a variable might represent different numbers?
2. How would you use a variable to represent an unknown quantity in a math problem you're familiar with?

# Generative AI’s opportunities for Personalization

Authoring personalized learning materials




Evaluating student learning process

Simulating real-world environment

# 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.

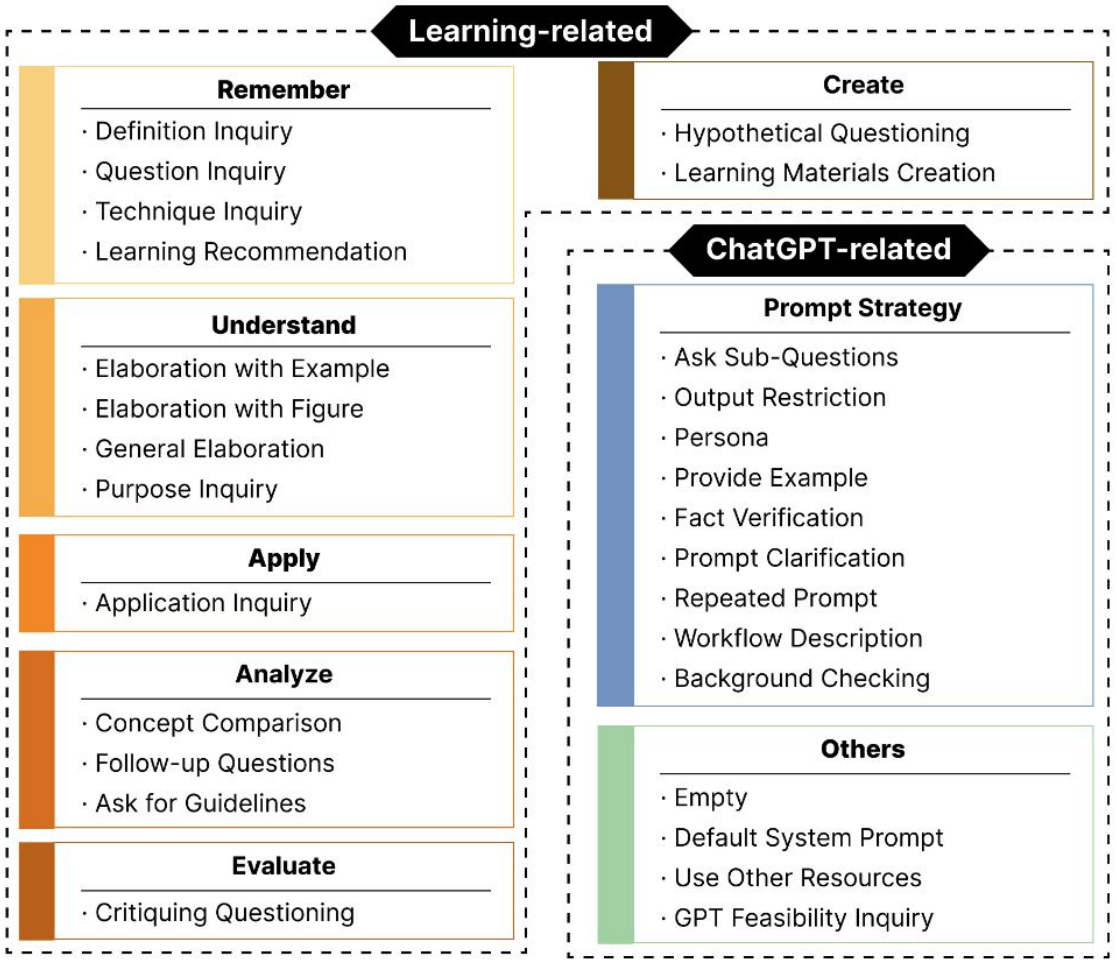
# Concerns from instructors:

- 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



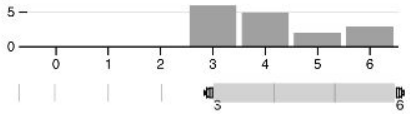
# StuGPTViz

## Filter View (A)

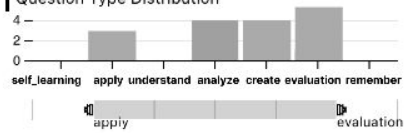
### a1 Task Overview

Search for Task ID | Set

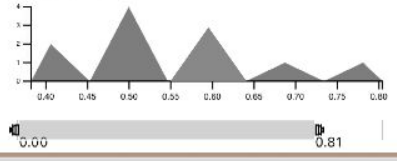
Difficulty Distribution




Question Type Distribution




Average Score Distribution




### a2 Student Overview

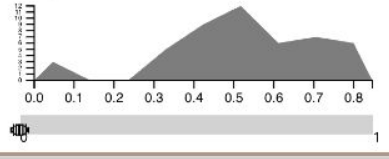
Search for Student ID | Set

Visualization Background


CS Background


ChatGPT Background

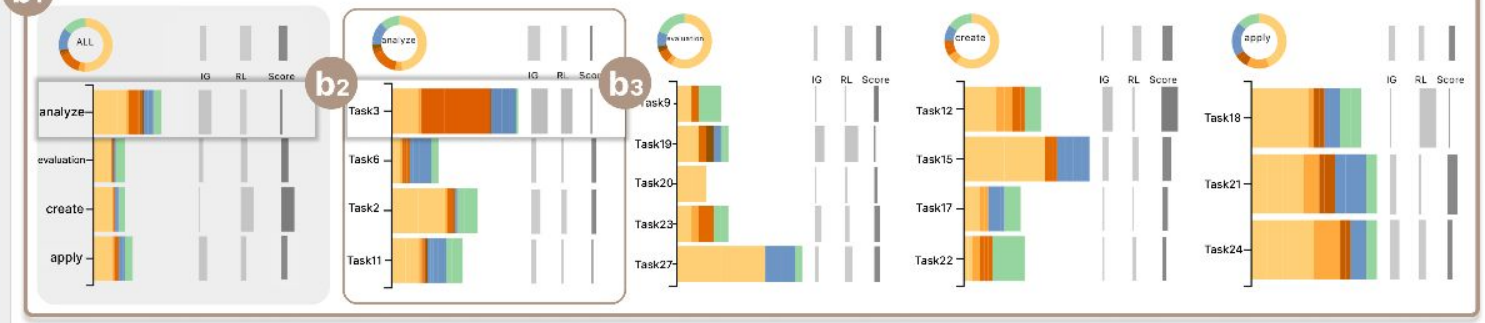

Average Score Distribution




## Pattern View (B)

StudentG | TaskG

Pattern Summary



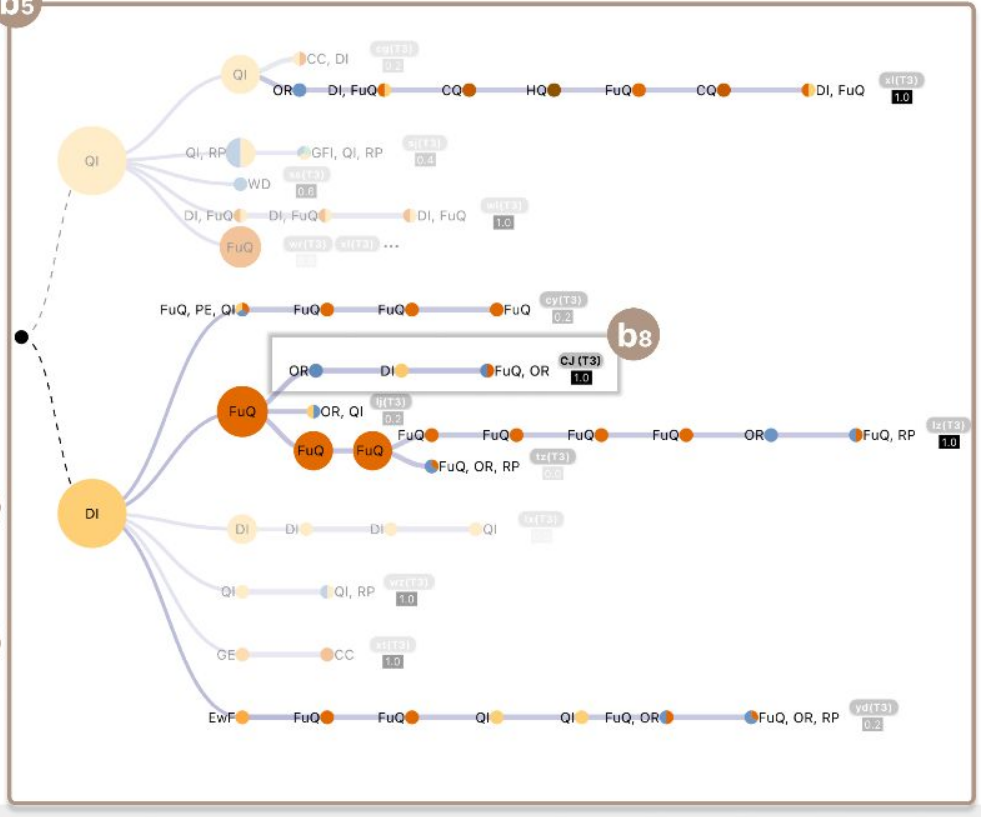

### b4 Pattern Table

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... |
| 2 | { Follow up Questions, Definition Inquiry } | 7 | 0.870... |
| 1 | { Output Restriction } | 7 | 0.828... |

First | Prev | Next | Last

### b5 Pattern Nuance




## Detail View (C)

### c1 Task Description

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

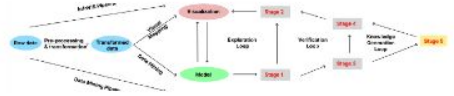
Stage 1 is [Stage_A]

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

### c2 Raw Conversation

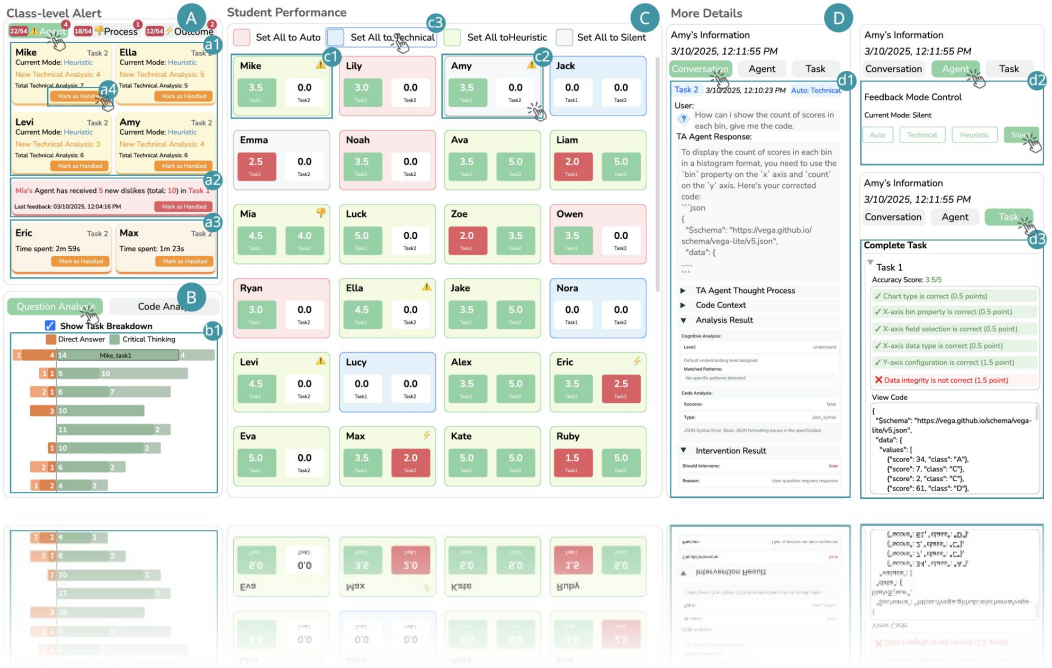
Do you know the diagram of updated visual analytics pipeline

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

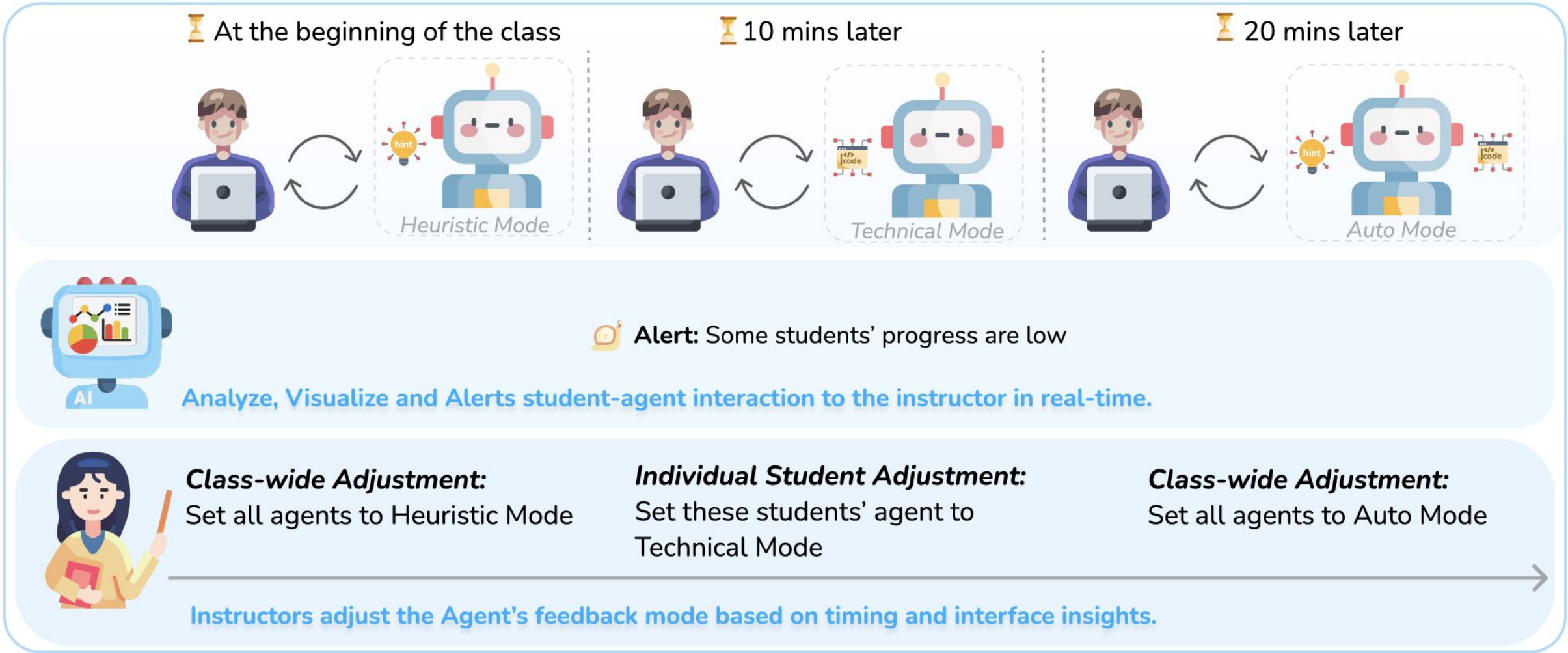
- Students' learning perspective:
  **More than 90%** students enjoy using ChatGPT in their learning process
- ChatGPT performance:
  **Strong positive correlation** between the IG (information gain) metric and experts' judgment of ChatGPT's response quality
- Expert interviews:
  ***"The ability to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

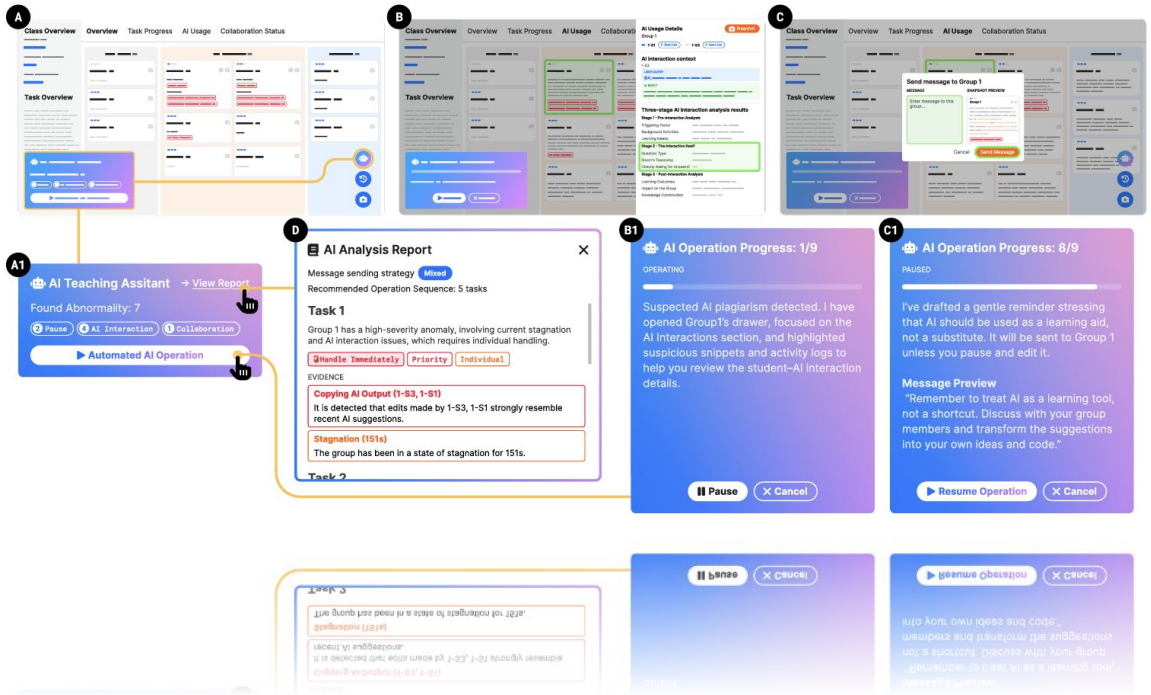
  *"The workflow's logical progression and the interconnection of each view were particularly impressive, enabling a diverse analytical focus through a unified procedure."*




# Real-time Classroom Orchestration for Students, AI, and Instructor

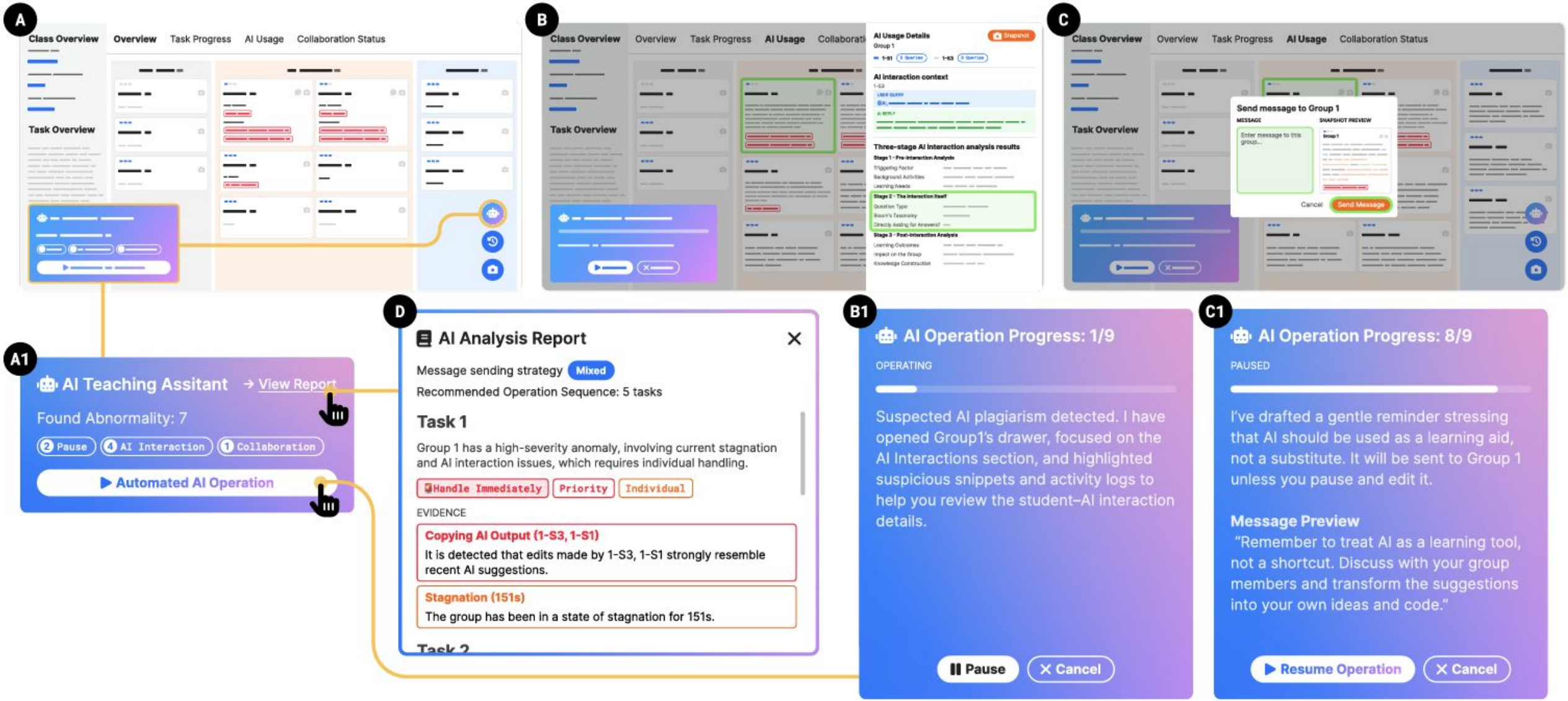
CHI 2026 (under review)


At the beginning of the class
10 mins later
20 mins later
hint
Heuristic Mode
</>
code
Technical Mode
hint
</>
code
Auto Mode
AI
Alert: Some students' progress are low
Analyze, Visualize and Alerts student-agent interaction to the instructor in real-time.
Class-wide Adjustment:
Set all agents to Heuristic Mode
Individual Student Adjustment:
Set these students' agent to Technical Mode
Class-wide Adjustment:
Set all agents to Auto Mode
Instructors adjust the Agent's feedback mode based on timing and interface insights.

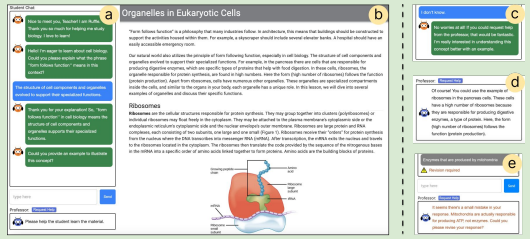



# Real-time Classroom Orchestration for Groups, AI, and Instructor

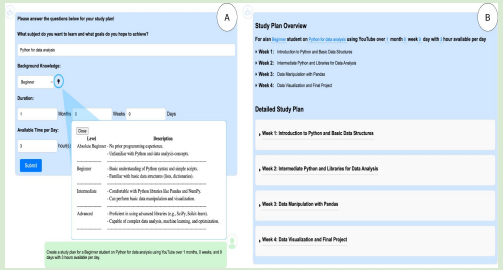
CHI 2026 (under review)


A
Class Overview
Overview
Task Progress
AI Usage
Collaboration Status
Task Overview
B
Class Overview
Overview
Task Progress
AI Usage
Collaborati
Task Overview
AI Usage Details
Snapshot
Group 1
AI interaction context
Three-stage AI interaction analysis results
C
Class Overview
Overview
Task Progress
AI Usage
Collaboration Status
Task Overview
Send message to Group 1
MESSAGE
SNAPSHOT PREVIEW
Enter message to this group...
Cancel
Send Message
A1
AI Teaching Assitant
→ View Report
Found Abnormality: 7
2 Pause
4 AI Interaction
1 Collaboration
Automated AI Operation
D
AI Analysis Report
Message sending strategy
Mixed
Recommended Operation Sequence: 5 tasks
Task 1
Group 1 has a high-severity anomaly, involving current stagnation and AI interaction issues, which requires individual handling.
Handle Immediately
Priority
Individual
EVIDENCE
Copying AI Output (1-S3, 1-S1)
It is detected that edits made by 1-S3, 1-S1 strongly resemble recent AI suggestions.
Stagnation (151s)
The group has been in a state of stagnation for 151s.
Task 2
B1
AI Operation Progress: 1/9
OPERATING
Suspected AI plagiarism detected. I have opened Group1's drawer, focused on the AI Interactions section, and highlighted suspicious snippets and activity logs to help you review the student–AI interaction details.
Pause
Cancel
C1
AI Operation Progress: 8/9
PAUSED
I've drafted a gentle reminder stressing that AI should be used as a learning aid, not a substitute. It will be sent to Group 1 unless you pause and edit it.
Message Preview
"Remember to treat AI as a learning tool, not a shortcut. Discuss with your group members and transform the suggestions into your own ideas and code."
Resume Operation
Cancel

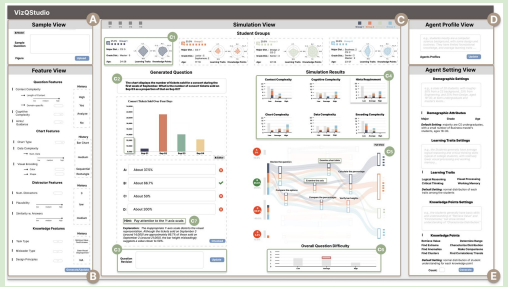
# Tutoring



Ruffle&Riley (AIED 2024)

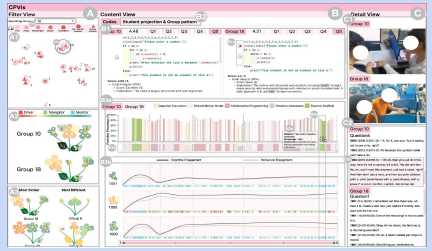


PlanGlow (L@S 2025)



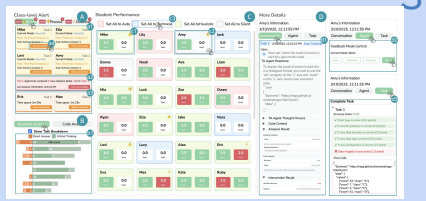
(under review)

# Evaluating

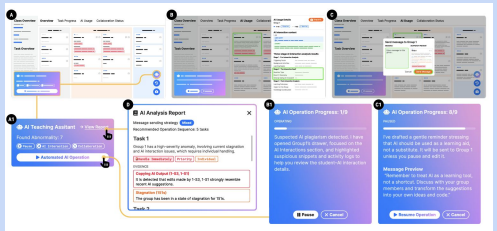


CPVis (CHI 2024)

StuGPTVis (TVCG 2024)

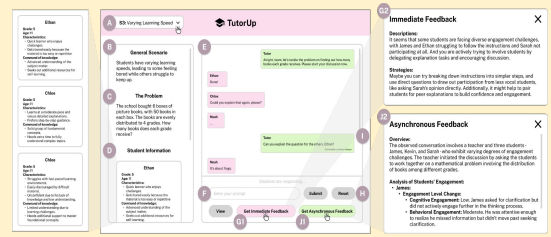


(under review)

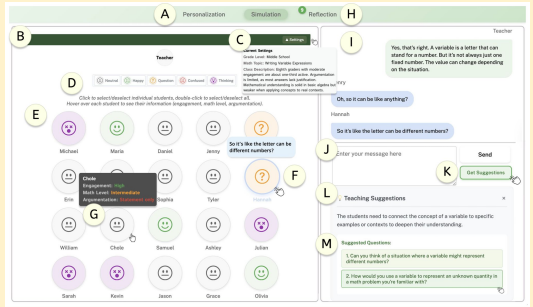


(under review)

# Practicing

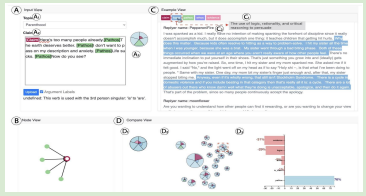


TutorUp (CHI 2025)

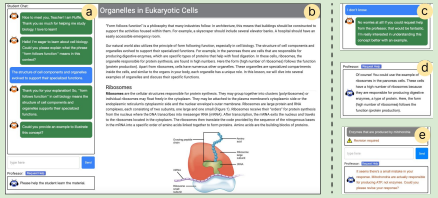


(under review)

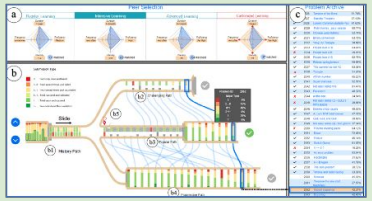
# Tutoring



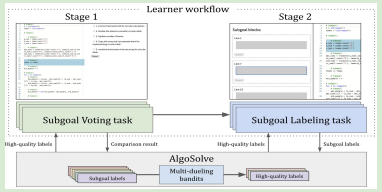
Persua (CSCW 2022)



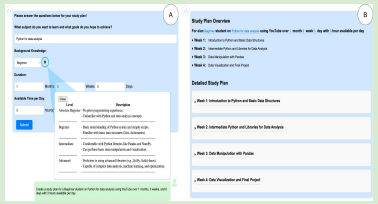
Ruffle&Riley (AIED 2024)



Peerlens (CHI 2019)

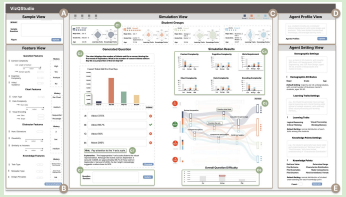



AlgoSolve (CHI 2022)



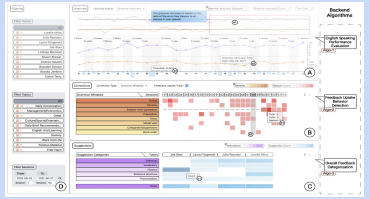
PlanGlow (L@S 2024)


“Gaming the system” (L@S 2020)

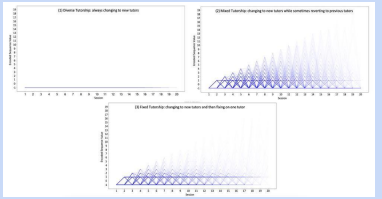


(under review)

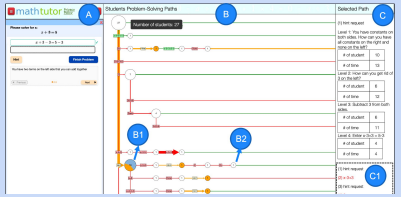
# Evaluating



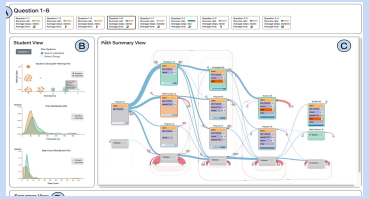
RLens (L@S 2022)



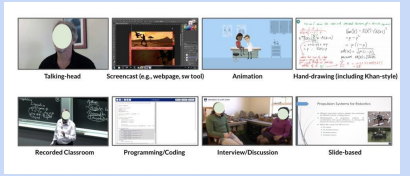
Distributed Tutorship (LAK 2022)



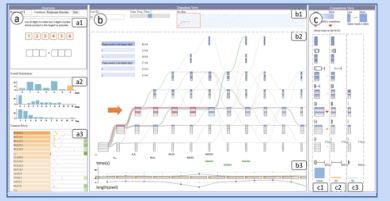
SolutionVis (AIED 2023)



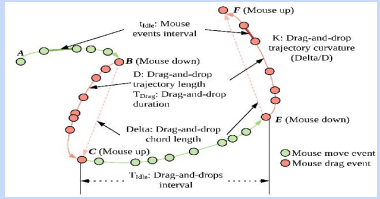
BlockLens (L@S 2022)



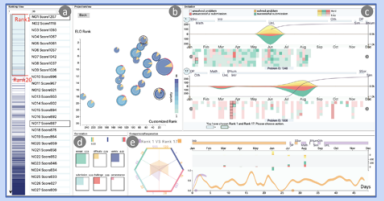

Mobile MOOCs (CHI 2022, **Best Paper Award**)



QLens (TVCG 2021)

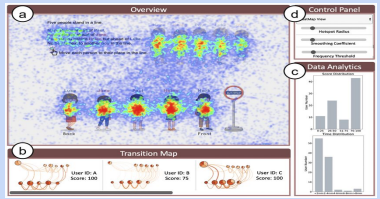



Predication (LAK 2020)



SeqDynamics (EuroVIS 2020)

StuGPTVis (TVCG 2024)



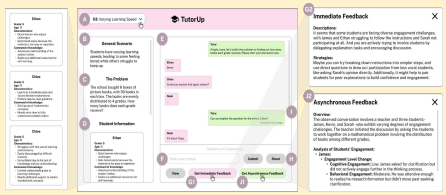

Visual Analytics K-12 (VIS 2019, **Best Poster Award**)

CPVis (CHI 2024)

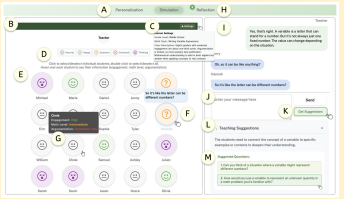
(under review)

(under review)

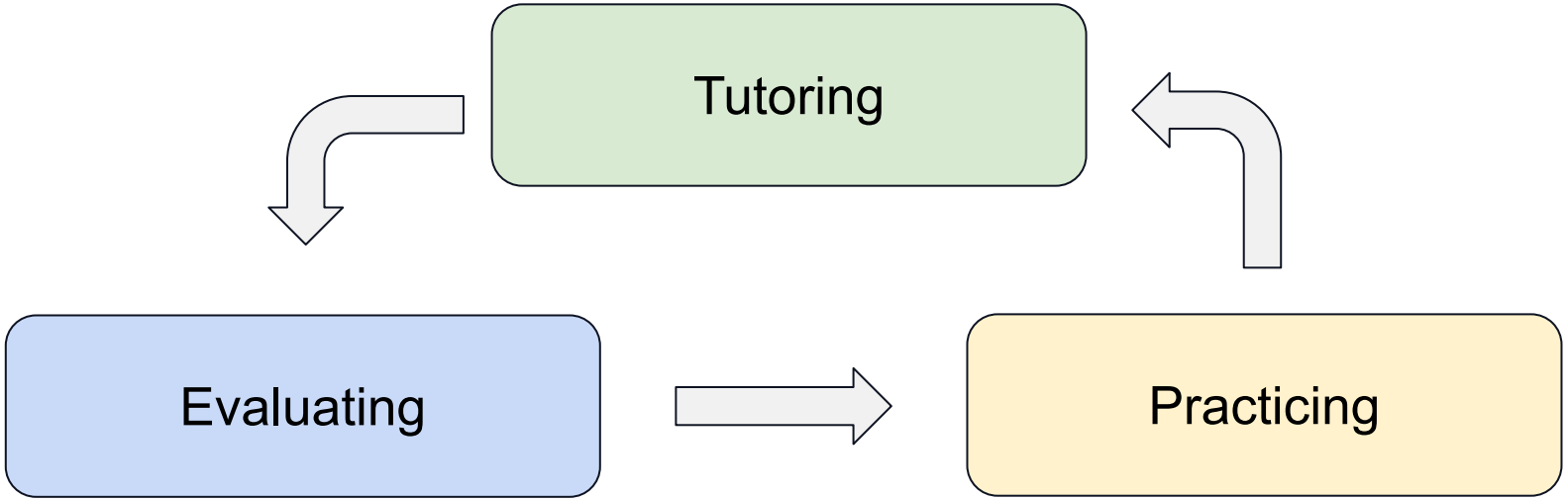
# Practicing




TutorUp (CHI 2025)



(under review)




Meng Xia
mengxia@tamu.edu
www.mengxia.org

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**!

DLab Dream Lab

TEXAS A&M UNIVERSITY