

THINK
CHANGE.
DO

Collaborative Learning Environments Supported by AI-Powered Feedback Systems

Nick Qi

Overview

- Introduction
- Method
- Experiment
- Results
- Conclusion and Future Work

INTRODUCTION

The emergence of large language models (LLMs) has opened new avenues for artificial intelligence (AI) based **learning progress diagnosis**.

Despite their potential, the application of LLMs in real-world diagnostic contexts remains limited due to their **challenges in proactively collecting learning interaction data**.

This study introduces a novel diagnostic system that leverages LLMs that **improves collaborative learning by simulating tutor-learner interaction**.

METHODS

A. Simulation Roles & Workflow

To simulate Online Tutoring for Coding, we conducted retrospective roles of a tutor agent and a learner agent.

Learner Agent: The learner agent is designed to simulate a typical student's submission, including step by step solution and final code

Tutor Agent: The tutor agent prompt the learner agent with questions and provides feedback that allows the learner agent to refine its answers.

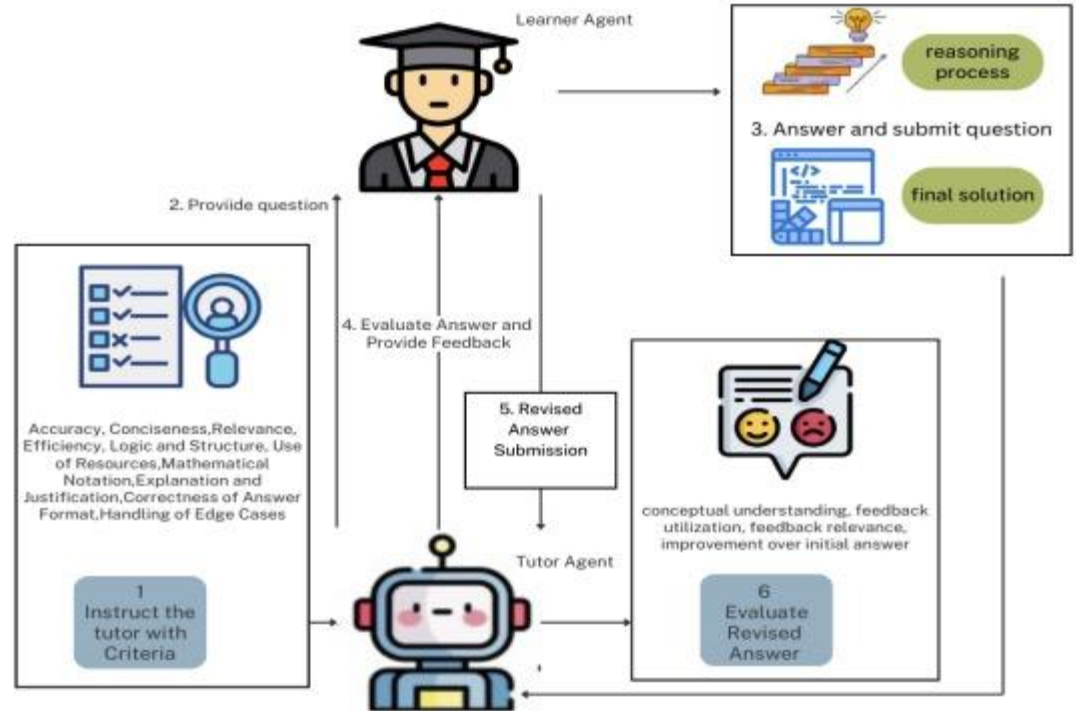


Fig. 1. Tutor Learner Interaction and Evaluation

METHODS

B. Simulation Clients & Libraries

1. In the **Task Construction**, **criteria** are used to guide the tutor agent in generating a task description for the learner agent, forming the basis for initiating a simulated conversation on Python tutoring.
2. The **LLM Client** is employed to invoke the language model and produce responses based on the constructed prompt, preparing the system for further LLM invocations as needed.

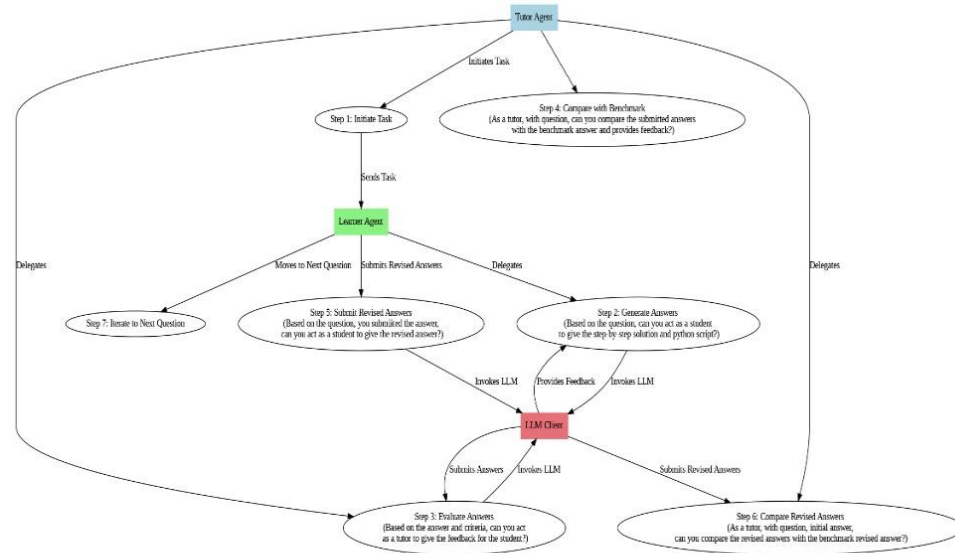


Fig. 2. Simulation Workflow

EXPERIMENT

C. Evaluation Metrics

Name	Description
Accuracy	The solution must be correct and adhere strictly to mathematical principles and techniques appropriate for the problem.
Conciseness	The explanation and method provided should be direct and to the point, avoiding unnecessary steps or complexity.
Efficiency	The solution should be derived in a time-effective manner, considering the complexity of the problem.
Logic and Structure	The reasoning should be logical and the information structured in a clear and understandable sequence.
Mathematical Notation	The use of proper and standard mathematical notation in the solution and explanation.
Explanation and Justification	There should be a clear explanation, rationale, or justification for each step taken towards the solution.
Handling of Edge Cases	The solution should correctly handle any special or edge cases that may arise in the problem.

METHODS

D. Differential Diagnosis

We are using **comparison prompt** to assess and compare the submission of students with the benchmark. It employs diagnostic techniques to **assess learners' performance and highlights the differences** in data analysis.

Code Analysis:

The system parses both the learner's code and the correct code to facilitate a detailed comparison, identifying key differences between the two versions, including **syntactic and semantic** discrepancies, such as incorrect operations, missing or extra steps, and logical errors

Step-by-Step Solution Analysis:

The evaluation process involves several key aspects:

1. It assesses whether the **logic** is correct and aligns with the intended solution.
2. It examines if the correct **operations** are performed at each step.
3. It also verifies that the **sequence** of operations follows the proper order of execution.
4. It determines whether the **code effectively** handles edge cases.

EXPERIMENT

A. Hypothesis

1. That answers generated by large language models (LLMs) will show **minimal differences** compared to those produced by human experts.
2. That LLM-generated **revised answers**, based on initial responses and feedback, will exhibit **limited differences** compared to human responses when evaluated using a set of revised answer performance metrics.

B. Experimental Settings

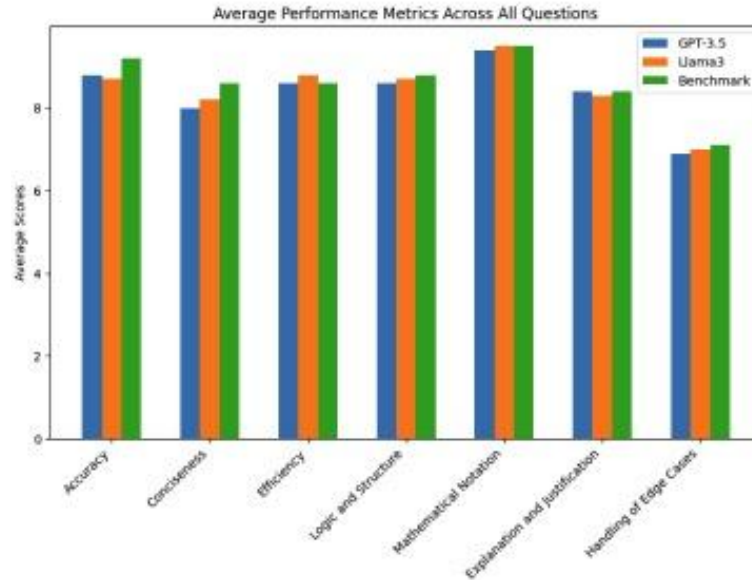
We used GPT-3.5 Turbo and Llama3 (8B-8192) as the foundation models for learner agents. These models are still highly proficient at **generating coherent and relevant responses**, making them well-suited for simulating the student role.

For the tutor agent, we opted to use GPT-4o due to its superior **language understanding, nuanced responses**, and ability to manage **long, context-rich conversations** that are crucial for a tutor role that often involves **explaining complex concepts** across multiple exchanges.

We chose to use the **Python_Code_Critic_21k** dataset to verify our hypothesis as it provides a comprehensive benchmark of annotated Python code submissions with expert-generated revisions and feedback.

RESULTS

A. Learner agent comparison and analysis



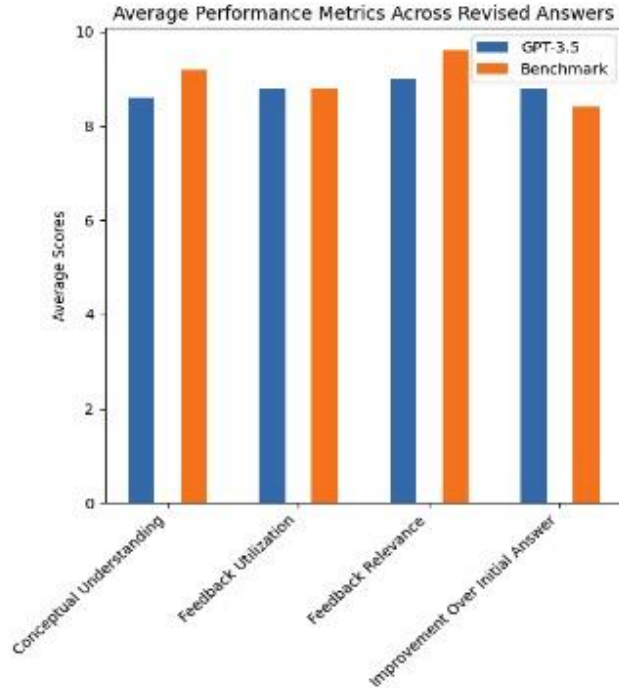
Human experts consistently score highest across all metrics, demonstrating their superior accuracy, conciseness, efficiency, logic and structure, mathematical notation, explanation, and edge case handling.

GPT3.5 has slightly higher accuracy and efficiency, while **Llama3 exhibits strong performance but with minor issues in mathematical notation and explanation.**

Human experts excel in conciseness and explanation, suggesting that AI models could benefit from further refinement in these areas.

RESULTS

B. Revised answer evaluation and analysis



The average scores in the figure show that **GPT-3.5** revised answers and **benchmark** revised answers perform **similarly across all metrics**.

In **feedback relevance**, GPT-3.5 performs better than Benchmark.

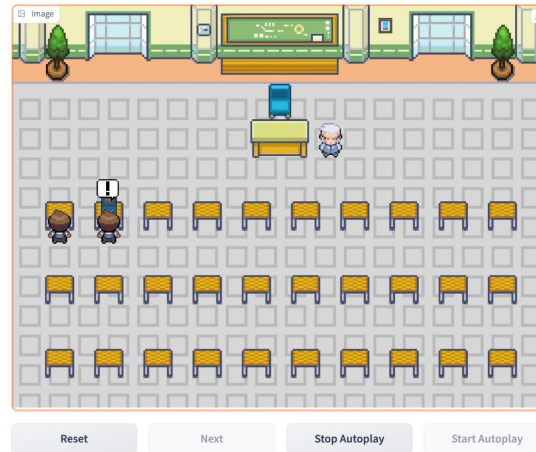
CONCLUSION AND FUTURE WORK

A. Conclusion

1. Explored the integration of AI-powered feedback systems on the use of LLMs to **enhance collaborative learning** which **boost the tutor-learner interactions**.
2. Developed a **simulation workflow** involving tutor and learner agents, where the tutor agent (supported by GPT-4) provided **iterative feedback** to learner agents (using GPT-3.5 and Llama3).
3. The experiment results indicated that there were **no statistically significant differences** between the performance of the LLM-generated initial/revised answers and the benchmark solutions.

B. Future Work

1. Incorporate **context-aware feedback** that adapts to the **learner's progress** over time to enhance the system's effectiveness
2. Expand the simulation to include **multiple learner agents** working collaboratively could uncover how AI-powered **peer-to-peer interactions** and **TA support** can improve learning outcomes



[Teaching Assistant Gamma]: "You're welcome, Beta. I'm glad you found the explanation helpful. Setting up a system of equations is a common approach to solving problems involving multiple variables. It allows us to represent the given information and find the values of the variables by solving the system. This problem was a good example of using deductive reasoning to solve a real-life problem. Deductive reasoning involves using logical principles and established facts to draw conclusions. In this case, we used the given

[Student Beta]: "Thank you, Professor Michael and Teaching Assistant Gamma, for explaining the solution to the problem. I understand now that we can set up a system of equations to solve for the number of student tickets and general admission tickets sold. By solving the system, we found that there were 268 student tickets and 257 general admission tickets sold. This was a great example of using deductive reasoning to solve a real-life problem."

[Professor Michael]: "To solve this problem, we can set up a system of equations. Let's call the number of student tickets 'S' and the number of general admission tickets 'G'. We know that the cost of a student ticket is \$4 and the cost of a general admission ticket is \$6. We also know that the total number of tickets sold is 525, so we can write the equation: $S + G = 525$. Additionally, we know that the total revenue collected is \$2876, which can be expressed as the equation: $4S + 6G = 2876$. Now we have a system of two equations with two variables. We can solve this system using substitution or elimination method to find the values of S and G, which will give us the number of student tickets and general admission tickets sold. By solving the system, we find that $S = 268$ and $G = 257$. Therefore, 268 student tickets and 257 general admission tickets were sold."

[Teaching Assistant Gamma]: "To solve this problem, we can set up a system of equations. Let's call the number of student tickets 'S' and the number of general



LinkedIn

