Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

LLM-based Collaborative Agents with Pedagogy-guided Interaction Modeling for Timely Instructive Feedback Generation in Task-oriented Group Discussions

Qihao Yang[1,2] , **Yu Yang**[2] , **Sixu An**[2] , **Tianyong Hao**[1†] and **Guandong Xu**[2†]
[1]School of Computer Science, South China Normal University, Guangzhou, China
[2] Centre for Learning, Teaching, and Technology, The Education University of Hong Kong, Hong Kong
yeung.qh@gmail.com, {yangyy, s1154340, gdxu}@eduhk.hk, haoty@m.scnu.edu.cn

Abstract

Large language models (LLMs) fundamentally reshape learning and teaching models, shifting tutoring systems from supporting individual learning to facilitating collaborative learning (CL) like task-oriented group discussions. However, existing AI tutors struggle to guide CL, as they seldom model the interactions between AI tutors and students. Therefore, they cannot scaffold students to complete tasks collaboratively, which impairs learning outcomes and pedagogy adaptability. Additionally, existing AI tutors fail to make use of CL theories to generate instructive feedback, which leads to undesirable interactions such as over-instruction and limits students' autonomy. In this paper, we propose an LLM-based collaborative agent that innovatively leverages pedagogical strategies to sense discussion stages, detect learning issues, identify the timing of intervention, and generate instructive feedback. To develop the agent, we first design a prompting strategy based on a CL theory, that is, the Community of Inquiry, to cultivate the agent to understand the discussion status. Second, a multi-agent interaction framework is proposed to simulate the collaborative learning behavior between AI tutors and students. Meanwhile, a synthetic task-oriented group discussion dataset, namely CLTeach, is generated, which consists of 27k manually-verified multi-party dialogues with fine-grained annotations of instructive feedback and explanations. Lastly, we use CLTeach to fine-tune the LLM agent, ultimately enabling it to generate instructive feedback at the right time to support students in CL. Extensive experiments demonstrate that our agent achieves state-of-the-art performance in feedback generation and has the potential to mimic human teachers effectively.

1 Introduction

Driven by large language models (LLMs), LLM-based tutoring systems designed for individual learning make significant advancements, promoting in-depth research [Liu *et al.*, 2024a] and practical applications [Dan *et al.*, 2023] in AI-empowered education. With the growing importance of 21st-century skills including teamwork and critical thinking, developing advanced AI tutors to facilitate collaborative learning (CL), such as task-oriented group discussion (ToGD), has received increasing attention [Wang *et al.*, 2024; Gan *et al.*, 2023].

Collaborative learning is an effective pedagogical approach for facilitating domain-specific knowledge construction and enhancing students' problem-solving skills through group interaction and communication [Laal and Laal, 2012]. Online ToGD is a typical means of CL, which has been widely adopted in frontline teaching. Students can fully lead the discussion. However, it is common for the discussion to go off-topic, and some students may not engage actively, eventually impairing the learning outcomes in CL. To improve the quality of CL, frontline teachers were sent to moderate discussions, but this also brought extra workload to the already overburdened teachers, the quality of the moderation largely depended on the teachers' abilities. Developing an AI tutor to assist frontline teachers in moderating ToGD remains an open problem.

Naively prompting the general-purpose LLMs for ToGD fails to generate instructive feedback for students at the right time because it does not have the guidance of CL theories or knowledge. One problem is that the LLMs will provide answers to students directly instead of guiding students to solve the problem collaboratively. Students will conform to the dominant ideas presented by the AI tutor, stifling critical thinking and independent analysis. The other problem is that the LLMs tend to reply to every student's post, which disrupts collaboration and makes CL become invalid. Providing instructive feedback at the right time without disrupting collaboration among students is the successful key of the AI tutor for CL [Holstein *et al.*, 2019].

Fine-tuning the LLMs using student-generated discussion data is not sufficient to achieve the above goal. The absence of teacher-student interactions hinders LLM from generating higher-quality instructive feedback through learning the impact of teacher feedback on the following discussion among students. Teacher-student interactions also tell the timing of intervention for LLM to learn when should give the instructive feedback. Although there exists several teacher-student

†Corresponding authors.

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

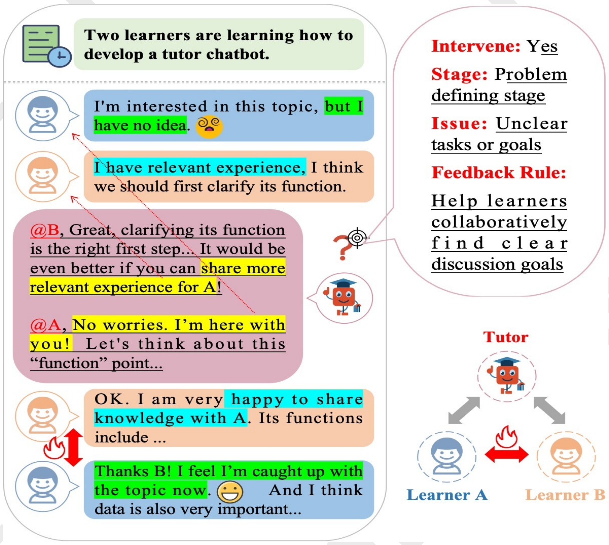

Figure 1: Illustration of our collaborative agent.

dialogue datasets [Macina *et al.*, 2023; Chen *et al.*, 2024; Shani *et al.*, 2024], unfortunately, they are one-on-one conversations and have nothing to do with collaboration. Conducting real-world experiments to collect large-scale teacher-facilitated discussion data is time-consuming and with extremely high cost.

In this paper, we study an important yet overlooked problem, which is timely instructive feedback generation in ToGD. To solve this problem, we confront two major technical challenges. The first is to discover and learn the unknown impact of AI tutors' feedback on the following discussion among students. The second challenge is to ensure the pedagogy adaptability of the generated instructive feedback. This requires the model to be able to understand the overall discussion situation like a real teacher, promptly identify the problems encountered by each student in the discussion, and generate guiding feedback accordingly.

To overcome these challenges, we propose a novel LLM-based collaborative agent, incorporating pedagogical strategies to sense discussion stages, detect learning issues, identify the timing of intervention, and generate instructive feedback. To ensure the pedagogy adaptability, we leverage a CL theory, the Community of Inquiry (CoI) [Garrison and Arbaugh, 2007], to supervise the entire design of our LLM-based collaborative agent. CoI is the primary framework for designing collaborative learning environments, which is crucial for establishing a sense of community and interpersonal connections among students [Garrison, 2022]. It encourages group members to reflect on their understanding and progress regularly, identify areas for improvement or adjust their learning strategies. According to the CoI, the existing study [Akyol and Garrison, 2011] gives specific discussion stages, types of issues students may encounter in the discussion, and the corresponding feedback rules. This inspires us to design a prompting strategy based on the CoI and the work [Akyol and Garrison, 2011] to cultivate the agent to understand the overall discussion situation and promptly identify the problems encountered by each student. We fine-tune the prompted LLM using a labeled student group discussion dataset, further improving its performance in identifying five discussion stages and forty learning problems and making intervention decisions. An example is shown in Figure 1.

To discover and learn the unknown impact of AI tutors' feedback on the following discussion among students, we devise a multi-agent framework to simulate the collaborative learning behavior between a tutor agent and multiple student agents by automatically performing the ToGD. It simulates how a teacher provides feedback to students and how students react collaboratively after receiving the instructive feedback. Meanwhile, a synthetic ToGD dataset, namely CLTeach, is generated, which consists of 27k manually-verified multi-party dialogues with fine-grained annotations of instructive feedback and explanations. These synthetic data enable us to further fine-tune the LLM agent, learning the impact of AI tutors on students' collaborative behavior to generate instructive feedback at the right time without disrupting collaboration among students.

Our main contributions are summarized as follows:

- A new holistic approach of building a collaborative LLM agent as an AI tutor to effectively moderate ToGD like a real teacher[1]. It contains a novel CoI-based prompting strategy with fine-tuning to ensure the pedagogy adaptability of the generated instructive feedback. A multi-agent framework is designed to discover and learn the impact of AI tutors on students' collaborative behavior to generate instructive feedback.
- A new ToGD dataset consists of 27k manually verified multi-party LLM-involved dialogues with fine-grained annotations, which can drive future research on studying human-LLM agent interactions.
- Extensive experiments demonstrate that our LLM-based collaborative agents achieve SOTA performance in providing students with timely instructive feedback without disrupting the collaboration process.

The literature review will be presented in the next section. Section 3 will introduce the technical details of our LLM-based collaborative agent, followed by the experimental results in Section 4 before concluding.

2 Related Work

2.1 LLM-based Tutoring Systems

LLM-based tutoring systems have gained attention in recent years due to their ability to provide personalized and scalable educational support [Liu *et al.*, 2024a; Gan *et al.*, 2023]. To activate LLMs' potential for diverse educational contexts, some studies explore the effectiveness of general-purpose LLMs like ChatGPT in facilitating knowledge acquisition, especially in STEM education [Ding *et al.*, 2024], language learning [Liu *et al.*, 2024b], and writing support [Han *et al.*, 2023]. A MWPTutor is proposed by [Pal Chowdhury *et al.*, 2024] to guide students for math word problems, which demonstrates that GPT4 [Achiam *et al.*, 2023]

[1]Source codes and datasets are available via https://github.com/CharlesYang030/Collaborative-Agents.

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

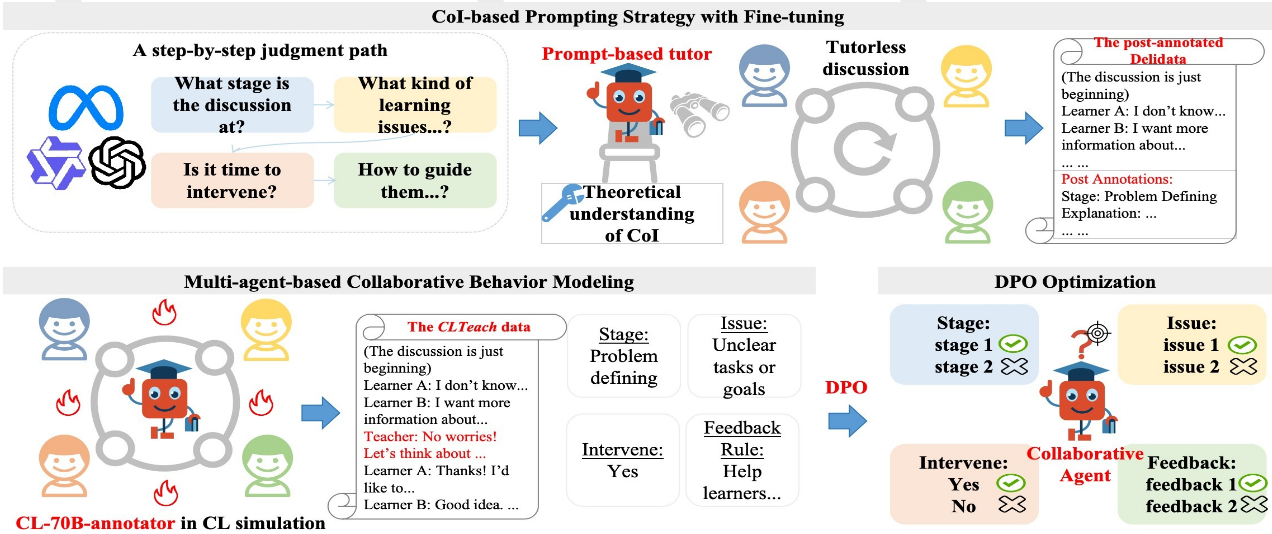

Figure 2: Overview framework of the proposed collaborative agent.

has good teaching quality on two datasets with mathematical problems and solutions. A scaffolding dataset and a CLASS framework are proposed by [Sonkar *et al.*, 2023] to enable LLMs to guide students in a step-by-step way. This highlights the LLMs' abilities to break down biology questions into manageable subproblems and provide encouraging responses to students. Although these studies shows the promising prospects to enhance LLMs with pedagogical strategies, they focus mainly on one-on-one teaching. The investigation of LLM-based tutoring systems within collaborative learning (CL) remains underexplored [Gan *et al.*, 2023; Cai *et al.*, 2024]. However, it is necessary to have a skilled facilitator to guide the discussions, ask questions that spark new ideas, cultivate an environment that values diverse perspectives, establish the tone of discussions, and redirect the conversation when it strays off course [Cai *et al.*, 2024]. Existing systems for CL rely on predefined scripts to initiate / conclude topics or pose simple follow-up questions. This often leads to undesirable interactions such as the reduction of student motivation due to rigid guidance. This does not meet the expectations that students can complete common goals collaboratively and efficiently in an engaged environment. In addition, these systems do not emphasize the collaborative performance between students and AI tutors, resulting in their inability to model multi-party collaboration well.

2.2 Teaching Datasets

Multi-party collaborative dialogues are the key factor for building a CL agent. Unfortunately, existing multi-party dialogues, such as AMI [Carletta *et al.*, 2005] and ICSI [Janin *et al.*, 2003], are mainly designed for meeting summarization and they lack role annotations. Thus, they are not suitable for training a CL agent. Constructing collaborative dialogues requires the interactions between a tutor and diverse students, while collecting data from the real world is costly. The work [Nguyen *et al.*, 2024] prompts GPT4 to role-play as different personas with various demographics and levels of concern about climate change. They highlight the feasibility of simulating tutorless discussions about social issues with multiple LLM agents. The work [Liu *et al.*, 2024a] utilizes dean-teacher-student agents to generate data but the part of multiparty dialogue only involves isolate interactions between the teacher with each student. This cannot present the collaborations among students under the guidance from teachers. However, this paves the way to leverage a multi-agent framework to model collaborative behaviors and generate synthetic dialogue data.

2.3 Community of Inquiry

The Community of Inquiry (CoI) theory, widely applied in online and blended learning, emphasizes cognitive, social, and teaching presences to foster collaborative learning [Shea and Bidjerano, 2010]. The research [Shea *et al.*, 2012] has explored CoI's role in discussion stages, highlighting its effectiveness in structuring collaborative dialogue from exploration to resolution [Akyol and Garrison, 2011]. Some studies show its potential in addressing learning issues, such as unclear goals and someone's lack of involvement, by enabling critical discourse and reflective thinking [Shea and Bidjerano, 2009]. Moreover, CoI promotes shared metacognition [Garrison, 2022], as students collaboratively regulate their learning processes, monitor progress, and resolve conflicts. Although no studies have yet integrated CoI with AI tools to support ToGD, this provides theoretical support for our work.

3 Our LLM-based Collaborative Agent

Inspired by CoI, we propose an LLM-based collaborative agent to effectively moderate ToGD like a real teacher, as shown in Figure 2. A CoI-based prompting strategy with fine-tuning is used to ensure the pedagogy adaptability for the ToGD status recognition. With the reliable instruction-following ability, the agent is then integrated into the multi-agent framework with diverse Student agents, to generate synthetic data. The agent is further fine-tuned to generate

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

instructive feedback at the right time without disrupting collaboration among students.

3.1 CoI-based Prompting Strategy with Fine-tuning for ToGD Status Recognition

Thanks to the expertise provided by educational scientists from the Education University of Hong Kong, we first designed a prompting strategy that unified the standards of stages, issues, intervention, and feedback rules in ToGD. Specifically, we define five progressive stages including problem defining, exploration, integration, resolution, and reflection. For each stage, there are eight common learning issues. For instance, at the problem defining stage, the issues include unclear tasks or goal, lack of motivation to participate and so on. At resolution stage, the issues contain lack of practical solution, difficulty in reaching consensus and so on. Then, the conditions for intervention are defined. For example, intervention is needed when students engage in superficial discussions or non-exploratory thinking without clear direction. When a discussion is just beginning and moving forward, there is no need to intervene. If a learning issue arises during the discussion and intervention is needed, there is a targeted feedback rule for this case. For example, the feedback rule for the unclear tasks or goal issue is to guide students to clarify the task goals and ensure a shared understanding. Each feedback rule also has several example sentences. All stages, issues, intervention conditions and feedback rules have clear definitions.

To alleviate the prompting burden for LLM, we propose a step-by-step judgment path. Specifically, we treat task-oriented group discussions as multi-party dialogues, where multiple students $L = \{l_1, l_2, ..., l_n\}(n >= 2)$ and a single agent A interact with each other under a given topic Θ. In limited turns $T = \{t_1, t_2, ..., t_n\}$, the dialogue consists of the sequence of utterances $U = \{u_{t_1}, u_{t_2}, ..., u_{t_n}\}$ and each utterance is represented as a tuple $u_t = (c_t, s_t, t)$, where c_t is the utterance content, s_t is one of the speakers $S = L \cup A$, and t is the turn. When a student speaks at a turn t, a dialogue history $H_t = \{u_1, u_2, ..., u_{u-1}, u_{t-2}\}$ contains utterances up to and including the turn t. Under the step-by-step judgment path, given a dialogue history H_t and a topic Θ, the agent A is required to successively output stage results R_t^{stage}, issue results R_t^{issue}, intervention results $R_t^{intervention}$, and instructive feedback $R_t^{feedback}$ by following Eq.1 to Eq. 4:

$$f_{stage}\left(H_t, \Theta\right) \rightarrow R_t^{stage} \tag{1}$$

$$f_{issue}\left(H_t, \Theta, R_t^{stage}\right) \rightarrow R_t^{issue} \tag{2}$$

$$f_{intervention}\left(H_t, \Theta, R_t^{stage}, R_t^{issue}\right) \rightarrow R_t^{intervention} \tag{3}$$

$$f_{feedback}\left(H_t, \Theta, R_t^{stage}, R_t^{issue}, R_t^{intervention}\right) \rightarrow R_t^{feedback} \tag{4}$$

where f denotes generative functions.

To validate the effectiveness of the proposed prompting strategy, we use a Llama3.1-70B-Instruct model to post-annotate Delidata [Karadzhov *et al.*, 2023]. Llama3.1-70B-Instruct is selected because it is an open-source model with impressive performance in various instruction tasks, which is good for reproducibility. Delidata is constructed under a psychology game environment where participants are required to complete the game collaboratively. It records 500 group discussions including 14,000 utterances with role labels (without tutor). Delidata is chosen since it presents the collaborations in a tutorless discussion, compared to other datasets. After that, we recruit four student helpers from The Education University of Hong Kong, whose background is Education Science, to manually validate the LLM-generated annotations. An evaluation of the post-annotated Delidata is conducted using Fleiss' Kappa, yielding an agreement score of 0.81 and a valid data ratio of 87.24%. This ensures the effectiveness of our prompting strategy and the reliability of these annotations. The data is further refined manually and its statistics are displayed in Table 1. The proportions of the exploration stage and the integration stage are 40.04% and 31.33% respectively. This reflects that students are indeed more likely to encounter more issues in the early stages of the discussion in a real-world CL environment. This also shows the importance of having a tutor agent to facilitate CL in a tutorless discussion.

Unfortunately, we identify 5.7% of cases involving garbled text and 7.06% involving failure to follow instructions. These errors are caused by the LLMs' inherent high sensitivity to prompts before they are fine-tuned. On one hand, this will impair the interaction ability of the agent, resulting in its poor performance in real teaching scenarios for CL. On the other hand, this will also hinder the agent from understanding CL theories to identify discussion stages and learning issues and make intervention decisions. Therefore, we fine-tune the Llama3.1-70B-Instruct model using the instruction tuning method [Wei *et al.*, 2021] by not masking the prompt, to further cultivate the agent to understand the overall discussion situation. The fine-tuned Llama3.1-70B-Instruct model is named CL-70B-vanilla, which have further theoretical understanding of CL and improved instruction-following ability for the ToGD task. It is deployed in the interaction framework as a vanilla tutor.

3.2 Multi-agent-based Collaborative Behavior Modeling and Data Generation

Previous studies [Nguyen *et al.*, 2024; Jin *et al.*, 2024; Xu and Zhang, 2023] demonstrate the potential of LLMs in simulating virtual students by defining the student profiles such as education, major, interests, speaking style and tone. Inspired by this, we propose a multi-agent interaction framework in which the CL-70B-valilla model interact with multiple student agents. Firstly, we select a set of topics from Debatepedia [Gottipati *et al.*, 2013] as the initial topics. Debatepedia is a collaborative online encyclopedia documenting arguments for and against various debates on a wide range of deep topics suitable for in-depth discussions. Instead, existing public question-answering datasets [Rajpurkar *et al.*, 2016; Joshi *et al.*, 2017] mainly cover simple questions, such as "What is the molecular formula of water?". Each virtual discussion room has a topic, where at least two and at most five student agents are instantiated with different profiles and prior knowledge about the topic. To reflect how students react collaboratively after re-

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

Item	The CLTeach dataset	
	The post-annotated Delidata	The interaction dataset
Discussion	454	319
Utterance	13,361	13,849
Average tokens of feedback	69.98	85.48
Problem defining stage	9.90% $\sigma = 1.77$	3.36% $\sigma = 0.31$
Exploration stage	40.04% $\sigma = 3.37$	10.83% $\sigma = 1.36$
Integration stage	31.33% $\sigma = 3.26$	25.71% $\sigma = 1.59$
Resolution stage	12.51% $\sigma = 2.37$	30.06% $\sigma = 1.84$
Reflection stage	6.22% $\sigma = 2.54$	30.04% $\sigma = 2.47$
Agreement score	0.81	0.83

Table 1: Statistics of the CLTeach dataset including the post-annotated Delidata and the interaction dataset. σ refers to the standard deviation.

ceiving the instructive feedback, student agents are required to monitor their own shared metacognition [Garrison, 2022]. Shared metacognition refers to the awareness and regulation for individual learning and group learning among participants in a community of inquiry [Akyol and Garrison, 2011; Garrison, 2022]. It reflects the change of social presence, teaching presence and cognitive presence of participants in CL. Following [Garrison, 2022], we respectively use 13 levels of self-regulation for individual learning and 13 levels of co-regulation for group learning, to prompt student agents to adjust their responses based on their levels of shared metacognition. Examples of the self-regulation for individual learning and the co-regulation for group learning are shown as follow:

When I am engaged in the learning process as an individual: SELF-REGULATION
I1: I am aware of my effort.
I2: I am aware of my thinking.
... ...
I12: I assess how I approach the problem.
I13: I assess my strategies.

When I am engaged in the learning process as a member of a group: CO-REGULATION
G1: I pay attention to the ideas of others.
G2: I listen to the comments of others.
... ...
G12: I help the learning of others.
G13: I monitor the learning of others.

This allows more authentic collaborative behavior modeling than forcing collaborators follow the one-input-one-output paradigm. The maximum number of discussion turns is set as 50. The discussion will end if it reach 50 turns.

After multi-agent simulation, an interaction dataset across a wide range of topics is generated. The dataset is refined and verified by the student helpers, yielding an agreement score of 0.83 and a valid data ratio of 91.14%. Its statistics is presented in Table 1. Compared to the post-annotated Delidata, the proportions of the integration stage, the resolution stage, and the reflection stage are 25.71%, 30.06 and 30.04% respectively. This suggests that students generally need guidance in the later stages of the discussion in the simulated CL environment. With the involvement of a tutor, this also reflects that students are more likely to get through the early stages that often involve a lot of information exploration and integration.

3.3 Instructive Feedback Generation

The post-annotated Delidata reflects the changes of discussion stages and learning issues encountered by collaborators in a real ToGD. The interaction dataset shows the nuances of teacher-student collaborative behaviors in a simulated ToGD. They both involve the learnable information for the agent to generate timely instructive feedback without disrupting collaboration among students. Therefore, they are combined into a dataset named CLTeach to optimize our final agent jointly.

However, we realize that human teachers often make a teaching decision depending on preferences in different situations. The teaching decision is not either-or [Sonkar *et al.*, 2024]. For example, when students exchange their ideas, this discussion is sometimes identified as the exploration stage and sometimes as the integration stage. when students are encountering conflicting viewpoints, human teachers perhaps provide a leading guidance to all students to alleviate the common conflict, or they perhaps encourage a specific student to lead the discussion and allow them try to resolve the issue on their own. Such decision margin can be changed by the teacher's preference. If the agent can simulate the learning from preferences, it will have more flexible decision-making ability to generate high-quality instructive feedback.

The supervised fine-tuning (SFT) approach is widely used to optimize generative language models by cross-entropy loss functions to minimize the difference between the labeled data distribution and the model generated distribution. The training data for SFT contains only one targeted answer (decision). In other words, given a training data (x, y) where x denotes the inquiry and y denotes the answer, language models are optimized in a mapping mode of only one absolutely correct answer y. Direct preference optimization (DPO) [Rafailov *et al.*, 2024] is a recently proposed efficient and simple fine-tuning method. Given training data with a pair of answers (x, y_w, y_l) where y_w indicates the preferred answer and y_l indicates the non-preferred answer, language models are optimized to maximize the probability of generating the preferred answer y_w by following:

$$\mathcal{L}_{DPO} = -\log \sigma \left(\beta \left(f_\theta \left(x, y_w \right) - f_\theta \left(x, y_l \right) \right) \right)$$

where f_θ denotes a scoring function for a given input x and output y, β denotes a temperature parameter for controlling the intensity of the preference difference, and σ denotes a sigmoid function used to represent the probability that y_w is preferred over y_l.

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

Model	The post-annotated Delidata						The interaction dataset					
	Intervention	Stage	Issue	Feedback			Intervention	Stage	Issue	Feedback		
	F1-score			BLEU	Rouge-L	BERTscore	F1-score			BLEU	Rouge-L	BERTscore
Mistral-7B-Instruct	44.61	10.24	4.79	2.55	6.0	2.59	38.24	7.99	0.0	4.19	2.85	0.80
Qwen2.5-7B-Instruct	42.97	8.73	4.20	2.78	6.21	3.31	40.55	7.11	1.43	5.73	4.22	1.69
Llama3.1-8B-Instruct	45.55	14.15	5.29	3.24	11.52	5.43	43.93	12.12	2.92	7.23	6.32	3.01
GPT-4o-mini	44.34	14.20	6.46	2.98	10.18	5.56	42.47	14.33	3.57	8.14	6.53	4.67
GPT-4o	50.14	27.98	11.95	4.04	14.82	8.78	49.80	21.24	6.23	11.65	10.98	6.77
Our agent	**68.99**	**56.67**	**50.12**	**12.34**	**30.10**	**33.22**	**66.71**	**58.14**	**44.59**	**25.22**	**37.16**	**41.74**
w/o post-annotated Delidata	56.14	42.68	33.15	7.29	22.32	26.68	56.04	44.32	32.57	15.15	22.35	30.97
w/o interaction data	52.81	39.00	24.10	8.68	23.15	26.77	49.78	40.59	28.98	13.40	20.71	24.56
w/o decision making	64.70	50.07	46.88	10.46	26.51	32.88	63.93	53.51	37.39	20.15	28.55	33.67

Table 2: Teaching performance. Bold numbers indicate the SOTA results. All values are in percentage.

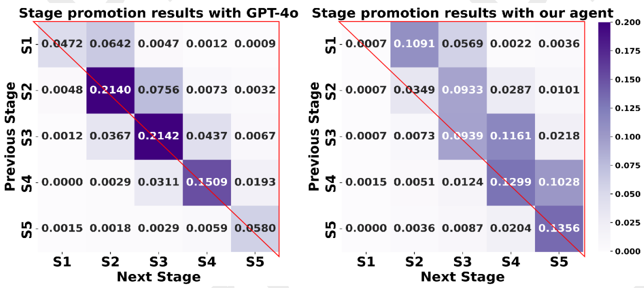

Figure 3: Stage promotion results with GPT-4o and our agent.

We believe that human teachers' decision-making process is closer to DPO than SFT. Therefore, we reform the CLTeach dataset into preference-style data required by DPO. The standard decision as y_w from CLTeach is paired with another random but not repeated decision as y_l. This is a straightforward method, and it indeed saves the cost of humans labeling preferences and scales up the learnable information. The agent can gain more learnable knowledge from the preference-style data. Considering the inference speed, a Llama3.1-8B-Instruct model is fine-tuned on the preference-style data, which is relatively efficient and meets the needs for the interaction with users in real-world ToGD.

4 Experiments and Results

4.1 Implementation details

Baseline. We compare several mainstream LLMs with our collaborative agent for this task, including close-source LLMs (GPT-4o and GPT-4o mini) and open-source lightweight LLMs (Llama3.1-8B-Instruct, Mistral-7B-Instruct [Jiang *et al.*, 2023], and Qwen2.5-7B-Instruct[Bai *et al.*, 2023]).
Dataset. We split 80% of the CLTeach dataset as the training set and 20% as the test set respectively. All data is in English.
Setting. The experiments are implemented Pytorch 2.4.0 and 4 A800 (80G) GPUs. We employ LoRA [Hu *et al.*, 2021] to fine-tune models, where the learning rate is 5e-5, the epoch is 3 and the compute type is bf16. For the close-source LLMs (GPT-4o and GPT-4o mini), we test them through OpenAI APIs. For the balance between inference speed and the dialogue history, the context length is set as 4096 tokens. A classification metric (F1-score) is used to assess agents' performance on intervention decisions, stage identification, and issue detection. The generative metrics (BLEU [Papineni *et al.*, 2002], Rouge-L [Lin, 2004], and BERTscore [Cai *et al.*, 2024]) are used to assess the quality of agents' feedback content.

4.2 Main Results

The comparison results between our collaborative agent and the baseline models on the test sets are shown in Table 2. Our agent achieves the state-of-the-art teaching performance, especially in feedback generation. Compared to GPT-4o, the agent attains 18.85% improvement for intervention, 28.69% for stage, 38.17% for issue, and 24.44% BERTscore improvement in the post-annotated Delidata. Similarly, it attains 16.91% for intervention, 36.9% for stage, 38.36% for issue, and 34.97% BERTscore improvement in the interaction dataset. These results shows that our collaborative agent has impressive pedagogical adaptability for facilitating CL. It outperforms GPT-4o and it is far superior to those general-purpose LLMs, demonstrating the effectiveness of the pedagogy-guided interaction modeling. The Llama3.1-8B-Instruct model is the backbone of our collaborative agent. Compared to it, our agent has improved in all metrics, proving the advantages of the manually-verified LLM-generated annotations to enhance the agent to explore and learn the collaboration information among AI tutors and students. In addition, we find that the these baseline models performs worse on the interaction data than on the post-annotated Delidata. This shows the difficulty of guiding CL in an interaction environment.

4.3 Ablation Study

An ablation study is conducted to reveal the contribution of each module of our approach, and the results are also reported in Table 2. When the agent is fine-tuned on only the post-annotated Delidata (*w/o* interaction data), the teaching performance of the agent drops 16.18% for intervention, 17.67% for stage, 26.02% for issue, and 6.45% for BERTscore in the post-annotated Delidata, 16.93% for intervention, 17.55% for stage, 15.61 for issue, and 17.18% for BERTscore in the interaction dataset. These results are reduced more than the situation where the agent is fine-tuned on only the interaction dataset (*w/o* post-annotated Delidata). This means that collaborative behavior modeling is the most crucial component for developing a CL agent, since the generated data enables the contextual pedagogy adaptability of the agent to imitate more appropriate teaching behaviors in a CL environment.

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

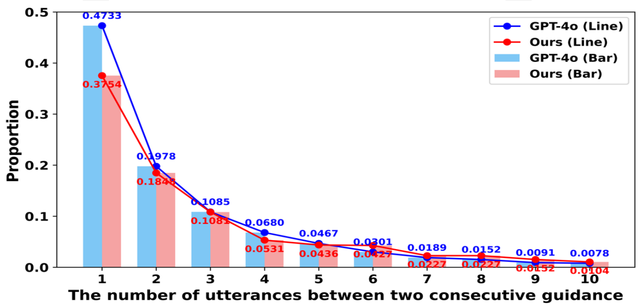

Figure 4: The distribution of the number of utterances between each two consecutive guidance with GPT-4o and our agent.

When the agent is fine-tuned by SFT (*w/o decision making*), its performance is not degraded as severely as in the above two cases, but it is still degraded. This shows that using DPO contributes to gaining more learnable knowledge from preference-style decisions.

4.4 Effectiveness of Facilitating CL

To further investigate the effectiveness of facilitating CL and the actual applicability of models, thirty additional topics from Debatepedia are selected to initiate new discussion rooms. Our collaborative agent and GPT-4o are respectively used to interact with the same set of student agents following the multi-agent interaction framework. We analyze every two consecutive stages identified by models to observe whether the discussion process is facilitated. The stage promotion result is shown is Figure 3. Stage(n,m) represents the promotion from previous stage n to next stage m ($n < m$). Theoretically, if a discussion is promoted ideally, the upper right corner of the stage-promotion heatmap will be darker. The stage promotion results with our collaborative agent achieve 10.91% Stage(1,2), 9.33% Stage(2,3), 11.61% Stage(3,4), and 10.28% Stage(4,5). Compared to GPT-4o, our agent can promote the discussion stages more smoothly. In other words, students are more likely to move forward to next stages without disrupting the collaboration. Moreover, GPT-4o tends to guide students in mid-stages (stage 2. exploration to stage 4. resolution), while our agent tends to guide in later stages (stage 3. integration to stage 5. reflection). This suggests that the discussion is more likely to enter the later stages under the guidance of our agent due to the effective feedback.

To reveal the teaching behaviors of models, we calculate the proportions of the number (1-10) of utterances between each two consecutive guidance for GPT-4o and our agent, as shown in Figure 4. Compared to our agent, the guidance frequency of GPT-4o is more intensive, which may lead to over-instruction and reduce Student autonomy. The average number of utterances of each student in discussions is 8.991 for GPT-4o and 12.465 for our agent respectively. This indicates that those students guided by our agent are more engaged.

We also visualize the promotion of students' shared metacognition with the guidance of our agent, as shown in Figure 5. Students often move from the cognition of individual learning I6 (I am aware of my existing knowledge) to the cognition of group learning G1 (I pay attention to the ideas of others) and G9 (I respond to the contributions that others make). This shows that students undergone a positive cognitive transformation from individual to collective in CL.

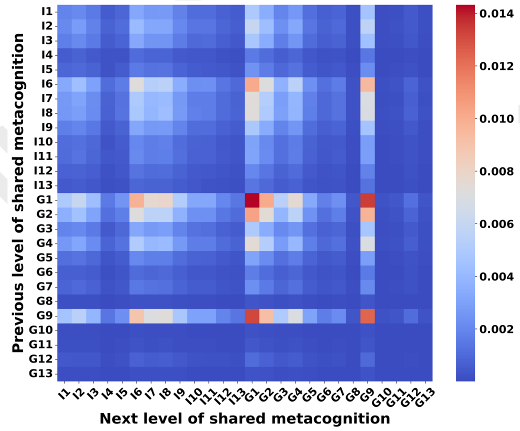

Figure 5: The promotion of students' shared metacognition with the guidance of our agent.

In summary, our agent can effectively drive discussions, enhance student engagement, and guide positive cognitive changes with high-quality instructive feedback.

5 Conclusion

This paper studies the problem of the AI tutor in moderating ToGD and proposes an LLM-based collaborative agent to generate instructive feedback at the right time without disrupting collaboration among students. It features an innovative CoI-based prompting strategy combined with fine-tuning to enhance the pedagogy adaptability of the generated instructive feedback. Additionally, a multi-agent framework is proposed to enable LLMs to explore and learn how AI tutors influence students' collaborative behaviors in order to produce effective instructive feedback. Meanwhile, a multi-party LLM-involved dialogue dataset with fine-grained annotations is produced. This dataset has the potential to advance future research on human-LLM agent interactions. Extensive experiments demonstrate that our LLM-based collaborative agent achieves SOTA performance in providing students with timely instructive feedback without disrupting the collaboration process.

Although our LLM-based collaborative agent performs very well in facilitating ToGD in empirical experiments, its effectiveness has not yet been validated in authentic classroom settings. Moreover, our LLM-based collaborative agent has no verification of the improvement of students' learning outcomes, e.g., academic performance. In the near future, we will conduct in-class trials and collect student data to further validate the effectiveness of our agents.

Acknowledgments

The work is fully conducted at The Education University of Hong Kong (EdUHK) and supported by grants from National Natural Science Foundation of China (No. 62372189) and EdUHK-ECNU projects (No. 02184 and No. 02185).

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

References

[Achiam *et al.*, 2023] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*, 2023.

[Akyol and Garrison, 2011] Zehra Akyol and D Randy Garrison. Assessing metacognition in an online community of inquiry. *The Internet and higher education*, 14(3):183–190, 2011.

[Bai *et al.*, 2023] Jinze Bai, Shuai Bai, Yunfei Chu, Zeyu Cui, Kai Dang, Xiaodong Deng, Yang Fan, Wenbin Ge, Yu Han, Fei Huang, et al. Qwen technical report. *arXiv preprint arXiv:2309.16609*, 2023.

[Cai *et al.*, 2024] Zhenyao Cai, Seehee Park, Nia Nixon, and Shayan Doroudi. Advancing knowledge together: integrating large language model-based conversational ai in small group collaborative learning. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–9, 2024.

[Carletta *et al.*, 2005] Jean Carletta, Simone Ashby, Sebastien Bourban, Mike Flynn, Mael Guillemot, Thomas Hain, Jaroslav Kadlec, Vasilis Karaiskos, Wessel Kraaij, Melissa Kronenthal, et al. The ami meeting corpus: A pre-announcement. In *International workshop on machine learning for multimodal interaction*, pages 28–39. Springer, 2005.

[Chen *et al.*, 2024] Jiahao Chen, Zitao Liu, Mingliang Hou, Xiangyu Zhao, and Weiqi Luo. Multi-turn classroom dialogue dataset: Assessing student performance from one-on-one conversations. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pages 5333–5337, 2024.

[Dan *et al.*, 2023] Yuhao Dan, Zhikai Lei, Yiyang Gu, Yong Li, Jianghao Yin, Jiaju Lin, Linhao Ye, Zhiyan Tie, Yougen Zhou, Yilei Wang, et al. Educhat: A large-scale language model-based chatbot system for intelligent education. *arXiv preprint arXiv:2308.02773*, 2023.

[Ding *et al.*, 2024] Yuyang Ding, Hanglei Hu, Jie Zhou, Qin Chen, Bo Jiang, and Liang He. Boosting large language models with socratic method for conversational mathematics teaching. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pages 3730–3735, 2024.

[Gan *et al.*, 2023] Wensheng Gan, Zhenlian Qi, Jiayang Wu, and Jerry Chun-Wei Lin. Large language models in education: Vision and opportunities. In *2023 IEEE international conference on big data (BigData)*, pages 4776–4785. IEEE, 2023.

[Garrison and Arbaugh, 2007] D Randy Garrison and J Ben Arbaugh. Researching the community of inquiry framework: Review, issues, and future directions. *The Internet and higher education*, 10(3):157–172, 2007.

[Garrison, 2022] DR Garrison. Shared metacognition in a community of inquiry. *Online learning*, 26(1):6–18, 2022.

[Gottipati *et al.*, 2013] Swapna Gottipati, Minghui Qiu, Yanchuan Sim, Jing Jiang, and Noah A. Smith. Learning topics and positions from Debatepedia. In David Yarowsky, Timothy Baldwin, Anna Korhonen, Karen Livescu, and Steven Bethard, editors, *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1858–1868, Seattle, Washington, USA, October 2013. Association for Computational Linguistics.

[Han *et al.*, 2023] Jieun Han, Haneul Yoo, Yoonsu Kim, Junho Myung, Minsun Kim, Hyunseung Lim, Juho Kim, Tak Yeon Lee, Hwajung Hong, So-Yeon Ahn, et al. Recipe: How to integrate chatgpt into efl writing education. In *Proceedings of the Tenth ACM Conference on Learning@ Scale*, pages 416–420, 2023.

[Holstein *et al.*, 2019] Kenneth Holstein, Bruce M McLaren, and Vincent Aleven. Designing for complementarity: Teacher and student needs for orchestration support in ai-enhanced classrooms. In *Artificial Intelligence in Education: 20th International Conference, AIED 2019, Chicago, IL, USA, June 25-29, 2019, Proceedings, Part I 20*, pages 157–171. Springer, 2019.

[Hu *et al.*, 2021] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021.

[Janin *et al.*, 2003] Adam Janin, Don Baron, Jane Edwards, Dan Ellis, David Gelbart, Nelson Morgan, Barbara Peskin, Thilo Pfau, Elizabeth Shriberg, Andreas Stolcke, et al. The icsi meeting corpus. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03).*, volume 1, pages I–I. IEEE, 2003.

[Jiang *et al.*, 2023] Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. Mistral 7b. *arXiv preprint arXiv:2310.06825*, 2023.

[Jin *et al.*, 2024] Hyoungwook Jin, Seonghee Lee, Hyungyu Shin, and Juho Kim. Teach ai how to code: Using large llnguage models as teachable agents for programming education. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–28, 2024.

[Joshi *et al.*, 2017] Mandar Joshi, Eunsol Choi, Daniel Weld, and Luke Zettlemoyer. TriviaQA: A large scale distantly supervised challenge dataset for reading comprehension. In Regina Barzilay and Min-Yen Kan, editors, *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1601–1611, Vancouver, Canada, July 2017. Association for Computational Linguistics.

[Karadzhov *et al.*, 2023] Georgi Karadzhov, Tom Stafford, and Andreas Vlachos. Delidata: A dataset for deliberation in multi-party problem solving. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW2):1–25, 2023.

Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

[Laal and Laal, 2012] Marjan Laal and Mozhgan Laal. Collaborative learning: What is it? *Procedia-Social and Behavioral Sciences*, 31:491–495, 2012.

[Lin, 2004] Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain, July 2004. Association for Computational Linguistics.

[Liu *et al.*, 2024a] Jiayu Liu, Zhenya Huang, Tong Xiao, Jing Sha, Jinze Wu, Qi Liu, Shijin Wang, and Enhong Chen. Socraticlm: Exploring socratic personalized teaching with large language models. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

[Liu *et al.*, 2024b] Zhengyuan Liu, Stella Xin Yin, Geyu Lin, and Nancy F. Chen. Personality-aware student simulation for conversational intelligent tutoring systems. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 626–642, Miami, Florida, USA, November 2024. Association for Computational Linguistics.

[Macina *et al.*, 2023] Jakub Macina, Nico Daheim, Sankalan Pal Chowdhury, Tanmay Sinha, Manu Kapur, Iryna Gurevych, and Mrinmaya Sachan. Mathdial: A dialogue tutoring dataset with rich pedagogical properties grounded in math reasoning problems. *arXiv preprint arXiv:2305.14536*, 2023.

[Nguyen *et al.*, 2024] Ha Nguyen, Victoria Nguyen, Saríah López-Fierro, Sara Ludovise, and Rossella Santagata. Simulating climate change discussion with large language models: Considerations for science communication at scale. In *Proceedings of the Eleventh ACM Conference on Learning @ Scale*, L@S '24, page 28–38, New York, NY, USA, 2024. Association for Computing Machinery.

[Pal Chowdhury *et al.*, 2024] Sankalan Pal Chowdhury, Vilém Zouhar, and Mrinmaya Sachan. Autotutor meets large language models: A language model tutor with rich pedagogy and guardrails. In *Proceedings of the Eleventh ACM Conference on Learning@ Scale*, pages 5–15, 2024.

[Papineni *et al.*, 2002] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: A method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318, 2002.

[Rafailov *et al.*, 2024] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 36, 2024.

[Rajpurkar *et al.*, 2016] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. SQuAD: 100,000+ questions for machine comprehension of text. In Jian Su, Kevin Duh, and Xavier Carreras, editors, *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2383–2392, Austin, Texas, November 2016. Association for Computational Linguistics.

[Shani *et al.*, 2024] Lior Shani, Aviv Rosenberg, Asaf Cassel, Oran Lang, Daniele Calandriello, Avital Zipori, Hila Noga, Orgad Keller, Bilal Piot, Idan Szpektor, et al. Multi-turn reinforcement learning from preference human feedback. *arXiv preprint arXiv:2405.14655*, 2024.

[Shea and Bidjerano, 2009] Peter Shea and Temi Bidjerano. Cognitive presence and online learner engagement: A cluster analysis of the community of inquiry framework. *Journal of Computing in Higher Education*, 21:199–217, 2009.

[Shea and Bidjerano, 2010] Peter Shea and Temi Bidjerano. Learning presence: Towards a theory of self-efficacy, self-regulation, and the development of a communities of inquiry in online and blended learning environments. *Computers & education*, 55(4):1721–1731, 2010.

[Shea *et al.*, 2012] Peter Shea, Suzanne Hayes, Sedef Uzuner Smith, Jason Vickers, Temi Bidjerano, Alexandra Pickett, Mary Gozza-Cohen, Jane Wilde, and Shoubang Jian. Learning presence: Additional research on a new conceptual element within the community of inquiry (coi) framework. *The internet and higher education*, 15(2):89–95, 2012.

[Sonkar *et al.*, 2023] Shashank Sonkar, Naiming Liu, Debshila Mallick, and Richard Baraniuk. CLASS: A design framework for building intelligent tutoring systems based on learning science principles. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 1941–1961, Singapore, December 2023. Association for Computational Linguistics.

[Sonkar *et al.*, 2024] Shashank Sonkar, Kangqi Ni, Sapana Chaudhary, and Richard Baraniuk. Pedagogical alignment of large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 13641–13650, Miami, Florida, USA, November 2024. Association for Computational Linguistics.

[Wang *et al.*, 2024] Shen Wang, Tianlong Xu, Hang Li, Chaoli Zhang, Joleen Liang, Jiliang Tang, Philip S Yu, and Qingsong Wen. Large language models for education: A survey and outlook. *arXiv preprint arXiv:2403.18105*, 2024.

[Wei *et al.*, 2021] Jason Wei, Maarten Bosma, Vincent Y Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M Dai, and Quoc V Le. Finetuned language models are zero-shot learners. *arXiv preprint arXiv:2109.01652*, 2021.

[Xu and Zhang, 2023] Songlin Xu and Xinyu Zhang. Leveraging generative artificial intelligence to simulate student learning behavior. *arXiv preprint arXiv:2310.19206*, 2023.