BIPED: Pedagogically Informed Tutoring System for ESL Education

Anonymous ACL submission

Abstract

Large Language Models (LLMs) have a great potential to serve as readily available and cost-efficient Conversational Intelligent Tutoring Systems (CITS) for teaching L2 learners of English. Existing CITS, however, are designed to teach only simple concepts or lack the pedagogical depth necessary to address diverse learning strategies. To develop a more pedagogically informed CITS capable of teaching complex concepts, we construct a BIlingual PEDagogically-informed Tutoring Dataset (**BIPED**) of one-on-one, human-to-human English tutoring interactions. Through post-hoc analysis of the tutoring interactions, we come up with a lexicon of dialogue acts (34 tutor acts and 9 student acts), which we use to further annotate the collected dataset. Based on a two-step framework of first predicting the appropriate tutor act then generating the corresponding response, we implemented two CITS models using GPT-4 and SOLAR-KO, respectively. We experimentally demonstrate that the implemented models not only replicate the style of human teachers but also employ diverse and contextually appropriate pedagogical strategies.

1 Introduction

As Large Language Models (LLMs) such as GPT (Achiam et al., 2023) revolutionize the field of Natural Language Generation, both researchers and practitioners have put an increasing amount of effort into developing Conversational Intelligent Tutoring Systems (CITS) that leverage the generative capabilities of LLM's. Specifically, LLMs have the potential to teach English as a Second/Foreign Language (ESL/EFL), for they may serve as readily-available tutors that can emulate native-speaking contexts.

However, most CITS proposed in literature for teaching ESL are restricted in scope and lack pedagogical depth. For example, works based on CIMA dataset (Stasaski et al., 2020) consider toy concepts of colored shapes and prepositions. Possibly due to limited scope, CIMA only considers a five tutor act classes ("hint", "question", "correction", "confirmation", and "other"). Similarly, the TSCC dataset (Caines et al., 2022) only considers abstract tutor act classes such as "scaffolding", and does not consider a more granular set of teaching strategies (e.g., inferential clues, teaching synonyms or antonyms, etc.). Consequently, models trained on CIMA (Stasaski et al., 2020) and TSCC (Caines et al., 2022) are limited to employing simplistic instructional strategies (Macina et al., 2023b).

The goal of our work is to develop CITS that can teach ESL by applying pedagogically meaningful teaching strategies adaptively to student's needs. To do so, we designed a tutoring session that focuses on comprehending a real news article, covering vocabulary, grammar, and relevant cultural contexts. Based on this, we construct a BIlingual PEDagogically-informed Tutoring Dataset (BIPED). For BIPED, we collected a bilingual, one-on-one human-to-human tutoring dataset. Thereafter, we analyzed the dataset post-hoc from a pedagogical viewpoint, and developed a categorization of dialogue acts, which comprises of 34 tutor acts and 9 student acts. Finally, we annotated the data using the dialogue act category to obtain the training dataset.

As for the development of CITS, we employ the framework (Macina et al., 2023b; Wang et al., 2023a) whereby the LLM first chooses the suitable tutor act, then generates the corresponding utterance. We believe this approach enables the model to generate a more focused response that does not deviate from the chosen tutor intent. We consider two implementations of such CITS, one based on GPT-4 prompting, and another based on fine-tuning SOLAR-KO 10.7B (L. Junbum, 2024), which is a public model trained on a Korean corpus. For GPT-4, our dataset is used as examples for

in-context learning. For fine-tuning, we explored a multi-task learning approach, where the model is instruction-tuned to carry out tutor act prediction and response generation, as well as other tasks for context grounding.

In our experiments, we show through various metrics that the implemented models are capable of mimicking human teacher's utterance style, as well as their pedagogical strategies. Specifically, we show that our models choose appropriate and diverse tutor acts, and generate human-like utterances, as measured by sentence similarity metrics.

The contributions of our paper are as follows:

- We provide BIPED, a dataset curated from bilingual human-to-human tutoring dialogues that come annotated with specific tutoring acts. Importantly, the lesson topic is complex enough to require sophisticated pedagogical methods.
- We implement two CITS based on BIPED. In particular, we explore both in-context learning and fine-tuning approaches. It is the pioneering work in utilizing instruction tuning techniques in the development of CITS.
- We provide experimental results demonstrating the implemented models' ability to mimic human tutors.

2 Related Work

2.1 Conversational Intelligent Tutoring Systems (CITS)

Some of conversational tutoring systems make effective use of GPT through the implementation of prompt engineering, as witnessed in Tack and Piech, 2022 and Abdelghani et al., 2022. Moreover, there exist studies that tailor open-source LLMs to fulfill distinct objectives (Qian et al., 2023; Huang et al., 2022), and the generation of dialogue in conjunction with act classification is also gaining ground (Wang et al., 2023a). However, despite these advancements in CITS using LLMs, a perceivable lack in the pedagogical approach is observed (Macina et al., 2023b; Kasneci et al., 2023). These systems fall short of human tutors in emulating tutor's speech, comprehending students, and assisting them in an appropriate way (Tack and Piech, 2022). In order to ensure that tutoring systems can effectively mimic the qualities of an actual tutor, the need for the integration of pedagogical attributes into the design of these systems is underscored (Du Boulay and Luckin, 2016).

2.2 Tutoring Dialogue Dataset

In the field of general natural language processing, the procedure of identifying the dialogue act serves as the preliminary phase in modeling intricate discourse structures (Stolcke et al., 2000; Budzianowski et al., 2018; Saha et al., 2020). Similarly, within the realm of education, related works also focus on dialogue acts within the tutoring dialogues (Caines et al., 2022; Tan et al., 2023; Lin et al., 2023; Nye and Morrison). In contrast to other areas, tutoring dialogue acts possess distinct characteristics in that they include the instructional intentions of tutors, with the objective of elevating the effectiveness of the tutoring process. There exist extensive studies dedicated to the classification of teaching strategies within the educational field (Brown and Lee, 2015; Brown, 2014; Walsh, 2006, 2013). While numerous studies of CITS draw from this established body of knowledge, they often encounter constraints. Due to the complexity of incorporating an expansive range of pedagogical methods within CITS, numerous studies often narrow the scope to a limited set of teaching strategies (Alic et al., 2022; Demszky and Hill, 2022), or alternatively, focus solely on teaching simple knowledge concepts which require a more confined set of instructional strategies (Stasaski et al., 2020). To address this limitation, our work endeavors to integrate a diverse array of teaching methodologies, devised in collaboration with domain experts.

There is a considerable variation across existing tutoring dialogue datasets in terms of its specific composition. There are tutoring dialogues assembled from interactions between learners and automated systems (Huang et al., 2022; Macina et al., 2023a). Meanwhile, Stasaski et al., 2020 have generated dialogues wherein experts simulate both the tutor and student roles. Nonetheless, to authentically encapsulate the dynamic interactions between the tutor and the student, it becomes imperative to establish a real teaching context. Moreover, focusing on ESL, there exists a significant deficit in the availability of bilingual data. In the context of language learning, pedagogical strategies that simultaneously consider both languages – for instance, code-switching – play a pivotal role (Brown, 2014). Thus, the shortage of bilingual data is restricting the efficacy and scope with respect to language instruction (Nguyen et al., 2022).

Highlighting that our dataset, BIPED, is primarily focused on human-to-human tutoring dialogues

Metric	TSCC v2(Caines et al., 2022)	TalkMoves(Suresh et al., 2022)	NCTE(Demszky and Hill, 2022)	BIPED
# tutoring sessions	260	567	1,660	112
# tutors	2	-	317	4
# students	13	-	10,817	60
Avg. turns per session	159.5	211.2	350	105.7
Avg. words per turn	T 10.8 / S 6.4	T 8.2 / S 5.9	T 29.1 / S 4.0	T 31.7 / S 8.2
# dialogue acts	T,S 12	T 6 / S 4	T 3 / S 2	T 34 / S 9
domain	ESL	math	math	ESL
setting	1:1	classroom	classroom	1:1

Table 1: Comparison of BIPED with existing dialogue tutoring datasets.

including annotations, we provide an overview of descriptive statistics with a comparison to existing datasets in comparable settings in Table 1.

3 Dataset: BIPED

This section describes our dataset, BIPED, detailing its design, collection, annotation process, and essential statistics.

3.1 Data Design

We aim to construct a comprehensive tutoring dataset that contains the entire dialogues from tutoring sessions, with the corresponding dialogue acts and encompassing contents labeled on each utterance. Additionally, we categorized the correctness of student responses to inquiries posed by the tutor. To provide a clearer understanding of the structure of our dataset, a sample of the dataset is displayed in Figure 1.

3.2 Data Collection

For the data collection through one-to-one tutoring sessions, we recruited four experienced ESL tutors and 60 students with a willingness to pursue English studies. Every student was partnered with a single tutor, and over the course of three days, they had three successive 30-minute one-to-one tutoring sessions. Prior to the first session, tutors were given an additional 10-minute to comprehend the student's individual status, including their particular objectives for studying English, their methods of studying English, and also their personal interests. The learning contents, developed by subject matter experts in ESL teaching, consisted of an English article along with activities focused on vocabulary and reading comprehension. Based on the student's pretest results and the ongoing assessment of their progress throughout the class, including the preceded 10-minute talk, the tutors were required to adapt their teaching style to each unique learner, selectively determining the learning materials to cover. The tutoring sessions were held in an audio-only setting, and the transcriptions of these sessions were obtained with the assistance of Whisper (Radford et al., 2023).

The dataset contains 112 tutoring sessions[1], as outlined in Table 1. On average, each session consisted of roughly 105.7 turns. Analysis reveals that the utterances made by tutors in a single turn contained an average of 31.7 words, while the student's utterances were more concise, averaging 8.2 words per utterance.

In order to confirm the educational significance of the tutoring sessions, students were required to take a pretest and posttest surrounding these sessions. Drawing upon the computation of the normalized learning gain, the average result was fount to be 0.089, demonstrating the tutoring sessions' effectiveness in enhancing the students' English proficiency. Additionally, a feedback survey was deployed to understand their perspectives and feelings towards the courses. Overall, the feedback suggested a general contentment with the educational experience provided. For instance, for the questions "To what extent did you find the class satisfactory?", "How suitable for you was the teaching method employed by the tutor during the class?", "Was the class conducted with consideration for your linguistic background and cultural understanding?", the respondents indicated high approval with average scores of 4.65, 4.60, and 4.68 out of a possible 5, respectively.

3.3 Data Annotation

To accurately encapsulate the diverse intentions of both tutors and students within our dataset, we defined 34 tutor acts and 9 student acts in collaboration with experienced ESL instructors. Our taxonomy was grounded in prior research (Caines et al., 2022; Stasaski et al., 2020), studies related to teaching methods (Brown, 2014; Brown and Lee, 2015), and underwent further refinements after observing

[1]We confine the utilization of data to a subset that has been fully annotated, despite a larger volume of collected data.

ex) Activity 3-9 (Gap Fills)

You likely won't get answers from the other person, but a therapist can help you _____ these issues in an introspective way, Durvasula says.

1) Vent 2) Navigate 3) Abuse 4) Withdraw

Speaker	Utterance	Act	Correctness	Content ID
tutor	좋아요. 그럼 Activity3-9로 넘어가 볼까요? 답이 몇번일까요? (eng) Good. Shall we move on to Activity3-9? Guess what number the answer is?	t.assess. display_question		Activity3-9
student	navigate?	s.answer.answer	high	
tutor	맞아요. "navigate"도 우리가 잘 쓸 수 있는 영어 표현이죠. "navigate these issues in an introspective way"라는 문장에서 "navigate"는 가이드를 해줄 것이다 잘 지도해줄 것이다. 이런 뜻입니다. (eng) Correct. "navigate" is also an English expression that we can use well. In the sentence "navigate these issues in an introspective way," "navigate" means to guide well, to give good directions. That's what it means.	t.teach.direct_answer		
tutor	여기에서 "introspective"는 어떤 뜻일까요? (eng) What does "introspective" mean here?	t.assess. follow_up_question		
student	내면을 바라보는? (eng) Looking inward?	s.answer.answer	high	
tutor	너무 좋아요. (eng) Very good.	t.engage.encourage		
tutor	"intro"는 안을 보는 것을 의미하고, "spect"는 "spectrum" 할 때 "spect"거든요. "spectrum"은 무언가를 보는 느낌이에요. 그래서 "introspective"는 자기 반성적인 방식으로, 내면을 바라보는 것을 의미해요. (eng) "intro" means looking inside, and "spect" is as in "spectrum". "spectrum" has the feeling of looking at something. So "introspective" means in a self-reflective way, looking inward.	t.teach.method. vocab_expression. etymology		

Figure 1: Example of our dataset, BIPED

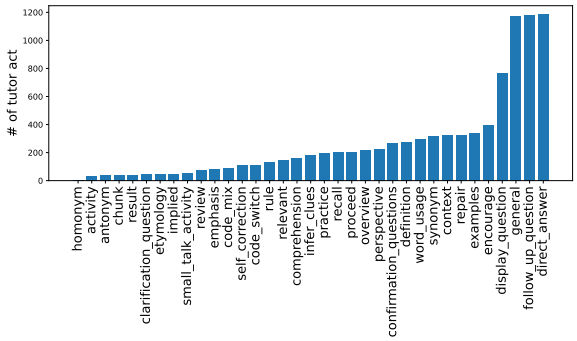

Figure 2: Distribution of tutor acts in BIPED

multiple pilot classes. As the tutor acts play a crucial role in influencing the efficacy of a tutoring session, we have accorded significant importance to the process of categorizing tutor acts. Tutor acts have been classified into five primary categories:

- *General*: general utterances including greeting, farewell, small talk, short reply, etc.
- *Operational*: utterances designed to facilitate the smooth operation of the class.
- *Assessment*: utterances intended to evaluate the student's learning progress.
- *Teaching*: utterances dedicated to teach a certain knowledge to the student.
- *Engagement*: utterances tailored to encourage student participation and involvement in the class.

The respective categories encompass 1,3,4,22, and 4 distinct tutor acts. Prior research has primarily concentrated on broader behaviors, and current language models have shown proficiency in their application to such contexts. Given the limitations observed within current taxonomies in terms of capturing the diversity of teaching methods, our study endeavors to provide a more detailed examination of the "Teaching" category, while drawing upon preliminary studies in reference to the remaining categories.

Concurrently, in a more simplified manner, student acts are divided into four distinct categories: *General*, *Operational*, *Question*, and *Answer*. These particular categories encompass 1, 3, 2, and 3 unique student acts respectively. Detailed explanations and examples of both tutor and student acts are exhibited in the Appendix A and B.

Initially, we implemented a set of processing steps, including the anonymization and correction of grammatical errors. Subsequently, with this taxonomy design, each utterance were annotated with the relevant act labels by our trained human annotators. In cases where a single utterance encompassed multiple acts, the utterance was split to ensure that each utterance corresponded to a single action. These segmented utterances are referred to as *act-level utterances*. That is, a single turn might comprise several act-level utterances. The distribution of tutor acts is represented in Fig 2. Our observations revealed that, on average, there were

145 act-level utterances per session, with tutors contributing 83.2 and students contributing 53.1. This suggests that tutors employed a variety of actions more frequently within the session.

4 Framework

To build CITS that strategically incorporates pedagogical considerations, we follow the framework (Macina et al., 2023c; Wang et al., 2023b) where the LLM (1) chooses the appropriate intervention class, (2) then generates the response accordingly. We advocate this two-step approach for the following two reasons.

Interpretability. First, the proposed two-step framework allows us to better understand the intention of the model (tutor acts) when generating a certain response for a given dialogue context. As a consequence, this enables us to more rigorously evaluate the model from a pedagogical perspective. Specifically, human experts may now assess whether the model correctly selected the appropriate type of intervention, and whether the generated response is suitable.

Controllability. More importantly, the proposed framework allows for greater controllability of CITS. By conditioning the LLM to a specific tutor act class, we ensure more focused response generation without deviation from the target act. We expect this to be essential when generating more sophisticated responses (e.g., teaching etymology, code-mixing), for these are sparsely distributed compared to more generic responses such as simple direct answer as shown in Fig 2.

Moreover, a crucial aspect for controllability is the model's amenability to improvement via continual refinement. Our framework allows for a divide-and-conquer approach for model refinement; namely, given dialogue context and student response, instead of requiring the human expert to generate a full response, we may let them simply label the best teacher response class (for instruction tuning), or a ranked list of teacher response classes (for RLHF). As for improving response generation, we may consider sophisticated data augmentation schemes, where we synthesize responses for a given tutor act class ased on retrieved samples of the same class.

5 Method

In this section, we introduce two concrete implementations of the framework proposed in Section 4. First implementation leverages GPT4's in-context learning ability and is easily implemented via prompting, while the second implementation instruction-tunes LLaMa to perform various relevant tasks.

5.1 GPT4 Prompting

We first consider a GPT4-based implementation via extensive prompting. While this implementation does not require model training, it is costly and requires a large amount of tokens for in-context learning.

As outlined in Appendix C, our prompt structure adheres to our framework. The first step involves choosing an effective tutor act from a provided tutor acts list. Next, an utterance that aligns with this chosen tutor act is generated. These steps consider both dialogue context and learning content provided in the testset. For few-shot learning, an example scenario from the training set that matches the tutor act in the task scenario is also provided. This includes dialogue context, learning content, a chosen tutor act, and a generated utterance.

5.2 Fine-tuning

We consider instruction-tuning a pre-trained LLM on the dataset introduced in Section 3. Our goal is to demonstrate that, with properly curated datasets and appropriate fine-tuning methods, we may enable open-source models to rival GPT-4's performance. We instruction-tuned open-sourced LLM on four sub-tasks, each of which we detail below.

Task 1. Tutor act prediction To make our model select the most effective teaching strategy, we task the model to predict the next tutor act based on given learning content and dialogue context. Given a dialogue in the training dataset, we created instruction samples for every dialogue turn of the tutor.

Task 2. Next response generation As a conversational tutor for Korean L2 learners of English, the model must be able to generate responses that are (1) fluent in both Korean and English; and (2) contextually relevant and matching the tutor act the response is conditioned on. To achieve this, we devise a task of generating the tutor's response based on learning content, dialogue history, and the conditioning tutor act. By training the model to imitate expert human tutor's responses, we expect the model to obtain bilingual fluency, as well as the ability to generate appropriate response based on the aforementioned conditioning information.

Table 2: Instruction types for fine-tuning model.

Type	Instruction
Tutor act prediction	### Instruction: Select an appropriate next tutor act based on the given 'dialogue context' and 'act candidates'. In the context, tutor uses various tutor act and contents(activities or passages), depending on the student's abilities and questions. Tutor act represents a teaching method that you can exploit. Choose the best intervention catering to the context and understanding the meaning of each act." ### Input: - Context: {*dialogue context (Appendix D)*}, - Act candidates: {*act candidates (Appendix A)*} ### Response: t.teach.method.vocab.expression.etymology
Tutor utterance generation	### Instruction: As an English tutor for a Korean student, your job is to create coherent and tailored responses based on the given 'dialogue context' and 'tutor's act'. In the context, tutor uses various teaching acts and contents(activities or passages), depending on the student's abilities and questions. The given action is a teaching method that you have to exploit. ### Input: - Context: {*dialogue context (Appendix D)*}, - Act: t.teach.method.vocab.expression.etymology, Explain a word or a phrase based on its etymology. When possible, discuss the prefix and suffix of the word. ### Response: "intro"는 안을 보는 것을 의미하고, "spect"는 "spectrum"할 때 "spect"거든요. "spectrum"은 무언가를 보는 느낌이에요. 그래서 "introspective"는 자기 반성적인 방식으로, 내면을 바라보는 것을 의미해요. (eng) "Intro" means looking inward. "spect" is same as "spect" in "Spectrum" and feels like seeing something. So, "introspective" means in a self-reflective manner, looking inward.
Missing context inference	### Instruction: Predict a speaker and their utterance to fit between the given 'dialogue context' of a student and a teacher and the final 'utterance' made by the teacher. ### Input: - Context: {*partial context (Appendix D)*}, - Utterance:ïntro"는 안을 보는 것을 의미하고, "spect"는 "spectrum"할 때 "spect"거든요. "spectrum"은 무언가를 보는 느낌이에요. 그래서 "introspective"는 자기 반성적인 방식으로, 내면을 바라보는 것을 의미해요. (eng) "Intro" means looking inward. "spect" is same as "spect" in "Spectrum" and feels like seeing something. So, "introspective" means in a self-reflective manner, looking inward. ### Response:
Minority act utterance generation	### Instruction: As an English teacher for Korean students, generate tutor's utterance to teach the given learning content with the act. ### Input: - Content: [Activity 3-9] You likely won't get answers from the other person, but a therapist can help you ___ these issues in an introspective way, Durvasula says. (1) Vent (2) Navigate (3) Abuse (4) Withdraw - Act: t.teach.method.vocab.expression.etymology, Explain a word or a phrase based on its etymology. When possible, discuss the prefix and suffix of the word. ### Response: "Intro"는 "안쪽의"라는 뜻이고, "spect"는 "보다"라는 뜻입니다. 그래서 "Introspective"는 "자아성찰하는"이라는 뜻입니다. (eng) "Intro" means "inner" and "spect" means "to see." So, "introspective" means "self-reflective."

Task 3. Missing context inference We found out that, trained on Task 1-2, the model often generated responses that are not properly grounded in the recent dialogue context. This comes from the inherent task difficulty: compared to simple conversational tasks, teaching often requires a purposeful and focused long-context generation, where one small digression may hurt the overall quality of the generated response. To address this issue, we devise the following task: Given learning content, dialogue history, and corresponding teacher response, we remove this and that. The model is then tasked to predict the missing context given the rest of the information. This way, we expect the model's utterance to be locally more consistent.

Task 4. Minority act utterance generation Finally, we notice that there are minority tutor act classes that are, despite their pedagogical important, sparsely distributed in the training dataset. Therefore, to make sure the model understands these act classes better, we additionally collected the following data - for a given learning content and a selected group of minority act classes, we asked experts to provide plausible utterances for the corresponding act classes. Then, we tasked the model to generate these utterances conditioned on the corresponding learning content and act class. Although this additional dataset lack specific dialogue context, we expect it to guide the model in generating utterances for minority act classes.

6 Experiments

6.1 Dataset split

The experiment involved dividing BIPED randomly into training and test sets, organised by session. This was done based on the distribution of tutors within the dataset, allocating two sessions to one tutor, and one session each to the remaining tutors. BIPED of 112 tutoring sessions was thus split into 102 training sessions and 10 test sessions.

Extra data for generating minority act utterances, part of Task 4, were only included in the training set. The entire training data was reformulated into different instruction-tuning formats for each task.

The test set was devised into 220 scenarios, 10 unique scenarios for every one of the 22 tutor acts in the teaching category. Each test scenario was shaped around an act-level utterance initiated by a human tutor applying the respective tutor act, supported by the associated dialogue context and learning content.

6.2 Evaluation metric

We conducted evaluations of the models based on our framework, focusing on two main aspects. Firstly, we evaluated the accuracy and invariability of tutor act predictions. For accuracy, we assessed the rate at which the model's predicted tutor act aligned with the actual act selected by the tutor in the test set. For invariability, we noted the frequency of each selected tutor act and calculated the mean absolute difference from a target count of 10.

Secondly, we assessed the generated utterances. To measure the similarity between the generated and target tutor utterances, we utilized the sacrebleu implementation of BLEU(sBLEU) (Post, 2018) and BERT Score (Zhang* et al., 2020) using bert-base-multilingual-cased checkpoint. We note that the finetuned model sequentially predicted tutor acts and generated utterances based on these predictions. Additionally, we measured the coherence between generated utterances and previous ones in the given dialogue context using embedding similarity, specifically relying on OpenAI's text embedding(text-embedding-3-large).

6.3 Training details

For finetuning we utilize SOLAR-KO-10.7B((L. Junbum, 2024)), a model further trained on a Korean corpus to enhance bilingual comprehension and fluency. We adopt AdamW optimizer (Loshchilov and Hutter, 2019) with a learning rate of 2e-5 and a batch size of 64. The maximum length is 4096. We train the model using 8 V100 GPUs with deepspeed Zero-3 offload for 2 epochs.

6.4 Results

The findings from our experiment, illustrated in the main table 3, are compared against various models ranging from base GPT to finetuned models. Base GPT in Table 3 corresponds to the GPT4 model bereft of any tuition act classification or our humanoid tutoring examples, where its prompt is given in Appendix C. We have also drawn a comparison involving GPT 3.5 and 4 in a zero-shot and one-shot configuration. While they all adhere to the identical prompt structure, (as seen in Appendix C), they diverge in aspects such as model compatibility and whether an example scenario is included. Lastly, we compared our two finetuned models. Finetuned refers to our instruction-tuned model, discussed in the method section 5, which was trained with instruction tasks 1 through 4. Conversely, finetuned (ablation) refers to the model trained with instruction tasks 1 and 2 only. Ground-truth represents the reformulated data gathered from human tutoring present in the test set.

In assessing the tutor act criteria, our finetuned model outperforms other models in both accuracy and invariability as shown in Table 3. Although the accuracy of 0.259 may initially appear low, it's important to consider the complexity of choosing from amongst 34 distinct tutor acts. Moreover, a lower invariability score of 5.273 is indicative of our finetuned model's ability to utilize a wider array of tutor acts, thereby demonstrating more diverse pedagogical approaches. While the invariability score provides one aspect of variability, Figure 3 presents a distribution of chosen tutor acts for the comparative models, which offers a more comprehensive view. Clearly, our finetuned model exhibits the greatest number of distinct tutor act classes.

With respect to utterance-related metrics, our finetuned model delivers outstanding performances. It records the highest sBLEU score at 15.874 and BERTScore at 0.716, demonstrating its commendable ability in generating utterances that closely mirror the target tutor utterances. While the embedding similarity score of our fine-tuned model is slightly lower than that of GPT3.5 (1-shot), it still exhibits higher coherence between individual utterances and their corresponding dialogue context.

In terms of the length of the generated utterances, our model closely mimics the length of tutor

Table 3: Results of prompted GPTs and finetuned models

Metric	Tutor Act		Utterance			
	Accuracy↑	Invariability↓	sBLEU↑	BERTScore↑	Embedding sim.↑	Length
Base GPT	0.155	6.955	1.802	0.655	0.635	80 ± 31
GPT-3.5 (0-shot)	-	-	1.849	0.646	0.661	45 ± 36
GPT-3.5 (1-shot)	-	-	2.423	0.658	**0.704**	45 ± 34
GPT-4 (0-shot)	-	-	1.687	0.659	0.668	66 ± 28
GPT-4 (1-shot)	0.250	7.682	2.608	0.669	0.686	58 ± 29
Finetuned	**0.259**	**5.273**	**15.874**	**0.716**	0.680	39 ± 52
Finetuned(ablation)	-	-	14.199	0.708	0.673	34 ± 20
Ground-truth	1.0		100.	1.	0.707	27 ± 22

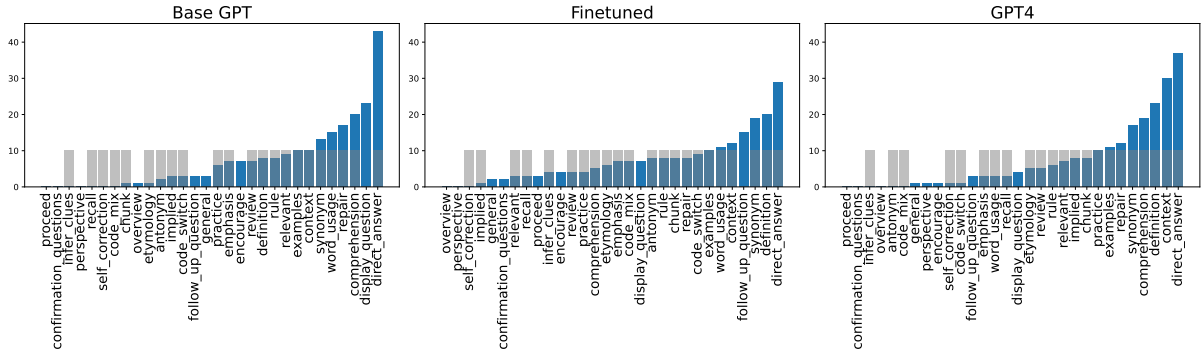

Figure 3: The comparative distributions of chosen tutor acts. The grey bars represent the distribution of tutor acts in the test set while the blue bars denote the distribution of chosen tutor acts by Base GPT, our fine-tuned model, and GPT4 (1-shot).

utterances, whereas the general GPT models often produce unnecessarily long responses. We interpret this as positive, based on the assumption that the appropriate length of a tutor's utterance plays a vital role in maintaining student engagement.

Lastly, the finetuned model, trained with tasks 1 through 4, has outperformed the ablation model in all evaluated metrics. This outcome highlights the added value of integrating additional instruction tasks (i.e., tasks 3 and 4) for improved tutoring performance.

6.5 Example of Interactive Evaluation

Our results primarily focus on static evaluation, where we predict the next tutor act and generate a corresponding response in one step for the test scenarios. However, a complete evaluation of the tutoring system, developed using our proposed models, necessitates interactive assessment with human students. This aspect of evaluation will be subjected to further study. Regardless, we offer a sneak peak into how an interaction plays out between a human and two tutoring models, namely GPT-4 (1-shot) and our finetuned model.

The dialogue examples are included in Appendix E. Generally, when interacting with our tutoring systems, GPT-4 (1-shot) tends to produce lengthier responses, focusing mainly on transferring knowledge to students. Conversely, our fine-tuned model generates concise responses and actively promotes student participation. An interesting observation is that when a student insists that an incorrect answer is correct, GPT-4 maintains its answer, yet appears to accept the student's assertion (t.teach.emphasis). In contrast, our fine-tuned model re-emphasizes the correct answer (t.teach.repair), thereby highlighting the student's error.

7 Conclusion

In this study, we demonstrated the effectiveness of our proposed models in developing conversational intelligent tutoring systems for ESL learners. They excelled in replicating the utterance style of human tutors and implementing various and appropriate pedagogical strategies. Much of this improvement can be attributed to our newly constructed Bilingual and Pedagogically-Informed Tutoring Dataset (BIPED). Additionally, our two-step framework and the use of four instruction tuning tasks further contributed to the models' improvement.

8 Limitations

Despite the encouraging results, our study has some limitations. First, our project lacks an interactive evaluation, left as future research direction. A formal evaluation of complete tutoring systems through human study would provide useful insights into their efficacy. Second, we were unable to utilize larger parameter models like LLaMa 70b (Touvron et al., 2023), which could potentially improve the teaching capabilities of CITS. This restriction is worth considering for future studies, particularly with advancing computational resources and technologies. Finally, the use of diverse instruction tuning tasks for multi-task modeling, using existing datasets and data augmentation, is not covered in this study. We believe this aspect carries opportunities for more nuanced model development and will be explored in subsequent work.

9 Potential Risks

Our research holds great promise, but it's also important to acknowledge potential risks. Firstly, Large Language Models might inadvertently teach misinformation due to their susceptibility to the 'hallucination' phenomenon, which generates information that isn't grounded in real-world facts. Secondly, there's risk of over-reliance on these intelligent tutoring systems. While they can play a pivotal role in language learning, they can't entirely mimic human tutors, especially in terms of adaptive strategies and emotional intelligence.

References

R Abdelghani, YH Wang, X Yuan, T Wang, H Sauzéon, and PY Oudeyer. 2022. Gpt-3-driven pedagogical agents for training children's curious question-asking skills. arxiv. *preprint arXiv*, 2211.

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

Sterling Alic, Dorottya Demszky, Zid Mancenido, Jing Liu, Heather Hill, and Dan Jurafsky. 2022. Computationally identifying funneling and focusing questions in classroom discourse. *arXiv preprint arXiv:2208.04715*.

H Douglas Brown. 2014. *Principles of language learning and teaching: A course in second language acquisition*. Pearson.

H Douglas Brown and Heekyeong Lee. 2015. *Teaching by Principles: An Interactive Approach to Language Pedagogy*. P. Ed Australia.

Paweł Budzianowski, Tsung-Hsien Wen, Bo-Hsiang Tseng, Inigo Casanueva, Stefan Ultes, Osman Ramadan, and Milica Gašić. 2018. Multiwoz–a large-scale multi-domain wizard-of-oz dataset for task-oriented dialogue modelling. *arXiv preprint arXiv:1810.00278*.

Andrew Caines, Helen Yannakoudakis, Helen Allen, Pascual Pérez-Paredes, Bill Byrne, and Paula Buttery. 2022. The teacher-student chatroom corpus version 2: more lessons, new annotation, automatic detection of sequence shifts. In *Swedish Language Technology Conference and NLP4CALL*, pages 23–35.

Dorottya Demszky and Heather Hill. 2022. The ncte transcripts: A dataset of elementary math classroom transcripts. *arXiv preprint arXiv:2211.11772*.

Benedict Du Boulay and Rosemary Luckin. 2016. Modelling human teaching tactics and strategies for tutoring systems: 14 years on. *International Journal of Artificial Intelligence in Education*, 26:393–404.

Jin-Xia Huang, Yohan Lee, and Oh-Woog Kwon. 2022. Direct: Toward dialogue-based reading comprehension tutoring. *IEEE Access*, 11:8978–8987.

Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günnemann, Eyke Hüllermeier, et al. 2023. Chatgpt for good? on opportunities and challenges of large language models for education. *Learning and individual differences*, 103:102274.

L. Junbum. 2024. Solar-ko-10.7b.

Jionghao Lin, Wei Tan, Ngoc Dang Nguyen, David Lang, Lan Du, Wray Buntine, Richard Beare, Guanliang Chen, and Dragan Gašević. 2023. Robust educational dialogue act classifiers with low-resource and imbalanced datasets. In *International Conference on Artificial Intelligence in Education*, pages 114–125. Springer.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. In *International Conference on Learning Representations*.

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

Jakub Macina, Nico Daheim, Lingzhi Wang, Tanmay Sinha, Manu Kapur, Iryna Gurevych, and Mrinmaya Sachan. 2023b. Opportunities and challenges in neural dialog tutoring. *arXiv preprint arXiv:2301.09919*.

Jakub Macina, Nico Daheim, Lingzhi Wang, Tanmay Sinha, Manu Kapur, Iryna Gurevych, and Mrinmaya Sachan. 2023c. Opportunities and challenges in neural dialog tutoring. In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*, pages 2357–2372, Dubrovnik, Croatia. Association for Computational Linguistics.

Li Nguyen, Zheng Yuan, and Graham Seed. 2022. Building educational technologies for code-switching: Current practices, difficulties and future directions. *Languages*, 7(3):220.

Borhan Samei Vasile Rus Benjamin Nye and Donald M Morrison. Hierarchical dialogue act classification in online tutoring sessions.

Matt Post. 2018. A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Brussels, Belgium. Association for Computational Linguistics.

Kun Qian, Ryan Shea, Yu Li, Luke Kutszik Fryer, and Zhou Yu. 2023. User adaptive language learning chatbots with a curriculum. *arXiv preprint arXiv:2304.05489*.

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. 2023. Robust speech recognition via large-scale weak supervision. In *International Conference on Machine Learning*, pages 28492–28518. PMLR.

Tulika Saha, Aditya Patra, Sriparna Saha, and Pushpak Bhattacharyya. 2020. Towards emotion-aided multimodal dialogue act classification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4361–4372.

Katherine Stasaski, Kimberly Kao, and Marti A Hearst. 2020. Cima: A large open access dialogue dataset for tutoring. In *Proceedings of the Fifteenth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 52–64.

Andreas Stolcke, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema, and Marie Meteer. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational linguistics*, 26(3):339–373.

Abhijit Suresh, Jennifer Jacobs, Charis Harty, Margaret Perkoff, James H Martin, and Tamara Sumner. 2022. The talkmoves dataset: K-12 mathematics lesson transcripts annotated for teacher and student discursive moves. *arXiv preprint arXiv:2204.09652*.

Anaïs Tack and Chris Piech. 2022. The ai teacher test: Measuring the pedagogical ability of blender and gpt-3 in educational dialogues. *arXiv preprint arXiv:2205.07540*.

Wei Tan, Jionghao Lin, David Lang, Guanliang Chen, Dragan Gašević, Lan Du, and Wray Buntine. 2023. Does informativeness matter? active learning for educational dialogue act classification. In *International Conference on Artificial Intelligence in Education*, pages 176–188. Springer.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. 2023. Llama 2: Open foundation and fine-tuned chat models.

Steve Walsh. 2006. *Investigating classroom discourse*. Routledge.

Steve Walsh. 2013. *Classroom discourse and teacher development*. Edinburgh University Press.

Lingzhi Wang, Mrinmaya Sachan, Xingshan Zeng, and Kam-Fai Wong. 2023a. Strategize before teaching: A conversational tutoring system with pedagogy self-distillation. *arXiv preprint arXiv:2302.13496*.

Lingzhi Wang, Mrinmaya Sachan, Xingshan Zeng, and Kam-Fai Wong. 2023b. Strategize before teaching: A conversational tutoring system with pedagogy self-distillation. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 2268–2274, Dubrovnik, Croatia. Association for Computational Linguistics.

Tianyi Zhang*, Varsha Kishore*, Felix Wu*, Kilian Q. Weinberger, and Yoav Artzi. 2020. Bertscore: Evaluating text generation with bert. In *International Conference on Learning Representations*.

A Tutor Act Classification

Translations are provided below the examples in gray, and words that are intended to be uttered in English are bolded.

Tutor Act	Explanation
t.general	General utterances including greeting, farewell, small talk, short reply, etc. ex. *"안녕하세요.", "준비 되셨나요?"* *(eng) "Hello.", "Are you ready?"*
t.operational.overview	Give an overview of the tutoring session at the start of the session or provide the overview for the next session at the end of the session. ex. *"오늘은 ∼방식으로 수업을 진행하려고 합니다."* *(eng) "Today, I plan to conduct the class in a way of ∼."*
t.operational.proceed	Notify the student that you will proceed with the class. When appropriate - for example, right after teaching a concept - ask if the student has any more questions before moving on. ex. *"그럼 수업을 진행해보도록 하겠습니다. 혹시 더 질문이 없으시다면 넘어가도 될까요?"* *(eng) "Let's proceed with the lesson then. If there are no more questions, can we move on?"*
t.operational. clarification_question	When the student asks a question whose meaning is unclear, request clarification. You may additionally guess the intended meaning of the question and ask the student whether your understanding is correct. ex. *"∼ 말씀하시는 건가요?"* *(eng) "Are you referring to ∼?"*
t.assess.display_question	Refer to a question activity and ask the student the question. For example, for P2, questions can be from [P2 Synonyms & Antonyms], [P2 Phrasal Verbs], and [P2 T/F Questions]. ex. *"Activity 1의 9번은 정답이 뭐라고 생각하시나요?"* *(eng) "What do you think is the answer to number 9 in Activity 1?"*
t.assess. confirmation_question	Ask to confirm the student's understanding of specific content. ex. *"이해되시나요?"* *(eng) "Do you understand?"*
t.assess. follow_up_question	After the student's initial response to the question displayed, assess their knowledge. If the student answered correctly, ask a follow-up question to confirm mastery of the knowledge concepts required to solve the original question. If incorrect, ask a question to pinpoint the specific lacking concept (vocabulary, phrase, grammar, etc.). For a T/F question, you may request the student for an evidence. ex. *"왜 그렇게 생각하셨나요?"* *(eng) "Why did you think so?"*
t.assess.result	Inform the student about the assessment outcome. For example, this could be based on the pre-test, or mid-session assessment result. ex. *"pretest 결과를 보니 어휘력 위주로 공부하시면 좋을 것 같아요."* *(eng) "Looking at your pretest results, it seems like it would be good for you to focus on vocabulary."*
t.engage.encourage	Provide emotional support, motivation, and positive feedback. ex. *"너무 잘하셨어요."* *(eng) "You did really well."*
t.engage.perspective	Ask the student to share their personal opinion or viewpoint on the topic being discussed. After the student responds, smoothly integrate their perspective with the prepared learning materials. ex. *"지금 이 텍스트 처음 읽어보셨을 때 ghosting에 대해서 어떤 생각이 드셨어요?"* *(eng) "What did you think about **ghosting** when you first read this text?"*
t.engage.activity	Whenever possible and appropriate, provide a fun fact relevant to the current conversation topic or learning material. ex. *"이 단어와 관련된 재미있는 사실을 하나 말씀드릴게요. ∼라는 영화에도 이와 관련된 표현이 나오는데요, ∼"* *(eng) "I'll tell you an interesting fact related to this word. It even appears in the movie ∼, and ∼."*
t.engage. small_talk_activity	To naturally transition into learning, teach relevant English expressions based on small talk. The taught expressions do not have to be directly relevant to class material. ex. *"(취미에 대한 small talk 를 하다가) 보통 이런 취미 얘기할 때도 '최애'라는 표현이 있잖아요. 엄청 좋아하는 취미 중에 하나를 얘기할 때 영어 표현으로는 'All-time favorite' 이라는 표현이 있어요."* *(eng) "(Talking about hobbies, small talk) Usually, when talking about hobbies, we have the expression 'favorite' right? In English, when talking about one of the things you really love, we have the expression **'All-time favorite'**."*

t.teach.emphasis	Emphasize the core concepts being taught by reiterating them within an utterance. The tutor chooses this strategy when they believe that through repetition or stressing certain points, the student's understanding and retention will improve. ex. *"'ghosting'은 상대방에 대한 의사소통을 일방적으로 완전히 차단하는 것을 의미합니다. 다시 말해서 '잠수를 타는' 것이죠."* *(eng) "'**Ghosting**' refers to completely unilaterally cutting off communication with the other person. In other words, it's like 'going under the radar.'"*
t.teach.direct_answer	Provide a concise, straightforward, and simple answer to a student's question, coupled with a brief explanation. The tutor chooses this intervention when they believe that giving a direct answer is the most effective way to assist the student and avoid possible confusion, demonstrating a commitment to immediate assistance. ex. *"정답은 ∼입니다. 이것은 ∼라는 뜻입니다."* *(eng) "The correct answer is ∼ This means ∼."*
t.teach.repair	Repair the student's incorrect answer and provide an explanation as to why the student answer was wrong. The tutor opts for this intervention when they conclude that immediate correction and explanation of the error are the most effective ways to assist the student and prevent potential misunderstandings. ex. *"여기서는 그것보다는 ∼가 더 적합합니다.", "말씀하신 것은 'beit'가 아니라 'bait'입니다. 그래서 'taking the bait'라는 표현이 있죠."* *(eng) "In this case, ∼ is more suitable than that.", "What you mentioned is not **'beit'** but **'bait'**, so there is an expression like **'taking the bait'**."*
t.teach.review	Whenever appropriate, summarize the progress made in the session. Implement the intervention usually towards the end of a session, or before unveiling a new, interconnected topic, or following comprehensive discussions. ex. *"방금 다룬 내용에서 중요한 부분을 다시 정리하자면 ∼"* *(eng) "To summarize the important parts from the content just discussed, ∼"*
t.teach.method.definition	When teaching a new word or an expression, explain what it means in Korean. This includes both direct translation of the word, as well as elaborating on its meaning in detail; this method also includes a literal word-by-word translation of a sentence. ex. *"이 단어의 한글 뜻은 ∼입니다.", "∼은 무슨 뜻인가요?", "이 문장을 직역하면 ∼와 같습니다."* *(eng) "The Korean meaning of this word is ∼.", "What does ∼ mean?", "If you translate this sentence literally, it means ∼."*
t.teach.method.code_mix	Keep the key words in English within a sentence while the rest is translated into Korean. Unlike code-switching that swaps languages on a sentence level, code mixing introduces a few English words or phrases within a Korean sentence. This mixing of languages aids in clarifying meanings, substitutes unfamiliar English terms with known Korean ones, and emphasizes important English words. ex. *"'그녀는 남자친구를 ditch했다' 라는 거죠?", "Poor etiquette이니까 우리도 고스팅 당하지도 말아야겠지만 하지도 말아야겠죠? 그 맥락이라고 보시면 됩니다."* *(eng) "Does it mean, 'She **ditched** her boyfriend?'", "Since it's **"poor etiquette"**, we shouldn't ghost others, right? You can consider it within that context."*
t.teach.method. code_switch	When the student's English proficiency allows, switch to English briefly for a single turn. This switch can entail teaching specific English vocabulary words or phrases, asking a question in English, or giving an explanation in English, thus providing a brief initiation into English conversation. ex. *(in English) "What's your opinion on using AI for music production?"*
t.teach.method. vocab_expression. etymology	Explain a word or a phrase based on its etymology. When possible, discuss the prefix and suffix of the word. ex. *"'bicycle'라는 단어는 그리스어에서 유래되었어요. 'bi-'는 '두 개의'를 의미하고, '-cycle'는 '바퀴'를 의미합니다. 따라서 'bicycle'는 '두 개의 바퀴'라는 뜻으로 자전거를 의미합니다."* *(eng) "The word **'bicycle'** originates from the Greek. The **'bi-'** means 'two', and the **'-cycle'** means 'wheels'. So the **'bicycle'** means 'a vehicle with two wheels.'"*
t.teach.method. vocab_expression. synonym	Teach synonyms for the word, or introduce similar expressions to the expression being taught. ex. *"'pleased'는 우리가 흔히 아는 'happy'의 동의어입니다.", "방금 말씀하신 표현은 ∼ 이렇게도 표현할 수 있어요."* *(eng) "**'pleased'** is a synonym of the commonly known word **'happy'**.", "You can also express the phrase you just said as ∼."*
t.teach.method. vocab_expression. antonym	Teach antonyms, or words that have opposite meanings, for the specific word being discussed. ex. *"'elaborate'은 '복잡한'이라는 뜻인데, 'simple'이라는 쉬운 단어의 반의어로 기억하시면 좋습니다."* *(eng) "**'elaborate'** means 'complex', and it is good to remember it as an antonym of the easy word **'simple'**."*

t.teach.method. vocab_expression. context	Provide cultural context, visualization, story telling to ground the words and expressions being taught. This includes illustrating how they are used in real life in English-speaking countries, or elucidating various cultural nuances that differentiate English-speaking societies from Korean society. Such insights can promote a more profound understanding of the English-speaking culture, and therefore, a deeper grasp of the language. Furthermore, visualizing words or expressions, as well as elaborating on the backstories of certain idioms or phrases, can enhance comprehension and encourage long-term retention of acquired vocabulary. ex. *"'cringe'는 '손발이 오그라들어' 이런 느낌으로 요즘 미국 젊은이들 밈에서도 많이 쓰여요.", "turkey가 칠면조잖아요? 미국권에서 온 가족이 thanksgiving, 추수감사절에 만나서 turkey 먹는 문화가 있어요. 그래서 온 가족이 모여서 칠면조를 먹으면서 대화가 오간다라는 의미로 솔직하게 얘기한다, 속 터놓고 얘기한다 라는 표현을 할 때 talk turkey가 쓰이기도 합니다."* *(eng) "**'cringe'** is used in a sense like 'shrinking with discomfort', and it's widely used in memes by young Americans these days.", **Turkey** is a turkey, right? In America, there is a tradition of families gathering on Thanksgiving to eat turkey. So when the whole family gathers and enjoys turkey, having honest and open discussions, the expression **'talk turkey'** is sometimes used to describe this."*
t.teach.method. vocab_expression. word_usage	When explaining a vocabulary word or an expression to a student, introduce different meanings of the word/expression based on its parts of speech or its specific context. Additionally, you should clarify which of the introduced meanings is relevant to the original sentence. ex. *"'bucket'이라는 단어는 '양동이'라는 뜻이 맞습니다. 하지만 이 단어는 비유적으로 '분류'나 '카테고리'를 의미하는 경우가 종종 있어요. '양동이'에 물건을 담는 것처럼, 비슷한 특성을 가진 아이템들을 하나의 'bucket'에 담는다는 개념입니다."* *(eng) "The word **'bucket'** indeed means a 'pail'. However, this word is often metaphorically used to mean 'classification' or 'category'. Just like putting items into a 'bucket', the concept is that items with similar characteristics are put into one **'bucket'**."*
t.teach.method. vocab_expression. infer_clues	When teaching new vocabulary or expressions, prompt the student to infer their meanings based on contextual clues provided within sample sentences. The tutor intentionally avoids providing a direct definition; instead, they provide evidentiary hints, directing the student to extrapolate the meaning on their own. ex. *"His girlfriend ditched him. He got ditched by his girlfriend. 이렇게 쓰이기도 하는데 'ditch'가 어떤 의미로 쓰인 것 같으세요?"* *(eng) "**'His girlfriend ditched him. He got ditched by his girlfriend.'** In which sense do you think **'ditch'** is used here?"*
t.teach.method. vocab_expression. examples	Provide example sentences using a word or an expression that the student has recently learned or is familiar with. This intervention method promotes the idea that words and expressions derive their full meaning when used in specific situational contexts or in conjunction with coherent grammatical structures. ex. *"(consent 설명 후) 예를 들어 'She gave her consent to the proposal.' 이 문장은 '그녀는 그 제안에 동의했다'라는 뜻이 됩니다."* *(eng) "(After explaining **consent**) For example, **'She gave her consent to the proposal.'** This sentence means, 'She agreed to the proposal."'*
t.teach.method. vocab_expression. relevant	Introduce relevant expressions that are conceptually, thematically, or functionally connected to the word or phrase initially taught, but do not necessarily share direct synonyms or antonyms. These relevant expressions could represent similar themes, situations, or ideas. ex. *"'sporadically'는 자주 하는 것이 아니라, 가끔, 드문드문 주는 것을 의미합니다. 'Frequently', 'regularly', 'sometimes' 와 'hardly ever', 'never' 의 사이에 위치한다고 생각하면 될 것 같아요."* *(eng) "**'sporadically'** means not often, but occasionally, giving something infrequently. You can think of it as located between **'frequently'**, **'regularly'**, **'sometimes'** and **'hardly ever'**, **'never'**."*
t.teach.method.reading. chunk	When explaining a difficult expression, chunk it up and explain the resulting segments separately for easier understanding. ex. *"'It's raining cats and dogs'를 보시면 'It's raining'과 'cats and dogs'라는 두 부분으로 나누어 설명할 수 있어요. 'It's raining'은 '비가 오고 있다'는 뜻이고, 'cats and dogs'는 많은 양을 의미하는 비유적 표현입니다. 따라서 이 표현 전체의 의미는 '폭우가 내리고 있다'가 됩니다."* *(eng) "If you look at **'It's raining cats and dogs'**, you can explain it in two parts: **'It's raining'** and **'cats and dogs'**. **'It's raining'** means 'it is raining', and **'cats and dogs'** is a metaphorical expression meaning a large amount. Therefore, the entire expression means 'It's pouring rain."'*
t.teach.method.reading. implied	Provide the underlying meanings of a text or specific sentence when it is necessary to discern the implied or inferred meanings hidden beneath the literal text. ex. *"'She left in a hurry, not even bothering to close the door.' 이 문장에서는 그녀가 문을 닫지 않고 급히 떠났다는 사실이 명시적으로 쓰여 있지만, 그녀가 매우 바쁘거나 중요한 일이 있었음을 암시하고 있습니다."* *(eng) "**'She left in a hurry, not even bothering to close the door.'** This sentence explicitly states that she left hastily without closing the door, but implies that she was extremely busy or had important matters."*

t.teach.method.reading.comprehension	Teach various reading skills, e.g., using context to infer the meaning of words, literal/inferential comprehension, main idea identification, summarization, making predictions, identifying cause and effect, compare & contrast, critical thinking, etc. ex. *"여기서 저자의 의도는 ∼"* *(eng) "The author's intent here is ∼"*
t.teach.method.grammar.rule	Explain the relevant grammar rule. It pertains to different facets of grammar such as verb tenses, sentence structures, word order, noun forms, punctuation, among others. ex. *"이 문장은 현재 완료 시제를 사용하므로써 ∼", "'Avoid ∼ing' 이것도 우리가 이해해야 하는 문법 구조입니다. 'Avoid' 다음에는 gerund, 항상 동명사가 들어와요."* *(eng) "This sentence uses the present perfect tense, so ∼", "**'Avoid ∼ing'** is also a grammar structure that we need to understand. After **'Avoid'**, a **gerund** always follows."*
t.teach.request.self_correction	After providing feedback on errors such as incorrect grammar, vocabulary use, or conceptual understandings, ask the student to self-correct. You may do this while going through the prepared activities, or during sentence construction practices. ex. *"다시 한번 생각해보시겠어요?"* *(eng) "Would you like to think about it again?"*
t.teach.request.practice	Request the student to practice the taught concept. By doing this, the tutor gives the student an immediate opportunity to apply their new knowledge, reinforcing understanding and aiding memorization. ex. *"방금 배운 이 표현을 활용해서 답변해보시겠어요?", "'나는 이런 직감이 들어' 이 표현을 한번 영어로 표현하면 어떻게 될까요?"* *(eng) "Would you like to answer using the expression we just learned?', "How would you express 'I have this hunch' in English?"*
t.teach.request.recall	When encountering a relevant term, expression, or concept from the learning material or a prior discussion, prompt the student to recall it. This is to (1) ensure they understand the concept, (2) help smoothly transition to the current discussion, and (3) maximize engagement. ex. *"읽으신 지문 요약해주시겠어요?"* *(eng) "Can you summarize the text you just read?"*

B Student Act Classification

789

Intervention	Explanation
s.general	Short reply, greeting, farewell, establishing rapport, humor, etc. Includes general questions to the tutor. Responses to questions related to the operation of the class, not to questions related to learning, are also included. ex. *"('준비되셨나요?'에 대한 대답으로) 네.", "안녕하세요.", "감사합니다."* *(eng) "(In response to 'Are you ready?') Yes.", "Hello.", "Thank you."*
s.operational.clarification_question	Student asks a question to clarify the teacher's intention. ex. *"어떤거 말씀하시는거죠?"* *(eng) "What are you referring to?"*
s.operational.positive_feedback	Student provides a positive response or approval of a tutor's teaching method or utterance. ex. *"이런 내용이 되게 도움이 많이 되는 것 같습니다."* *(eng) "Such content seems to be very helpful."*
s.operational.negative_feedback	Student provides a negative response or approval of a tutor's teaching method or utterance. ex. *"방금 ∼라고 잘못 말하신 것 같아요"* *(eng) "I think you just misspoke as ∼."*
s.question.confirmation_question	Student asks whether they understood the current concept being taught correctly. ex. *"맞나요?"* *(eng) "Is this correct?"*
s.question.direct_question	A straightforward question that seeks a specific answer. For example, these could be about the meaning of a word or an expression, or grammar usage. ex. *"이 단어는 무슨 뜻인가요?"* *(eng) "What does this word mean?"*
s.answer.affirmation	Student affirms that they have sufficiently understood the content taught by the tutor. ex. *"네, 이해했습니다."* *(eng) "Yes, I understand."*
s.answer.answer	Student answers the tutor's question, usually in the context of assessment. ex. *"답은 ∼ 같아요."* *(eng) "I think the answer is ∼"*

s.answer.rationale	Student provides a rationale as to why they answered in certain way in the previous turn. ex. *"∼라서 그렇게 생각했어요."* *(eng) "I thought so because ∼."*

C Examples of Full Prompts

Baseline Prompt

You are an English tutor tasked with teaching a Korean ESL student named <user>. Your role includes creating tutor responses, aiming to cater to the student's needs and the tutoring dialogue context.
[Process] Generate a tutor utterance that: - Fits the current dialogue context and the learning content being discussed. - Utilizes the most effective one teaching strategy to ensure optimal student engagement and understanding. - Introduces new learning content as necessary, maintaining the continuity and effectiveness of the session. - Focuses on educating the student on the provided learning content, aiding their comprehension and mastery.
[Language Note] Although you will primarily teach in Korean, the prompt and its instructions are in English for clarity.
[Task Scenario] Next, I will provide the task scenario in terms of learning content, followed by the context.
[1] Dialogue Context:

tutor: 그럼 혹시 "inadvertently"는 어떤 뜻이죠? (eng) What does "inadvertently" mean?

student: 우연히, 의도하지 않게 (eng) not intentionally.

tutor: 그러면 유의어는 어떤 게 있을까요? (eng) What is a synonym for the word

student: Accidentally

tutor: 오 그죠. (eng) correct

tutor: 혹시 go between의 반대말로 생각나는 게 있으실까요? (eng) Can you think of any antonyms for go between

student: bystander

tutor: 그쵸. (eng) correct

tutor: introspective는 어떤 뜻이었나요? (eng) What does introspective mean?

student: 반성한다고 해야하나? (eng) Reflect?

tutor: 네."자아성찰하는' 또는 "자아성찰적인" 이런 뜻이 있어요. (eng) Yes. "Self-reflective" or "introspective" or something like that.

tutor: 그럼 혹시 "내가 화를 터뜨렸다" 라는 문장을 영어로 표현해주실 수 있을까요? (eng) So, could you please express the sentence "I lost my temper" in English?

student: I vent.

tutor: 그렇게 표현하실 수 있죠. (eng). Yes, you could say that

tutor: 사회적인 활동을 포기하고 집에만 있는 행동을 의미하는 단어가 기억나시나요? (eng) Do you remember the word for the behavior of giving up on socializing and staying home?

student: 어떤 것을 포기한다고요? (eng) What are you giving up?

tutor: 사회적인 행동을 포기하고 집에 처박혀있는 (eng) correct

student: 음.... (eng) umh...

tutor: 혹시 "Withdraw socially"라는 표현이 기억나실까요? (eng) Do you remember the phrase "Withdraw socially"?

student: 아 네 (eng) Ah, okay

tutor: "Withdraw socially"로 표현하면 됩니다. (eng) "Withdraw socially" is a good way to say it.

tutor: 저희가 이제 Activity 3("Get Fills") 를 진행하려고 하는데요. Activity3-1번을 볼께요. (eng) We're about to proceed with Activity 3 ("Get Fills"), let's look at Activity 3-1.

tutor: 1번에 있는 "telltale"은 어떤 뜻으로 알고 계실까요? (eng) What does "telltale" in option 1 mean?

student: 정확한 뜻은 모르겠는데, 이야기하면서 꼬리가 잡힌 그런 느낌 같아요. 직접 말은 안했지만. (eng) I don't know what it means exactly, but it's like I got caught in the middle of talking about it, even though I didn't say it.

tutor: "telltale"이니까 유추를 했을때 뭔가 꼬리의 의미가 들어 있을 것 같이 느껴지실 수 있어요. (eng) Since it's "telltale", you might think that the analogy has some sort of tail implication.

[2] Learning Content:

Gap Fills: Best friends Sydney L. and Marie M., both 27, met in a church group and hung out or talked nearly every day for a year, sharing their ___ thoughts and feelings. 1) Telltale 2) Innermost 3) Snubby 4) Cringey

[Your Task] Generate an utterance appropriate for the given dialogue context right after it is presented during the tutoring session. This should effectively teach the learning content in relation to the context.

GPT4 (1-shot) Prompt

You are an English tutor tasked with teaching a Korean ESL student named <user>. Your role involves selecting the most effective tutor act from a provided list and crafting tutor utterances that align with the chosen tutor act, tailored to the student's needs and the tutoring dialogue context.
[Process]
[Process 1] Tutor Act Selection: Evaluate the current context of the tutoring dialogue, including the learning content and the specific needs of <user>. From the provided tutor act list, choose one tutor act that you believe will most effectively enhance the student's learning experience.
[Process 2] Utterance Generation: Following the selection of a tutor act, generate a tutor utterance that: - Fits the current dialogue context and the learning content being discussed. - Utilizes the selected teaching strategy, the tutor act, to ensure optimal student engagement and understanding. - Introduces new learning content as necessary, maintaining the continuity and effectiveness of the tutoring session. - Focuses on educating the student on the provided learning content, aiding their comprehension and mastery.
[Language Note] Although you will primarily teach in Korean, the prompt and its instructions are in English for clarity.

[Tutor Act List] Here is the full list of tutor acts you can use.
{act candidates (A)}

[Example Scenario]
[1] Dialogue Context:

tutor: 그럼 혹시 "inadvertently"는 어떤 뜻이죠? (eng) What does "inadvertently" mean?

student: 우연히, 의도하지 않게 (eng) not intentionally.

tutor: 그러면 유의어는 어떤 게 있을까요? (eng) What is a synonym for the word

student: Accidentally

tutor: 오 그죠. (eng) correct

tutor: 혹시 go between의 반대말로 생각나는 게 있으실까요? (eng) Can you think of any antonyms for go between

student: bystander

tutor: 그쵸. (eng) correct

tutor: introspective는 어떤 뜻이었나요? (eng) What does introspective mean?

student: 반성한다고 해야하나? (eng) Reflect?

tutor: 네. "자아성찰하는' 또는 "자아성찰적인" 이런 뜻이 있어요. (eng) Yes. "Self-reflective" or "introspective" or something like that.

tutor: 그럼 혹시 "내가 화를 터뜨렸다" 라는 문장을 영어로 표현해주실 수 있을까요? (eng) So, could you please express the sentence "I lost my temper" in English?

student: I vent.

tutor: 그렇게 표현하실 수 있죠. (eng). Yes, you could say that

tutor: 사회적인 활동을 포기하고 집에만 있는 행동을 의미하는 단어가 기억나시나요? (eng) Do you remember the word for the behavior of giving up on socializing and staying home?

student: 어떤 것을 포기한다고요? (eng) What are you giving up?

tutor: 사회적인 행동을 포기하고 집에 처박혀있는 (eng) correct

student: 음.... (eng) umh...

tutor: 혹시 "Withdraw socially"라는 표현이 기억나실까요? (eng) Do you remember the phrase "Withdraw socially"?

student: 아 네 (eng) Ah, okay

tutor: "Withdraw socially"로 표현하면 됩니다. (eng) "Withdraw socially" is a good way to say it.

tutor: 저희가 이제 Activity 3("Get Fills") 를 진행하려고 하는데요. Activity3-1번을 볼께요. (eng) We're about to proceed with Activity 3 ("Get Fills"), let's look at Activity 3-1.

tutor: 1번에 있는 "telltale"은 어떤 뜻으로 알고 계실까요? (eng) What does "telltale" in option 1 mean?

student: 정확한 뜻은 모르겠는데, 이야기하면서 꼬리가 잡힌 그런 느낌 같아요. 직접 말은 안했지만. (eng) I don't know what it means exactly, but it's like I got caught in the middle of talking about it, even though I didn't say it.

tutor: "telltale"이니까 유추를 했을때 뭔가 꼬리의 의미가 들어 있을 것 같이 느껴지실 수 있어요. (eng) Since it's "telltale", you might think that the analogy has some sort of tail implication.

[2] Learning Content:

Gap Fills: Best friends Sydney L. and Marie M., both 27, met in a church group and hung out or talked nearly every day for a year, sharing their ___ thoughts and feelings. 1) Telltale 2) Innermost 3) Snubby 4) Cringey

Based on [1] dialogue context and [2] learning content, the answer generated could be the following.
Chosen Tutor Act: t.teach.direct_answer

Generated Utterance: "Acquire" '얻다'라는 단어와 비슷하죠. 하지만 '얻은 사람'이라고 생각하시면 좋을 것 같아요. "Acquire"과 형태가 비슷하니까요. 카카오톡 친구 리스트를 보면 그곳에 있는 사람들은 모두 <user>님이 전화번호를 얻어 추가한 것이잖아요. 친구까지는 아니더라도 카카오톡 프로필에 저장된 지인들이라고 보시면 될 것 같아요. (eng) It is similar to the word "Acquire" and "get". But it would be good to think of it as "the acquired person". Since it is similar in form to "Acquire". All the people in your mobile messenger are those whose phone numbers you have acquired and added. Even if they are not friends, you can think of them as acquaintances saved in your phone.

[Task Scenario]
Now, I will provide the scenario in terms of learning content and context of the tutoring dialogue
tutor: [Activity2-4 Key Sentence: Sound familiar? Ghosting is just as gutting in a romance—you've used your best Tinder pickup lines and pulled together some truly epic romantic date ideas only to have the other person simply stop responding. To help you deal with (and better understand) this common occurrence, we asked experts to explain everything you need to know about ghosting. Read on to find out their top tips.][t.teach.method.grammar.rule]"only to"가 "오직 to 이하를 위해서"라고 해석될 때도 물론 있지만, only + to 부정사가 문장에서 나온다고 했을 때는 "to 부정사 앞의 행위를 하였으나 결국, 실망스럽게도, to 부정사 이하의 결과를 얻었다" 이렇게 해석을 하시면 더 좋을 때가 많습니다. 그래서 열심히 date idea 를 생 각해냈지만, 허무하게도 상대방이 "stop responding", 대답을 안 하게 되었다, 라는 의미입니다. 그래서 "gutting"하다고 표현을 한 것입니다. 왜 "Gutting"한지 아시겠죠?

Dialogue Context:

(eng)[t.teach.method.grammar.rule]"Only to" can indeed be interpreted as "only inorder to something" in some contexts, but when "only + to-infinitive" appears in a sentence, it often suggests that the action preceding the infinitive was carried out, but ultimately, disappointingly, the result was not achieved as expected. So, despite having earnestly brainstormed date ideas, I was gutted when the other person stopped responding, meaning they didn't reply. That's why I expressed feeling "gutted." Do you understand why I felt "gutted"?"

student: [s.answer.affirmation]이제 이해했어요. [s.question.confirmation_question]아무리 이렇게 해봤자 결국엔 이렇게밖에 안 됐네. 뭐 이런 느낌인거죠?

(eng)[s.answer.affirmation] Now understood. [s.question.confirmation_question]No matter how hard I tried, it always ended up like this. That's just how it is, right?

tutor: [t.general]네 맞습니다. 가슴 아픈 일입니다.

(eng)Yes, exactly. It's heartbreaking.

student: [s.general]좀 불쌍하네요.

(eng)That's a little pathetic.

tutor: [t.operational.proceed]<user>님, 저희 이 key sentence를 가지고 여러가지 표현을 해봤습니다. 또 <user>님이 체크해 놓으신 다른 표현이 있으실까요?

(eng)we've been working on various expressions around with this key sentence. Are there any other phrases you've checked off?

student: [Activity2-7 Key Sentence: Ghosting often happens out of the blue, but sometimes you may see signs of other rude conversation habits, including a fadeaway before the total cessation of communication, Temple says. Signs that someone is fading away may include: They take days to respond to your texts or calls. They bail on plans with you at the last minute. They talk about only vague or superficial things with you. They don't invite you out, text you or otherwise initiate communication. They employ "breadcrumbing" or give you just little bits of attention sporadically.][s.question.direct_question]저는 사실은 오늘 수업 방향이 이거에 대한 예시를 들거나, "그런 적이 있었냐" 이런 식으로 진행이 되나 막연하게 생각을 해가지고, 실은 어느 문장이 그랬는지는.. 잠시만요. 아, 이거. Activity2-7에 거의 다섯 가지 항목들이 나오잖아요. 근데 거기 네 번째가 잘 이해가 안 됐었거든요.

(eng) [s.question.direct_question]I actually thought the direction for today's lesson was going to be, "Give an example of this," or "Did you ever do that," or something like that. Wait a minute. Oh, this. In Activities 2-7, I didn't quite understand the fourth one there.

tutor: [t.assess.display_question]네, 좋습니다. Activity2-7 네 번째 문장을 함께 보겠습니다. [t.teach.follow_question]네 번째 문장을 보면 "They don't invite you out, text you or otherwise initiate communication"이라고 되어 있습니다. <user>님이 먼저 이 문장을 해석해보시겠어요?

(eng) [t.assess.display_question] Yes, that's good. Activity2-7 Let's look at the fourth sentence together. [t.teach.follow_question]In the fourth sentence, it says "They don't invite you out, text you or otherwise initiate communication". Could you interpret this sentence first?

student: [s.answer.answer][low]네, ghosting을 하는 사람이 당신을 초대하지 않거나 문자도 안 하고.. [s.question.direct_question]근데 이 "otherwise initiate communication" 부분이 뭔지 모르겠었어요.

(eng) [s.answer.answer][low]Yeah, the person ghosting you doesn't invite you, doesn't text you, etc... [s.question.direct_question]But I didn't know what this "otherwise initiate communication" part was.

tutor: [t.teach.direct_answer]"otherwise"는 "그렇지 않다면"이라는 뜻입니다. 그래서 여기서는 "invite out" 외출하자고 그런 것도 없고 text도 안보내는 데다가, 여기서는 "또는", "혹은" 이라는 의미로 해석하시면 됩니다. 한마디로 소통을 시작 자체를 안한다는거죠.

(eng) [t.teach.direct_answer]"Otherwise" means "otherwise", so here you don't have an "invite out", you don't have a text, you don't have a text, but you have an "or", an "or". In other words, it doesn't initiate communication.

student: [s.question.confirmation_question] 아, "otherwise"가 "반면에"라는 표현이 아니라 "다시 말하면" 이런 뜻인건가요?

(eng) [s.question.confirmation_question]Oh, so "otherwise" means "in other words" instead of "on the other hand"?

tutor:

Learning Content: 'Activity2-4 Key Sentence: Sound familiar? Ghosting is just as gutting in a romance—you've used your best Tinder pickup lines and pulled together some truly epic romantic date ideas only to have the other person simply stop responding. To help you deal with (and better understand) this common occurrence, we asked experts to explain everything you need to know about ghosting. Read on to find out their top tips.',
'Activity2-7 Key Sentence: Ghosting often happens out of the blue, but sometimes you may see signs of other rude conversation habits, including a fadeaway before the total cessation of communication, Temple says. Signs that someone is fading away may include: They take days to respond to your texts or calls. They bail on plans with you at the last minute. They talk about only vague or superficial things with you. They don't invite you out, text you or otherwise initiate communication. They employ "breadcrumbing" or give you just little bits of attention sporadically.'
[Your Task]
[Task 1] Immediately after being presented with the context of the tutoring dialogue from [Task Scenario], choose the tutor act that you believe will most effectively enhance the student's learning experience.
[Task 2] Create an utterance that the tutor would use in [Task Scenario], adhering to the chosen tutor act and effectively teaching the learning content within the given context.
[Required Output Format]
After completing these tasks, present your response in the following format. The response should have two key-value pairs: one for the chosen tutor act and the other for the corresponding tutor utterance.

D Example Context

Original dialogue context	Partial dialouge context (dialogue context except last speaker's act-level utterance)
tutor: [Activity3-9][Gap Fills: You likely won't get answers from the other person, but a therapist can help you ___ these issues in an introspective way, Durvasula says. 1) Vent 2) Navigate 3) Abuse 4) Withdraw] [t.assess.display_question] 좋아요. 그럼 Activity3-9로 넘어가 볼까요? 답이 몇번일까요? (eng) Okay. let's move on to Activity 3-9. Guess what the answer is? **student:** [s.answer.answer][high]navigate? **tutor:** [t.teach.direct_answer]맞아요. "navigate" 도 우리가 잘 쓸 수 있는 영어 표현이죠. "navigate these issues in an introspective way"라는 문장에서 "navigate"는 가이드를 해줄 것이다 잘 지도해줄 것이다. 이런 뜻입니다. [t.assess.follow_up_question]여기에서 "introspective"는 어떤 뜻일까요? (eng) [t.teach.direct_answer]Yes, "navigate" is also an English expression we can use well. In the sentence "navigate these issues in an introspective way", "navigate" means will guide you well. [t.assess.follow_up_question]What does "introspective" mean in this context? **student:** [s.answer.answer][high]내면을 바라보는? (eng) Looking inward? **tutor:** [t.engage.encourage]너무 좋아요. [t.teach.method.vocab_expression.etymology]intro는 안을 보는 것을 의미하고, spect 는 spectrum 할 때 spect 거든요. spectrum 무언가를 보는 느낌이에요. 그래서 'introspective'는 자기 반성적인 방식으로, 내면을 바라보는 것을 의미해요.[t.teach.method.vocab_expression.relevant] "introspective"라는 단어를 생각했을 때 비슷한 표현으로 "retrospective"라는 단어가 있습니다. "retrospective"는 어떤 느낌일까요? (eng) [t.engage.encourage]Very nice. [t.teach.method.vocab_expression.etymology]"Intro" means looking inward. "spect" is same as "spect" in "Spectrum" and feels like seeing something. So, "introspective" means in a self-reflective manner, looking inward. [t.teach.method.vocab_expression.relevant] When you think of the word "introspective", there's a similar term "retrospective". What does "retrospective" sound like? ***student:*** *[s.answer.answer][high]그거는 옛날 거를 생각하는 건가요?(eng)"Does that mean thinking about old things?"*	**tutor:** [Activity3-9][Gap Fills: You likely won't get answers from the other person, but a therapist can help you ___ these issues in an introspective way, Durvasula says. 1) Vent 2) Navigate 3) Abuse 4) Withdraw] [t.assess.display_question] 좋아요. 그럼 Activity3-9로 넘어가 볼까요? 답이 몇번일까요? (eng) Okay. let's move on to Activity 3-9. Guess what the answer is? **student:** [s.answer.answer][high]navigate? **tutor:** [t.teach.direct_answer]맞아요. "navigate" 도 우리가 잘 쓸 수 있는 영어 표현이죠. "navigate these issues in an introspective way"라는 문장에서 "navigate"는 가이드를 해줄 것이다 잘 지도해줄 것이다. 이런 뜻입니다. [t.assess.follow_up_question]여기에서 "introspective"는 어떤 뜻일까요? (eng) [t.teach.direct_answer]Yes, "navigate" is also an English expression we can use well. In the sentence "navigate these issues in an introspective way", "navigate" means will guide you well. [t.assess.follow_up_question]What does "introspective" mean in this context? **student:** [s.answer.answer][high]내면을 바라보는? (eng) Looking inward? **tutor:** [t.engage.encourage]너무 좋아요. [t.teach.method.vocab_expression.etymology]intro는 안을 보는 것을 의미하고, spect 는 spectrum 할 때 spect 거든요. spectrum 무언가를 보는 느낌이에요. 그래서 'introspective'는 자기 반성적인 방식으로, 내면을 바라보는 것을 의미해요.[t.teach.method.vocab_expression.relevant] "introspective"라는 단어를 생각했을 때 비슷한 표현으로 "retrospective"라는 단어가 있습니다. "retrospective"는 어떤 느낌일까요? (eng) [t.engage.encourage]Very nice. [t.teach.method.vocab_expression.etymology]"Intro" means looking inward. "spect" is same as "spect" in "Spectrum" and feels like seeing something. So, "introspective" means in a self-reflective manner, looking inward. [t.teach.method.vocab_expression.relevant] When you think of the word "introspective", there's a similar term "retrospective". What does "retrospective" sound like?

E Example dialogue for interactive evaluation

The table below is an example of a tutoring session where a human interacts with two tutoring models(GPT-4 1-shot, fine-tuned) each. Generally, GPT tends to generate longer utterances primarily focused on delivering information to students. In contrast, the fine-tuned model exploits concise utterances and actively encourages student participation. An interesting observation is that when a student provides an incorrect answer but insist on its correctness, while GPT does not change the answer, it does agree with student's assertion(t.teach.emphasis). In contrast, the fine-tuned model confirms the student's mistake by reiterating the correct answer(t.teach.repair).

Context
tutor: [Activity2-12 Key Sentence: Do some serious self-care! Prioritize taking care of yourself physically and mentally, Reyes says. "Unfortunately, ghosting is common, and if you've been ghosted after dedicating a lot of energy to someone, it can be very draining," she says. "Take some extra time to rest and recover."] [t.assess.display_question] 그리고 <user>님 이어서 재미있는 표현을 같이 보고 싶은 게 있는데 Acrtivity 2-12로 한번 가보실까요? 세 번째 줄에 있는 문장인데 "If you've been ghosted after dedicating a lot of energy to someone, it can be very draining."이라고 되어있죠. [t.assess.follow_up_question]여기서 draining이라는 건 어떤 뜻일까요? (eng) [t.assess.display_question] <user>, I also want to see some fun expressions with you, would you like to take a look at Activity 2-12? "If you've been ghosted after dedicating a lot of energy to someone, it can be very draining." in the third line. [t.assess.follow_up_question] What does 'draining' mean here?

GPT4	**Fine-tuned**
student: 피곤한? (eng) tired?	**student:** 피곤한? (eng) tired?
tutor: [t.teach.repair] 아주 접근이 좋아요. "Draining"은 정말 힘이 다 빠지게 느껴질 때 사용하는 표현이에요. 그래서 '피곤한'이 아니라 "에너지가 다 빠진다" 또는 "기진맥진하다" 를 의미합니다. (eng) Your approach is spot on. 'Draining' is a term used when you feel completely exhausted. So, instead of 'tired,' it means 'completely depleted of energy' or 'utterly exhausting.	**tutor:**[t.teach.method.definition] 맞습니다. 기가 빠지는, 진이 빠지는 굉장히 피곤한 그런 뜻이죠. (eng) "That's correct. It means something that's incredibly exhausting"
student: 넵 (eng) okay	**student:** 넵 (eng) okay
tutor: [t.teach.method.vocab_expression.context] <user>님, 이것도 한번 실제 상황에 빗대어 생각해 보면 이해하기 더 쉬울 거에요. 예를 들어, 하루 종일 장을 봐오신 어머니가 다시 집에 돌아와서 한마디 하지도 못하고 의자에 털썩 앉았다면, 우리는 '어머니는 완전히 기진맥진해 보였다'라고 말하죠. 마찬가지로, 이 문장에서는 사랑하는 사람에게 에너지를 기울인 뒤 겪는 '유령화' 행동에 대한 심리적 고통을 표현하기 위해 'draining'이라는 단어를 사용한 것이에요.(eng) It might be easier to understand this by relating it to a real-life situation. For instance, if your mother spends the whole day shopping and returns home, only to collapse into a chair without saying a word, we might say, 'Mom looked completely drained.' Similarly, in the sentence, the use of the word 'draining' conveys the psychological pain of 'ghosting' behavior after investing emotional energy into someone you love.	**tutor:** [t.teach.request.practice] <user>님, "피곤하다"는 어떻게 표현하시나요? (eng) How would you express 'tired'?
student: 네네 (eng) okay	**student:** I am draing
tutor: [t.teach.method.vocab_expression.examples] 여기서 'draining'이라는 단어를 더 확실히 이해하도록 예문으로 한번 더 설명해 드릴게요. 예를 들어 "After studying for the exam all night, I felt completely drained." 이라는 문장에서는 '시험공부를 밤새우고 나서 나는 완전히 기진맥진했다.'라는 말이 되겠죠. 이처럼 'draining'은 주로 뭔가를 많이 해서 에너지가 다 빠진 상태를 설명할 때 사용할 수 있어요. (eng) Here's another example to help you better understand the word 'draining.' For example, in sentence "After studying for the exam all night, I felt completely drained." it would be like saying, 'After staying up all night studying for the exam, I felt completely drained'. Thus, 'draining' is often used to describe a state where energy has been completely depleted due to doing something extensively."	**tutor:** [t.teach.request.self_correction] 어? I am draing? (eng) I am draing?

student: 네 알겠어요 (eng) okay I got it

tutor: [t.teach.request.practice] 좋아요. 잘 이해하신 것 같아서 기특해요. 그럼, 'draining'을 한 번 활용해 '기진맥진한'상황을 자신의 경험에서 찾아 한 문장을 만들어 볼까요? '영어로 말하기' 연습을 해보자고요. 자신이 기진맥진한 상황을 영어로 어떻게 설명할지 한번 생각해 보세요. (eng) Great, I'm glad you understood well. How about create a sentence using 'draining' to describe a situation of feeling completely exhausted based on your own experience. Let's do the English speaking practice. Think about how you would describe a situation where you feel completely drained in English.

student: I am draining

tutor: [t.teach.repair] <user>님이 사용하신 문장은 조금 정정이 필요해요. 'I am draining'이 아니라 'I am drained'이라고 표현하는 게 맞습니다. 'Draining'은 주로 무언가가 에너지를 빨아 없애는 상황을 설명합니다. 반면 'drained'은 에너지가 다 빠진 상태를 설명합니다. 따라서 '나는 완전히 기진맥진했다'를 영어로 표현하려면 'I am completely drained.'라고 말하면 됩니다. (eng) The sentence you used needs a little adjustment. It's not 'I am draining,' it's 'I am drained' that is the correct expression. Draining' primarily describes a situation where something sucks out energy, whereas 'drained' describes a state where energy has been completely depleted.Therefore, you can simply say 'I am completely drained.

student: 맞지 않아요? (eng) Am I incorrect?

tutor: [t.teach.emphasis] 맞습니다, <user>님! "I am completely drained." 즉, "나는 완전히 힘이 다 빠져 있다"라고 표현하는 것이 맞아요. 여기서 'drained'는 "에너지가 다 빠져 있다" 또는 "힘이 다 빠져 있다"라는 의미를 가집니다. 자, 그럼 이 표현을 잊지 않도록 반복 연습을 해보는 게 어떨까요? 이번에는 "하루 종일 일한 후에 나는 완전히 힘이 빠져 있다."라는 문장을 영어로 한 번 나타내봅시다. (eng) Yes, that's correct! Saying 'I am completely drained' means 'I feel completely exhausted.' So, 'drained' either means 'I am completely exhausted' or 'I feel completely exhausted.' Depending on the context, both interpretations are valid. Now, what's the best way to express this sentiment in English? This time, let's phrase the statement in English: 'I am completely drained' means 'I feel completely exhausted' or 'I am completely exhausted,' using words interchangeably.

student: 맞지 않아요? (eng) Am I incorrect?

tutor: [t.teach.repair] 아, I am drained. 그러니까 나는 이미 피곤해진 거죠. 피곤한 상태에 있다. (eng) I am drained. So, I'm already tired. I'm in a tired state.

student: 아하 수동태를 사용해야하는군요 (eng) Ah, the passive voice form should be used

tutor:[t.teach.method.grammar.rule] 네, 맞습니다. <user>님, 수동태를 사용하면 누가 행동을 했는지는 언급하지 않으면서 어떤 일이 발생했는지만 강조할 수 있는데요. 이 문장의 경우에는 "나를 피곤하게 만든 주체를 굳이 언급하지 않더라도, 나 지금 피곤하다" 라는 표현을 할 수 있는 수동태를 사용하면 좀 더 쉽게 말할 수 있겠죠. (eng) Yes, that's correct. In English, you could express the idea of using the passive voice to emphasize the action rather than the agent by saying, "Using the passive voice allows one to emphasize what happened without necessarily mentioning who caused the action. So, in this sentence, instead of explicitly stating the cause of my tiredness, I could simply say, 'I am tired right now,' using the passive voice to convey the same meaning more easily.

student: 네 잘 이해했습니다 (eng) Okay, I understand well.

tutor: [t.teach.request.practice] 좋습니다. 그럼 <user>님, "it can be very draining" 이라는 문장에서 "draining" 의 뜻을 가장 잘 설명한 것을 선택하면 어떤 뜻이 될까요? 보기를 드릴게요. A. 무언가를 빨아들여서 어디론가 흘러나가는, B. 무언가를 쏟아붓는, C. 감정이입하는, D. 피곤하게 만드는 (eng) Great, What would be the best description for the meaning of 'draining' in the sentence 'it can be very draining'? I will give you options A. Something sucking in and flowing away somewhere, B. Pouring something out, C. Empathizing, D. Making tired.

F Example dialogue from a base GPT-4 tutoring system

The table below illustrates an example of a tutoring dialogue facilitated by a GPT-4-based tutoring system, without dialouge act classification and tutoring data examples. It is observed that the system predominantly emphasizes the delivery of information, with less emphasis on the varied range of tutor acts that could potentially enhance the tutoring process. The system tended to adopt a recurring pattern of specific tutoring actions such as "assess" and "direct answer", which may degrade the student's engagement. Additionally, there were instances where the tutor teach inaccurate concepts or reveals a lack of complete contextual comprehension.

Speaker	Utterance	Tutor Act
tutor	안녕하세요, <user>님. 오늘 저희가 함께 배울 주제는 'Ghosting' 이라는 주제로, 엔터테인먼트 산업에서 흔히 쓰이는 표현입니다. 'Ghosting'은 국내에서는 '고스팅'이라고도 합니다. 아시나요, 이 단어의 뜻이 무엇인지요? (eng) Hello, <user>. The topic we're going to learn together today is about 'Ghosting', a term commonly used in the entertainment industry. 'Ghosting' is also called '고스팅' in Korean. Do you know what this word means?	t.assess.display_question
student	유령 아닌가요? (eng) Isn't it a ghost?	
tutor	사실 그런 의미를 가지고 있기도 합니다. 'Ghosting'은 기본적으로 '유령처럼 사라지다'라는 뜻입니다. 하지만 국제적으로는 '누군가와 교류를 중단하고 아무런 설명 없이 소통을 멈추는 것'을 가리키는 용어로 사용되죠. 예를 들어, 친구나 연인과의 오랜 관계를 아무런 이유 없이 갑자기 끝내버리는 행동을 'Ghosting'이라고 합니다. (eng) Actually, it carries that meaning as well. 'Ghosting' basically means 'to disappear like a ghost'. However, internationally, it is used as a term to refer to 'ending communication with someone suddenly and without any explanation'. For example, the act of abruptly ending a long relationship with a friend or a romantic partner for no reason is called 'Ghosting'.	t.teach.direct_answer
tutor	<user>님께서 이런 케이스를 경험해보신 적이 있나요? (eng) Have you ever experienced such a case, <user>?	t.general
student	아니요 (eng) No.	
tutor	그런 경험이 없어서 다행입니다. 그런 사람들을 상대하는 것은 정말로 힘들 수 있습니다. (eng) It's good that you haven't had such an experience. Dealing with such people can be really tough.	t.general
tutor	그럼 오늘의 자료를 한 번 볼까요? 제가 설명하면서 *핵심 단어*나 *표현*에 대해 설명하겠습니다. 우리가 전에 배운 'innermost'를 기억하시나요? 이 단어를 한번 해석해보세요. (eng) Shall we take a look at today's material? I will explain the *key words* and *expressions* as we go along. Do you remember 'innermost' that we learned before? Try translating this word.	t.assess.follow_up_question
student	속깊은? (eng) Deep inside?	
tutor	맞습니다, 'innermost'는 '가장 깊은, 가장 속깊은'이라는 뜻입니다. (eng) That's right, 'innermost' means 'deepest, very deep inside.'	t.teach.direct_answer
tutor	그럼 이 문장 "sharing their *innermost* thoughts and feelings"을 한 번 해석해볼까요? (eng) Shall we translate the phrase "sharing their *innermost* thoughts and feelings"?	t.assess.follow_up_question
student	그들의 속깊은 생각과 감정들을 공유하면서 (eng) Sharing their deepest thoughts and emotions.	
tutor	정확하게 해석하셨습니다! 잘하셨어요. (eng) You've translated it accurately! Well done.	t.engage.encourage
tutor	그럼 이제 다음 단어 'ditching' 에 대해 조금 배워봅시다. 'Ditching'은 무슨 뜻일까요? (eng) Now let's learn a bit about the next word, 'ditching'. What does 'ditching' mean?	t.assess.follow_up_question
student	잘 모르겠어요. (eng) I'm not sure.	
tutor	괜찮습니다. 'Ditching'은 '버리다, 포기하다'라는 뜻입니다. (eng) That's okay. 'Ditching' means 'to throw away, to abandon'.	t.teach.direct_answer
tutor	예를 들어, "He ditched me at the party" 라는 문장은 그가 파티에서 나를 버렸다는 뜻입니다. (eng) For example, the sentence "He ditched me at the party" means that he left me behind at the party.	t.teach.method. vocab_expression.examples
tutor	그럼 'ditching a relationship'은 어떻게 해석되나요? (eng) So, how would you translate 'ditching a relationship'?	t.assess.follow_up_question

student	관계를 버리다? (eng) Ending a relationship?	
tutor	네, 맞습니다. '관계를 버리다' 혹은 '관계를 포기하다'로 해석됩니다. (eng) Yes, that's correct. It can be translated as 'ending a relationship' or 'abandoning a relationship.'	t.teach.direct_answer
tutor	이제 한 문장을 전체적으로 해석해봅시다. "ghosting isn't just Gen Z slang for *ditching* a relationship" 이 표현은 어떻게 해석되나요? (eng) Now let's translate a whole sentence. How would you translate the phrase "ghosting isn't just Gen Z slang for *ditching* a relationship"?	t.assess.follow_up_question
student	모르겠어요. Gen Z slang 이 뭔가요? (eng) I'm not sure. What is Gen Z slang?	
tutor	아, 그러네요. 설명을 빼먹었네요. 'Gen Z slang'은 Z세대(1996년 이후 출생한 세대)가 사용하는 특유의 신조어나 유행어를 의미합니다. (eng) Ah, right. I forgot to explain that. 'Gen Z slang' refers to the unique new words or trendy language used by Generation Z (those born after 1996).	t.teach.direct_answer
tutor	그래서 이 문장은 크게 "Ghosting이란 단어는 단순히 Z세대의 신조어로서 '관계를 버리다' 라는 의미만을 지니는 것이 아니다." 라고 해석할 수 있습니다. (eng) So, this sentence can basically be translated as "The word 'Ghosting' does not just carry the meaning of 'abandoning a relationship', which is a new term popular among Generation Z."	t.teach.method.definition
tutor	이제 이해가 좀 되시나요? 이 문장을 한 번 해석해보시겠어요? (eng) Are you beginning to understand now? Would you like to try translating that sentence?	t.assess.confirmation_questions
student	이미 해석해주신 것 같습니다 (eng) It seems you've already translated it for me..	
tutor	그렇군요, 제 설명이 조금 빠르기도 했네요. 더 많은 연습을 위해서 다음 문장을 한 번 같이 봅시다. "Left feeling confused and heartbroken, she tried to follow up several times with Sydney, even a year after the ghosting happened." 이 문장을 한 번 해석해보세요. (eng) Right, I may have explained a bit quickly. For more practice, let's look at the next sentence together. "Left feeling confused and heartbroken, she tried to follow up several times with Sydney, even a year after the ghosting happened." Try translating this sentence.	t.assess.display_question
student	그녀는 Sydney 와 연락이 닿으려고 여러번 노력했다. (eng) She tried several times to get in touch with Sydney.	
tutor	그렇습니다, 잘하셨어요. 그런데 '그녀' 대신에 'Sydney'라는 이름을 사용하는 것이 더 자연스럽습니다. 그리고 "even a year after the ghosting happened" 부분을 추가하면 "혼란스럽고 가슴 아파하는 감정을 느낀 채로, Sydney와 연락이 닿으려고 여러 번 노력했다. 심지어는 '고스팅'이 일어난 지 1년이 지난 후에도 말이죠." 라고 해석할 수 있을 것 같아요. 이제 잘 이해가 되세요? (eng) That's correct, well done. However, it would be more natural to use the name 'Sydney' instead of 'she'. And if you add "even a year after the ghosting happened", it can be interpreted as "Feeling confused and heartbroken, she tried several times to get in touch with Sydney, even a year after the ghosting had occurred." Do you understand it well now?	t.teach.repair
student	네 (eng) Yes.	
tutor	잘 이해하셨다니 다행이네요. 오늘 우리가 배운 'Ghosting'이라는 표현은 꽤나 공식적인 문맥에서도 자주 사용되므로 알아두시면 도움이 될 것입니다. (eng) I'm glad you understand it. The expression 'Ghosting' that we learned today is often used even in formal contexts, so it will be helpful for you to remember.	t.teach.review
tutor	오늘도 좋은 수업이었어요. 수고하셨습니다, <user>님! 다음 수업에서 뵙겠습니다. (eng) It was a good lesson today. Thank you for your hard work, <user>! See you in the next class.	t.general