

Empowering Private Tutoring by Chaining Large Language Models

Anonymous ACL submission

Abstract

Artificial intelligence has been applied in various aspects of online education to facilitate teaching and learning. However, few approaches has been made toward a complete AI-powered tutoring system. In this work, we explore the development of a full-fledged intelligent tutoring system powered by state-of-the-art large language models (LLMs), covering automatic course planning and adjusting, tailored instruction, and flexible quiz evaluation. To make the system robust to prolonged interaction and cater to individualized education, the system is decomposed into three interconnected core processes-*interaction*, *reflection*, and *reaction*. Each process is implemented by chaining LLM-powered tools along with dynamically updated memory modules. Tools are LLMs prompted to execute one specific task at a time, while memories are data storage that gets updated during education process. Statistical results and feedback from human users testify the advantage of the proposed system in long-term interaction, showcasing the benefits from structured memory control and stable reflection and reaction.

1 Introduction

Online education, along with artificial intelligence (AI) technology, brought the aspiration of personalized tutoring within reach (Bloom, 1984). AI has been used to assist education in multiple aspects, ranging from adaptive content recommendation (Costello and Mundy, 2009), automatic performance evaluation (McDonald et al., 2013; Grivokostopoulou et al., 2017), to personalized instruction and dynamic feedback (Bhutoria, 2022; Tang et al., 2020; Gordon et al., 2016; Grawemeyer et al., 2016; Dzikovska et al., 2014). Although a few early approaches have been made towards a stand-alone dialogue-based intelligent tutoring system (Graesser et al., 2001; Rus et al., 2014), most

of them are domain-specific and focus primarily on guiding the users to solve a pre-defined problem. Nevertheless, a more ultimate exploration lies in the pursuit of a fully-fledged AI-driven tutoring system with greater flexibility and generalizability that teaches in a systematic and consistent manner on a much broader range of knowledge.

While previous works often employ diverse techniques jointly, including learner style classification (Nihad et al., 2017), data mining (Echeverria et al., 2015), Bayesian learning (Grawemeyer et al., 2016), etc, the recent emergence of large language models (LLMs) (Devlin et al., 2019; Raffel et al., 2019; Brown et al., 2020; Bommasani et al., 2021; Han et al., 2021), like ChatGPT (OpenAI, 2022), has broadened our imagination on new designs of intelligent tutoring systems. LLMs impressed people firstly with the ability to generate and transform information following human instructions, then with the potential in task planning and tool usage. Given their extraordinary ability in providing detailed and accessible content and summarizing information, LLMs become great sources for consulting a wide range of knowledge. Additionally, LLMs have shown remarkable capabilities for interacting in a conversational, dialog-based manner, which makes them well-suited for tutoring applications where there is a back-and-forth exchange with a student. Although challenges like hallucination remain (Bang et al., 2023), LLMs can be adapted by further fine-tuning to have expertise in specific domains and pedagogical strategies tailored to different instructional needs.

In this work, we explore the potential of employing generative large language models to build a full-fledged dialogue-based personalized tutoring system. One specialty about a tutoring system compared to existing LLM-powered systems is that, education is a long-term cooperative process accomplished by AI and human users jointly. A well-designed tutoring system is required to cor-

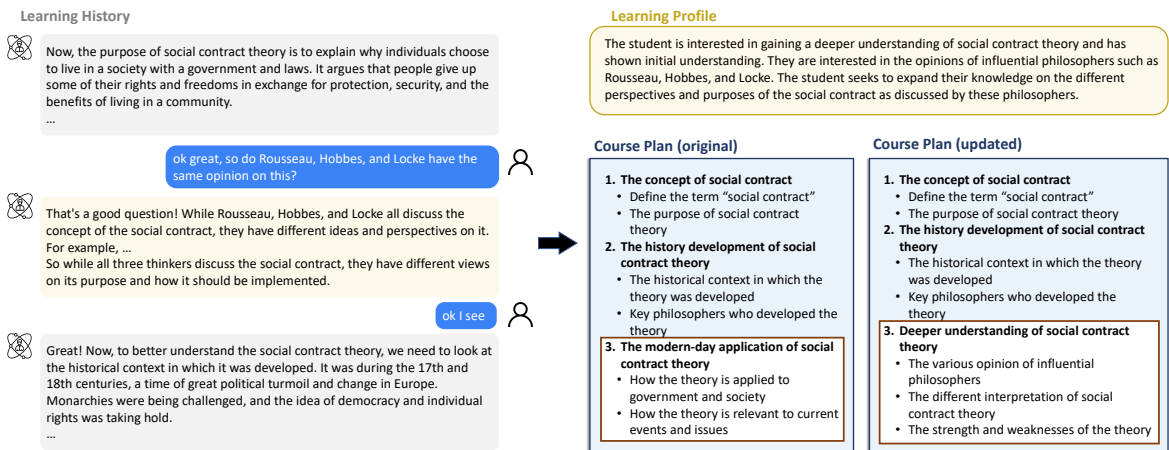


Figure 1: An example of the learning progress. Left side is the user interface directly controlled by interaction process. Right side is the backend memory changes brought by reflection and reaction processes.

rectly infer about human user’s mental states to achieve adaptive teaching, and meanwhile the user should be informed about the learning progress in order to cooperate more effectively. Therefore, the system faces some unique challenges in how to maintain an explainable and consistent control over the the learning progress, and how to attain an effective way in capturing human user’s behavior and responds correspondingly.

To solve the problem, the proposed system has a modularized design, encompassing three core processes-*interaction*, *reflection*, and *reaction*, each further composed of chained LLM-powered tools capable of performing specific tasks. The processes are connected to each other through various memory modules, which store the essential data describing the overall progress and support update and retrieval. Such design enables **structured memory control** and **stable reaction and reaction** on status quo. The proposed system carries out every stage in education systematically and dynamically, including instructing, question answering, exercise offering and evaluation.

Evaluation of the proposed system is conducted by analyzing statistics collection from learning logs and subjective human feedback. Results show that the system can satisfactorily handle various educational activities, including adaptive course plan design and updating, consistent instructing, impromptu question answering, and self-adaptive quiz designing and evaluation. Compared to ablation systems with partial functions, the designed system proves more advantageous in long-term educational process in stability and consistency, further supporting the plausibility of the design principle.

2 System Overview

The system is essentially a dialogue-based tutoring system that aims to help learners acquire knowledge on one given topic systematically. As shown in Figure 1, the whole learning process is carried out in natural language conversations, with time-to-time backend reflection and reaction to updates the memories. This section gives a general picture of the system working flow. We start with explaining the design principles by introducing three underlying processes within the system. Then we briefly go over components employed to realize each process. Finally, we provide a complete introduction of how each process and components work together. Note that the proposed system mainly explores personalized tutoring and adaptive system reflection with chained LLMs, while we do acknowledge the potential of fallacious and biased output due to inherent flaws in LLMs.

2.1 Design Principles

The system design demonstrates the breakdown into three core system processes: *Interaction*, *Reflection*, and *Reaction*. They each has a modularized implementation and is connected to one another to form an execution loop that empowers the whole education process.

Interaction. The interactive dialogue between the system and the user is the media for tutoring and learning, and therefore is the major process of the designed system. LLMs like ChatGPT can interact with users in a responsive and robust way in daily chit-chat. However, tasking them with long-term purposeful interaction is still tricky given restric-

082
083
084
085
086
087
088
089
090
091
092
093
094
095
096
097
098
099
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116

117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149

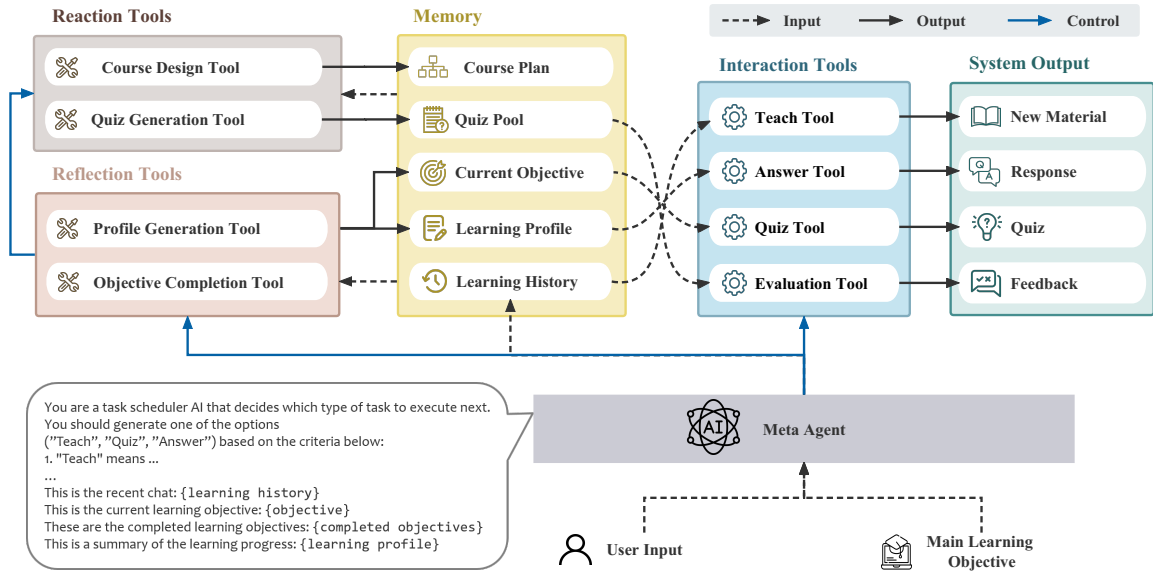


Figure 2: An overview of the system’s modularized implementation and execution in a single round of conversation.

tion on context length. As for educational purposes, it is especially important to keep the interaction on track, meanwhile ensuring its accessibility and informativeness.

Reflection. To facilitate interaction, we devise a reflection process to generate high-level insights on the learning progress, which serves as global information (Park et al., 2023). Meanwhile, it is expected to help adjust system response dynamically based on user preference and behavior to achieve personalized tutoring.

Reaction. Along with reflection, reaction refers to the automatically triggered system behavior afterwards, including adjustment of course plan and quiz generation. It differs from the interaction process in that *interaction* is always triggered directly by a new round of response, while *reaction* is performed at the backend from time to time, subsequent to reflection process.

2.2 Components

Tools. Under the principled design, each process is embodied by a set of tasks performed either sequentially or in parallel. For instance, there are diverse ways of engaging with the student, such as providing instructions, addressing questions, administering quizzes, and offering feedback. This variation in approaches complicates the development of a single unified solution. We therefore devise separate modules for each specific task to ensure performance. We term those modules as “tools”, and that each tool is a task-specific prompted LLM responsible for generating system output or updat-

ing memories, as shown in Figure 2. For example, *interaction* is broken down into four types of response in terms of education function, each hosted by one well-prompted tool. At each round, only one tool is used to generate the response.

Memories. Apart from tools, data storage is required to host information generated by reflection and reaction processes, while also supporting querying and retrieving. We propose four types of memories to record the progress and current status of learning, each stored in distinct data format and supports different ways of querying and updating. Another critical feature of the memories is that they serve as a linkage between different sets of tools to pass on information to control tool output. The detailed description of each tool and memory can be found in the next section.

Meta Agent. Above all three processes, we introduce meta agent, the single access of the control flow. It is powered by LLM and prompted to decide what specific tasks to execute next. See Figure 2 for an example prompt for controlling the interaction process. The template contains helpful information retrieved from the memory and asks for an output deciding the type of interaction process. In our implementation, the meta agent only controls the interaction tools, while we set a fixed time interval for the execution of reflection process.

2.3 Overview of Control Flow

Above all, all designs serve for the ultimate goal of better interaction with the users. The system reflects from time to time to update cognition on

the overall progress, and in turn refines the interaction production with new insights. At the frontend, the user first inputs what to learn with desired difficulty level. Then the system automatically calls the course design tool to generate the initial course plan, and starts the conversation accordingly. Upon receiving a new round of user input, the meta agent decides which interaction tool to use and the tool executes the task correspondingly to generate a new response with queried information from memories. At the backend, the reflection tools are triggered to reflect on the status quo and update the learning profile and current objective, after which the reaction tools will be triggered immediately to generate new quiz questions and update course plan.

As shown in Figure 2, the right side represents the interaction process that is presented on user interface, while the left side demonstrates the backend processes that are responsible for generating and updating memory modules. Practically, throughout each dialogue session, the reflection and reaction processes run alternatively at the backend, where the output result is periodically utilized by the interaction process to produce the final response to the user in each round. The learning proceeds until all objectives in the course plan have been completed.

3 Key Implementations

In this section, we present details of system implementation. Table 1 presents detailed usage of each tool in the three processes, including the input and output content, and the condition for tool execution. Based on the implementation, we further describe the key features of our system while explaining the rationale and mechanism for dynamics between different tools and memory modules.

3.1 Structured memory control

The interactive and cooperative feature of a tutoring system calls for the need to communicate with the users effectively about current and future progress. Meanwhile, it is also important to keep the system itself aware of the progress to ensure better stability. We therefore design various memory modules in different storage format and function to support the mutual communication. Designed to be both human understandable and machine operable, the course plan is stored in a tree structure, with each node representing an atomic topic in the course, and its child nodes representing the sub-topics. The

course is expected to be taught and learnt in depth-first traverse order. Current objective is a pointer pointing to the next uncompleted objective node in the tree that denotes current progress. The learning goes on until all leaf objectives in the current course plan have been completed. Such structure allows for presentation to the users, informing them of the overall status of learning, while enable mechanistic operation by the system.

Algorithm 1: Structured control and update of course plan and current objective pointer at each round of conversation.

Input: course plan p , learning profile l , current objective o

Output: p, o

if $ObjectiveCompletionTool(o)$ **then**
 | UpdateStatus(p);
end if
 $p' \leftarrow CourseDesignTool(p, l)$;
 $p \leftarrow Merge(p, p')$; // to mitigate generation error
 $o \leftarrow NextObjective(p)$;
return p, o

To facilitate such design, specialized tools are created to assist course plan generation and update, and current objective update. Specifically, the course design tool is used at the beginning of the learning to generate the initial course plan based on user’s desired topic and difficulty level. In each new round of conversation, objective completion tool is called to update status of the current objective based on the recent and relevant learning history. Then, the course design tool is asked to update the current course plan while maintaining the completed objectives. As for learning histories, the recent history is stored as plain text that can be directly fed into the LLM, whereas the relevant history is stored along with their embedding and queried with cosine similarity with embeddings of current objective upon usage. The detailed mechanism can be seen in Figure 3. Algorithm 1 shows the structured control and update of course plan in each round of conversation. Note that the merge operation on generated course plan and the old one ensures the consistency of completed objectives and mitigate potential instability in ChatGPT’s generation results. Meanwhile, the benefit of explicitly collecting completed objectives also extends to more effective quiz offerings, which will be detailed in the next section.

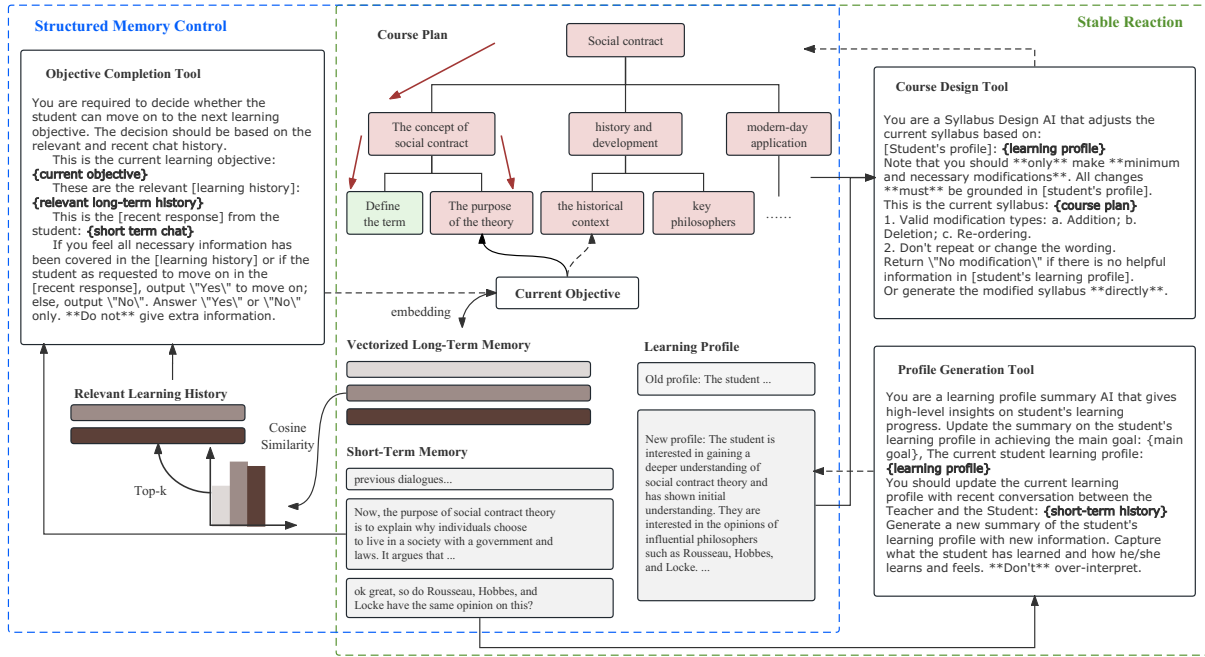


Figure 3: A detailed illustration of how course plan is stored and manipulated structurally and how reflection process helps stabilize the reaction followed.

Process	Tool Name	Execution Condition	Input	Output/Update
Interaction	Teach Tool	Meta agent	Learning history, Current objective, Learning profile	System output
	Answer Tool	Meta agent	Learning history	System output
	Quiz Tool	Meta agent	Quiz pool, Learning profile	System output
	Evaluation Tool	Quiz	Learning history	System output
Reflection	Profile Generation	Each round	Learning history, Learning profile	Memory: Learning profile
	Objective Completion	Each round	Learning history, Current objective	Memory: Current objective
Reaction	Course Design	Profile generation	Course plan, Learning profile	Memory: Course plan
	Quiz Generation	Objective completion	Learning history, Current objective	Memory: Quiz pool

Table 1: A summary of the detailed tool usage. “Input” means the memories are part of the prompt.

3.2 Stable reflection and reaction

Apart from reflecting on objective status, an important component is user’s learning profile. Learning profile summarizes what the user has learned and gives high-level insight on the user’s preference based on conversation history. Though not directly presented to users, it is crucial to the stability of system’s memory update and overall understanding of the learning process. It is especially useful as part of the input to course design tool to provide direction for course plan adjustment.

At each round of conversation, the system automatically reflects on the recent dialogue and updates learning profile with profile generation tool. The tool is a prompted LLM that takes recent dialogues and current profile summary as input and outputs a new version of learning profile, summarizing the learned knowledge, the user’s reaction and request mainly. Then it is fed into the course design

tool for a new version of course plan generation. Figure 3 provides an example of profile generation tool generating high-level insight of “the student seeks to expand their knowledge on the different perspectives and purposes of the social contract as discussed by the philosophers.” after the user asks a follow-up question about different philosophers’ opinion. This reflection process in turn triggers the course design tool, which takes the new learning profile as input and updates the course plan.

The reaction process “quiz generation” is similarly triggered by objective completion tool. Whenever the objective is judged as completed, the quiz generation tool is called to generate 3 ~ 5 quiz questions for the current objective, with relevant queried learning history provided. It ensures the stability and relevance of the generated quiz question. The questions are stored in the memory until the meta agent decides it is time for a quiz, where

the corresponding quiz questions are retrieved from the storage for the completed objectives and further filtered and organized by the quiz tool in the interaction process to generate final quiz questions.

4 Experiments

To demonstrate and analyze the features of our tutoring system, the experiments are conducted in two folds. We invite a number of users to learn a series of pre-defined topics using the system. During interaction, we collect critical statistics and record the conversation for future analysis. After learning completes, the users are required to answer a questionnaire to rate their experience with the system from multiple perspectives. We also develop ablation systems to better understand the effect of each process and module.

4.1 Experimental Design

System Setup. In addition to the main system, we implement two ablation systems with only partial functions. Specifically, we have one system without reflection process and another with both reflection and reaction processes removed. The detailed description of the two ablation systems can be found in Appendix A.1. We collect 80 different topics as the main learning objectives for evaluation and randomly assign them with different difficulty levels. The topics cover a wide range of academic subjects and some daily life phenomena, varying in granularity and language format. In evaluation, each topic is learned independently with three systems, making up altogether 240 courses. Further details about topic collection and tool prompt design and used can be found in Appendix A.2.

Participants. We invite 13 average adult users who are proficient in English to participate in learning. Every single course is randomly assigned to one user, while we make sure that each participant does not get repeated course topics.

Statistical Analysis. We collect various statistics for analysis, including (1) **Complexity of course plan** reflects the ability to design adaptive course plan; (2) **Average length of system response** and **Average number of objectives per response** are indicators of instruction informativeness; (3) **Frequency of course plan update** shows the reflective feature of the system; and (4) **Frequency of in-course quiz** explores the pattern of quiz offerings.

Survey Design. After completing the course, the learner is required to answer a survey composed of

9 questions targeting different aspects of the system. Each question is a statement to be rated on a 1~5 scale, where higher scores mean better qualities. Table 2 presents the statements by category.

Course Plan

1. Relevance: The course plan is relevant to the main objective.
2. Coherence: The course plan is coherent and logical.
3. Compatibility: The course plan is compatible with difficulty level.

Instruction

4. No Repetition: There is no repeated instruction.
5. Accessibility: The language used is easy to understand.

Question Answering

6. Timeliness: The learner's questions always get immediate response.
7. Consistency: The response is consistent with learning material.

Quiz

8. Relevance: The quiz questions match what has been covered.
 9. Judgment: The quiz evaluation is accurate in parsing and scoring.
-

Table 2: The complete survey questions. Learners are asked to rate the compatibility of each statement on a scale of 1~5.

4.2 Results

Statistical Results. Figure 4 presents the statistical characteristics related to teach tool, including average length of output and the average number of objectives covered in each generation. Overall all systems can generate tailored output according to difficulty level. Higher difficulty comes along with longer and more informative output. It means the teach tool is successfully aware of the dynamic prompting controlled by difficulty. What is worth noting is that the main system generates significantly longer output with the smallest variation. It demonstrates that the main system is able to consistently generate informative content on the given topic, which is further testified by the number of objectives covered in each output. This phenomenon shows the benefits of structured memory control, where the objective completion tool reflects on and updates the current objective so that the teach tool is prompted to give new materials on one specific atomic objective.

Figure 5 showcases the behavior of course design tool. Note that all systems start with the same initial course plan. The system without reflection relies on learning history in course plan updates, while the main system takes advantage of the learning profile. Similarly, the tool can generate difficulty-aware course plans, seen in a positive correlation between complexity and difficulty. It is also observed that the future updates in course plan tend to do more addition than deletion, making the course plan more complex than the initial

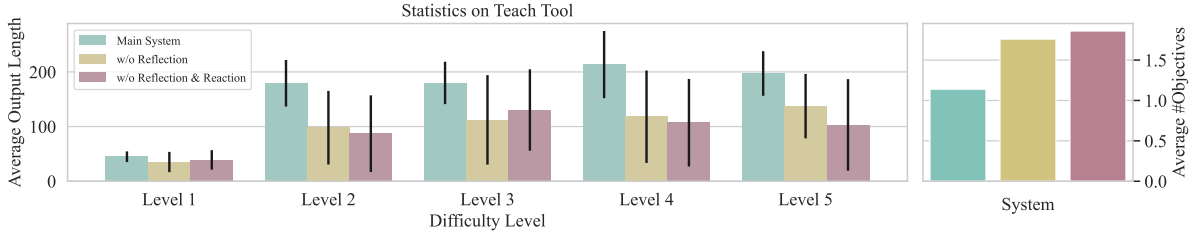


Figure 4: Average output length (calculated by the number of words) and the number of objectives covered in each output for different systems. Average number of objectives are manually annotated with 50 randomly sampled response from each system.

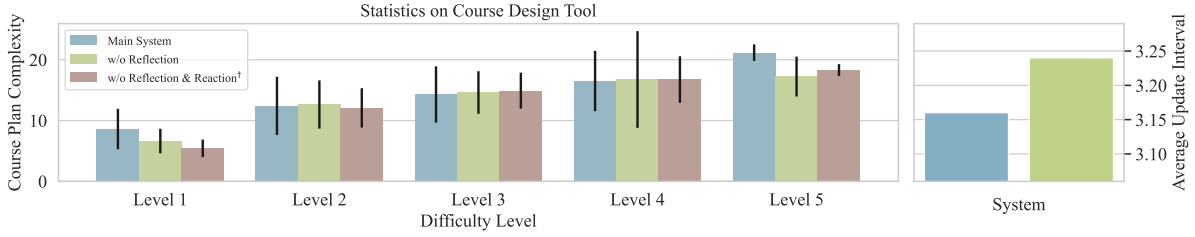


Figure 5: Average course plan complexity (calculated by the number of objectives) and update interval (calculated by the number of conversation rounds in between) by course design tool for different systems.† means this is the baseline statistics as the system without reflection or reaction processes has a fixed course plan throughout learning.

state. Note the main system shows more stability in complexity control compared to ablation system without intermediate learning profile generation step, demonstrating the effectiveness of reflection process stabilizing the reaction process. In terms of frequency of updates, we observe that the main system seems more prone to making updates, meaning that it is easier to detect helpful information in a compressed learning profile compared to verbose learning history.

As for in-course quizzes, the average interval between quizzes for the main system and ablation system are 6.91 and 10.20, respectively. Without the backend processes, the meta agent has little information available to suggest quiz generation. We also calculate the average number of objectives covered in each quiz for the main system, the result is reasonably 3.96 with good stability.

Survey Results. Table 3 and Table 4 present results on survey questions. Overall, with powerful ChatGPT, all systems demonstrate promising usability and quality. The designed reflection and reaction processes offer advantages in complex learning settings. For intricate course plan updates, the main system achieves better coherence and compatibility with difficulty levels. Through reflection and memory querying, the system maintains better control with fewer repetitive instructions during longer conversations. While all three systems seem to

generate highly relevant quiz questions, the main system generates much more fine-grained questions querying about detail information in the learning material, while the questions from ablation systems are largely answerable based on the course plan (see Appendix B). This, together with better evaluation accuracy, showcases the benefit of online quiz storage. However, it should also be noted that although the main system performs satisfactorily on QA consistency, the timeliness in response is compromised by backend processes. It indicates that prompts with global information may interfere with the LLM’s ability to focus locally and generate timely and coherent response.

5 Related Work

Ever since the development of artificial intelligence techniques, methods and tools have been proposed to assist in teaching and learning process. AutoTutor (Graesser et al., 2001) is the first conversation-based intelligent tutoring system, which inspires a number of works followed (McNamara et al., 2006; D’mello and Graesser, 2012; Graesser et al., 2003; Rus et al., 2013; VanLehn, 2011; Dzikovska et al., 2014). In addition to AutoTutor’s application to various fields, enhancement of specific aspects of education are also investigated, including adaptive feedback (Dzikovska et al., 2014; Roscoe and McNamara, 2013), learning material recommen-

System	Course Plan			Instruction		Question Answering		Quiz	
	Relevance	Coherence	Compatibility	No Repetition	Accessibility	Timeliness	Consistency	Relevance	Judgment
Main System	4.72	4.51	4.65	4.29	4.77*	4.41	4.82	4.88	4.24
w/o Reflection	4.71	4.62	4.77	4.42	4.66	4.64	4.85	4.75	4.65
w/o Reflection & Reaction	4.97 [†]	4.77 [†]	4.78 [†]	4.38	4.77	4.75	4.95	4.86	4.36

Table 3: Survey results for learning courses at difficulty level 1~3. † means the score evaluates the initial course plan only, as no changes in course plan happen throughout the learning process. It could be viewed as the static quality evaluation of course plan generated from scratch. * means p-value < 0.1 using t-test.

System	Course Plan			Instruction		Question Answering		Quiz	
	Relevance	Coherence	Compatibility	No Repetition	Accessibility	Timeliness	Consistency	Relevance	Judgment
Main System	4.87*	4.87**	4.73	4.27	4.53	3.67	5.00	4.87	4.20
w/o Reflection	4.67	4.60	4.67	3.87	4.87	4.80	4.67	4.40	4.13
w/o Reflection & Reaction	4.93 [†]	4.73 [†]	4.80 [†]	4.07	5.00	4.33	4.93	4.93	4.13

Table 4: Survey results for learning courses at difficulty level 4~5. † means the score evaluates the initial course plan only, as no changes in course plan happen throughout the learning process. It could be viewed as the static quality evaluation of course plan generated from scratch. * and ** means p-value < 0.1 and < 0.05 using t-test.

479 dation (S., 2014; Mohammed and Mohan, 2015),
480 and classifying learners (Grawemeyer et al., 2016;
481 Nihad et al., 2017; J., 2014). Commonly adopted
482 techniques include data mining (Echeverria et al.,
483 2015), condition-action rule based (J., 2014; S.,
484 2014), and bayesian based methods (Grawemeyer
485 et al., 2016), and reinforcement learning (Malpani
486 et al., 2011; Gordon et al., 2016). NLP-specific
487 techniques like semantic analysis (Graesser et al.,
488 2001) and textual entailment (Rus and Graesser,
489 2006; VanLehn et al., 2007; McCarthy et al., 2008)
490 are also adopted. In terms of application field, ex-
491 isting systems often rely on well-structured knowl-
492 edge bases and therefore only target a single do-
493 main, most popular among which are health (Mc-
494 Donald et al., 2013; El Saadawi et al., 2008), com-
495 puter science (Grivokostopoulou et al., 2013; Mo-
496 hammed and Mohan, 2015), and language learn-
497 ing (S., 2014; Emran and Shaalan, 2014).

498 As for applications with LLMs, with proper
499 prompting and chaining, a number of works have
500 exploited LLMs in decomposing tasks (Wei et al.,
501 2023), refining answers (Madaan et al., 2023; Shinn
502 et al., 2023), using external tools (Qin et al., 2023),
503 and simulating human behaviors (Park et al., 2023).
504 While our work focuses on building an interactive
505 tutoring system that works and teaches coopera-
506 tively with human users, featuring reflection on
507 user’s behavior and adaptive course design.

508 6 Discussion

509 AI in education is an intriguing yet tricky topic,
510 especially for a stand-alone AI-powered education
511 system. This work is a pioneering exploration of an
512 LLM-powered intelligent tutoring system, with an

513 emphasis on the possibility of employing LLMs to
514 complete complex and dynamic tasks based on nat-
515 ural language conversation with users. Although
516 our evaluation reveals mechanism of the adaptive
517 behaviors and proves the advantage of memory
518 mechanism and process design in long-term in-
519 teraction, we acknowledge that comprehensively
520 evaluating an intelligent tutoring system is far more
521 tricky. In addition to the ongoing debate and discus-
522 sion about what makes good education (Graesser
523 et al., 1995; Cohen et al., 1982), AI techniques
524 face their own special concerns such as the validity
525 of generated education content and the potential
526 bias from training data (Feng et al., 2023; Nozza
527 et al., 2022). Meanwhile, quantitatively assessing
528 the learning outcome is also difficult, as our sys-
529 tem teaches open-ended topics and devises quiz
530 questions by itself.

531 Therefore, one important direction for future
532 work is to incorporate more techniques to avoid
533 hallucination and mitigate bias in LLMs, for in-
534 stance, through domain-specific fine-tuning and
535 search-engine-based fact-checking. As for evalu-
536 ation, apart from potential falsehood and bias, we
537 observe that properly evaluating multi-turn conver-
538 sation is in itself a topic worth investigating. More
539 standard metrics should be developed other than hu-
540 man evaluation which is rather subjective. Despite
541 that, as a first approach in chaining LLMs to build
542 a full-fledged intelligent tutoring system, this work
543 demonstrates the promising application of a pure
544 LLM-based purposeful interaction system based on
545 reflection and memory mechanism, which might
546 inspire future efforts in employing LLMs to build
547 increasingly interactive and reflective systems.

A Ethical Statement

The implemented system realizes a full-fledged intelligent education process, which can help achieve efficient online personalized tutoring and in turn, promote education equality. On the other hand, the experimental system faces the risk of generating inaccurate information given the hallucination problem in LLMs. The inherent bias may also mislead the users. It is important to address these issues before a practical application is launched. The participants in the evaluation process are all fairly paid with agreed salaries.

References

- Yejin Bang, Samuel Cahyawijaya, Nayeon Lee, Wenliang Dai, Dan Su, Bryan Wilie, Holy Lovenia, Ziwei Ji, Tiezheng Yu, Willy Chung, Quyet V. Do, Yan Xu, and Pascale Fung. 2023. [A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity](#).
- Aditi Bhutoria. 2022. [Personalized education and artificial intelligence in the united states, china, and india: A systematic review using a human-in-the-loop model](#). *Computers and Education: Artificial Intelligence*, 3:100068.
- Benjamin S Bloom. 1984. The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational researcher*, 13(6):4–16.
- Benjamin S Bloom and David R Krathwohl. 2020. *Taxonomy of educational objectives: The classification of educational goals. Book 1, Cognitive domain*. longman.
- Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. 2021. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.
- Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. In *Proceedings of NeurIPS*.
- Peter A Cohen, James A. Kulik, and Chen-Lin C. Kulik. 1982. [Educational outcomes of tutoring: A meta-analysis of findings](#). *American Educational Research Journal*, 19(2):237–248.
- R. Costello and D.P. Mundy. 2009. [The adaptive intelligent personalised learning environment](#). In *2009 Ninth IEEE International Conference on Advanced Learning Technologies*, pages 606–610.

- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Sidney D’mello and Art Graesser. 2012. [Autotutor and affective autotutor: Learning by talking with cognitively and emotionally intelligent computers that talk back](#). 2(4):1–39.
- Myroslava O. Dzikovska, Natalie B. Steinhauer, Elaine Farrow, Johanna D. Moore, and Gwendolyn E. Campbell. 2014. [Beetle ii: Deep natural language understanding and automatic feedback generation for intelligent tutoring in basic electricity and electronics](#). *International Journal of Artificial Intelligence in Education*, 24:284–332.
- Vanessa Echeverria, Bruno Guamán, and Katherine Chiluita. 2015. [Mirroring teachers’ assessment of novice students’ presentations through an intelligent tutor system](#). In *2015 Asia-Pacific Conference on Computer Aided System Engineering*, pages 264–269. IEEE.
- Gilan M El Saadawi, Eugene Tseytlin, Elizabeth Legowski, Drazen Jukic, Melissa Castine, Jeffrey Fine, Robert Gormley, and Rebecca S Crowley. 2008. [A natural language intelligent tutoring system for training pathologists: Implementation and evaluation](#). *Advances in health sciences education*, 13:709–722.
- Mustafa Al Emran and Khaled Shaalan. 2014. [A survey of intelligent language tutoring systems](#). In *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 393–399.
- Shangbin Feng, Chan Young Park, Yuhan Liu, and Yulia Tsvetkov. 2023. [From pretraining data to language models to downstream tasks: Tracking the trails of political biases leading to unfair NLP models](#). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 11737–11762, Toronto, Canada. Association for Computational Linguistics.
- Goren Gordon, Samuel Spaulding, Jacqueline Kory Westlund, Jin Joo Lee, Luke Plummer, Marayna Martinez, Madhurima Das, and Cynthia Breazeal. 2016. [Affective personalization of a social robot tutor for children’s second language skills](#). 30(1).
- Arthur C. Graesser, G. Tanner Jackson, Eric Mathews, Heather H. Mitchell, Andrew M. Olney, Matthew Ventura, Patrick Chipman, Donald R. Franceschetti, Xiangen Hu, Max M. Louwerse, and Natalie K. Person. 2003. [Why/autotutor: A test of learning gains from a physics tutor with natural language dialog](#).

656	Arthur C. Graesser, Natalie K. Person, and Joseph P. Magliano. 1995. Collaborative dialogue patterns in naturalistic one-to-one tutoring. <i>Applied Cognitive Psychology</i> , 9:495–522.	tutoring system, istart. In <i>FLAIRS Conference</i> , pages 165–170. Citeseer.	712 713
657			
658			
659			
660	Arthur C. Graesser, Kurt VanLehn, Carolyn P. Rose, Pamela W. Jordan, and Derek Harter. 2001. Intelligent tutoring systems with conversational dialogue. 22(4):39.	Jenny McDonald, Alistair Knott, Sarah J. Stein, and Richard Zeng. 2013. An empirically-based, tutorial dialogue system: design, implementation and evaluation in a first year health sciences course.	714 715 716 717
661			
662			
663			
664	Beate Grawemeyer, Manolis Mavrikis, Wayne Holmes, Sergio Gutierrez-Santos, Michael Wiedmann, and Nikol Rummel. 2016. Affecting off-task behaviour: How affect-aware feedback can improve student learning. In <i>Proceedings of the Sixth International Conference on Learning Analytics & Knowledge</i> , LAK '16, page 104–113, New York, NY, USA. Association for Computing Machinery.	Danielle S McNamara, Tenaha P O'Reilly, Rachel M Best, and Yasuhiro Ozuru. 2006. Improving adolescent students' reading comprehension with istart. <i>Journal of Educational Computing Research</i> , 34(2):147–171.	718 719 720 721 722
665			
666			
667			
668			
669			
670			
671			
672	Foteini Grivokostopoulou, Isidoros Perikos, and Ioannis Hatzilygeroudis. 2013. An intelligent tutoring system for teaching fol equivalence. In <i>The First Workshop on AI-supported Education for Computer Science (AIEDCS 2013)</i> , volume 20.	Phaedra Mohammed and Permanand Mohan. 2015. Dynamic cultural contextualisation of educational content in intelligent learning environments using icon. <i>International Journal of Artificial Intelligence in Education</i> , 25:249–270.	723 724 725 726 727
673			
674			
675			
676			
677	Foteini Grivokostopoulou, Isidoros Perikos, and Ioannis Hatzilygeroudis. 2017. An educational system for learning search algorithms and automatically assessing student performance. <i>International Journal of Artificial Intelligence in Education</i> , 27:207–240.	Elghouch Nihad, Yassine Zaoui Seghroucheni, et al. 2017. Analysing the outcome of a learning process conducted within the system als_corr [lp]. <i>International Journal of Emerging Technologies in Learning (Online)</i> , 12(3):43.	728 729 730 731 732
678			
679			
680			
681			
682	Xu Han, Zhengyan Zhang, Ning Ding, Yuxian Gu, Xiao Liu, Yuqi Huo, Jiezhong Qiu, Liang Zhang, Wentao Han, Minlie Huang, Qin Jin, Yanyan Lan, Yang Liu, Zhiyuan Liu, Zhiwu Lu, Xipeng Qiu, Ruihua Song, Jie Tang, Ji-Rong Wen, Jinhui Yuan, Wayne Xin Zhao, and Jun Zhu. 2021. Pre-trained models: Past, present and future. <i>AI Open</i> .	Debora Nozza, Federico Bianchi, and Dirk Hovy. 2022. Pipelines for social bias testing of large language models. In <i>Proceedings of BigScience Episode #5 – Workshop on Challenges & Perspectives in Creating Large Language Models</i> , pages 68–74, virtual+Dublin. Association for Computational Linguistics.	733 734 735 736 737 738 739
683			
684			
685			
686			
687			
688			
689	Gautier Izacard, Mathilde Caron, Lucas Hosseini, Sebastian Riedel, Piotr Bojanowski, Armand Joulin, and Edouard Grave. 2021. Unsupervised dense information retrieval with contrastive learning.	TB OpenAI. 2022. Chatgpt: Optimizing language models for dialogue. <i>OpenAI</i> .	740 741
690			
691			
692			
693	Hsieh S. J. 2014. Algorithm and intelligent tutoring system design for programmable controller programming. <i>International Journal of Advanced Manufacturing Technology</i> , 71:1099.	Joon Sung Park, Joseph C O'Brien, Carrie J Cai, Meredith Ringel Morris, Percy Liang, and Michael S Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. <i>arXiv preprint arXiv:2304.03442</i> .	742 743 744 745 746
694			
695			
696			
697	Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegrefe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, Shashank Gupta, Bodhisattwa Prasad Majumder, Katherine Hermann, Sean Welleck, Amir Yazdanbakhsh, and Peter Clark. 2023. Self-refine: Iterative refinement with self-feedback.	Yujia Qin, Shengding Hu, Yankai Lin, Weize Chen, Ning Ding, Ganqu Cui, Zheni Zeng, Yufei Huang, Chaojun Xiao, Chi Han, Yi Ren Fung, Yusheng Su, Huadong Wang, Cheng Qian, Runchu Tian, Kunlun Zhu, Shi Liang, Xingyu Shen, Bokai Xu, Zhen Zhang, Yining Ye, Bo Li, Ziwei Tang, Jing Yi, Yu Zhu, Zhenning Dai, Lan Yan, Xin Cong, Ya-Ting Lu, Weilin Zhao, Yuxiang Huang, Jun-Han Yan, Xu Han, Xian Sun, Dahai Li, Jason Phang, Cheng Yang, Tongshuang Wu, Heng Ji, Zhiyuan Liu, and Maosong Sun. 2023. Tool learning with foundation models. <i>ArXiv</i> , abs/2304.08354.	747 748 749 750 751 752 753 754 755 756 757 758
698			
699			
700			
701			
702			
703			
704	Ankit Malpani, Balaraman Ravindran, and Hema A Murthy. 2011. Personalized intelligent tutoring system using reinforcement learning. In <i>FLAIRS Conference</i> , pages 561–562.	Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2019. Exploring the limits of transfer learning with a unified text-to-text transformer. <i>ArXiv preprint</i> , abs/1910.10683.	759 760 761 762 763
705			
706			
707			
708	Philip M McCarthy, Vasile Rus, Scott A Crossley, Arthur C Graesser, and Danielle S McNamara. 2008. Assessing forward-, reverse-, and average-entailer indices on natural language input from the intelligent	Rod D Roscoe and Danielle S McNamara. 2013. Writing pal: Feasibility of an intelligent writing strategy tutor in the high school classroom. <i>Journal of Educational Psychology</i> , 105(4):1010.	764 765 766 767
709			
710			
711			

768 Vasile Rus, Sidney D’Mello, Xiangen Hu, and Arthur
769 Graesser. 2013. [Recent advances in conversational
770 intelligent tutoring systems](#). 34(3):42–54.

771 Vasile Rus and Arthur C. Graesser. 2006. Deeper nat-
772 ural language processing for evaluating student an-
773 swers in intelligent tutoring systems. In *Proceedings
774 of the 21st National Conference on Artificial Intelli-
775 gence - Volume 2, AAAI’06*, page 1495–1500. AAAI
776 Press.

777 Vasile Rus, Dan Stefanescu, Nobal Niraula, and
778 Arthur C. Graesser. 2014. [Deeptutor: Towards
779 macro- and micro-adaptive conversational intelligent
780 tutoring at scale](#). In *Proceedings of the First ACM
781 Conference on Learning @ Scale Conference*, pages
782 209–210. ACM.

783 Gómez S. 2014. Context-aware adaptive and person-
784 alized mobile learning delivery supported by uolmp.
785 *Journal of King Saud University – Computer and
786 Information Sciences*, 26:47.

787 Noah Shinn, Federico Cassano, Beck Labash, Ashwin
788 Gopinath, Karthik Narasimhan, and Shunyu Yao.
789 2023. [Reflexion: Language agents with verbal rein-
790 forcement learning](#).

791 Ying Tang, Joleen Liang, Ryan Hare, and Fei-Yue Wang.
792 2020. [A personalized learning system for parallel
793 intelligent education](#). *IEEE Transactions on Compu-
794 tational Social Systems*, 7:352–361.

795 Kurt VanLehn. 2011. The relative effectiveness of
796 human tutoring, intelligent tutoring systems, and
797 other tutoring systems. *Educational psychologist*,
798 46(4):197–221.

799 Kurt VanLehn, Pamela Jordan, and Diane Litman. 2007.
800 Developing pedagogically effective tutorial dialogue
801 tactics: Experiments and a testbed. In *Workshop
802 on Speech and Language Technology in Education*.
803 Citeseer.

804 Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten
805 Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and
806 Denny Zhou. 2023. [Chain-of-thought prompting elic-
807 its reasoning in large language models](#).

A Experimental Details 808

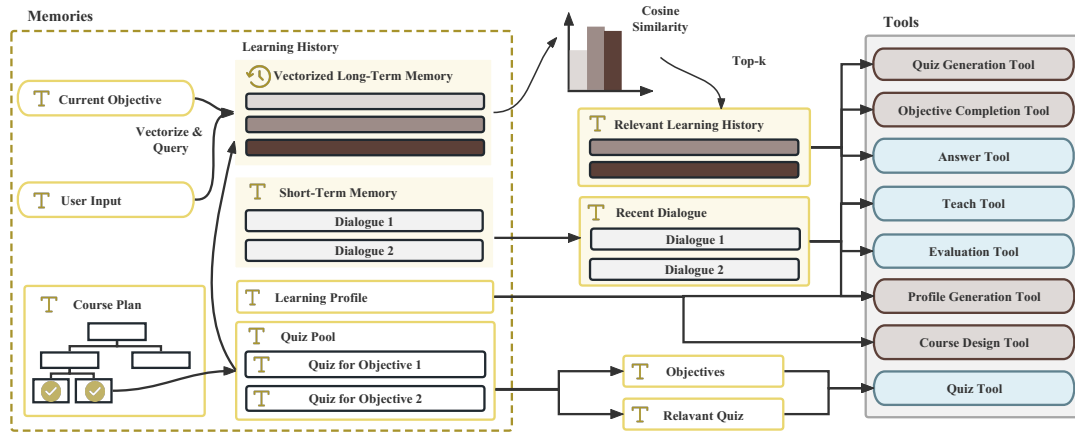
A.1 Ablation systems 809

810 For the first ablation system (**w/o Reflection**), we
811 remove the reflection process only. It means no
812 learning profile is generated throughout the whole
813 process, and the system does not explicitly judge
814 the objective status. Only the course plan and learn-
815 ing history are available for retrieval. As a result,
816 the reaction process is triggered at a fixed time in-
817 terval with limited input. For example, the course
818 design tool has only the recent learning history as
819 input without the learning profile, and the teach
820 tool has no access to the current ongoing objective
821 but can only infer from the course plan and learn-
822 ing history implicitly. At the same time, to mitigate
823 the loss of learning profile, we increase short-term
824 learning history volume to up to 10 rounds of con-
825 versation. This, however, will exceed the maximum
826 input length for ChatGPT in some extreme cases.

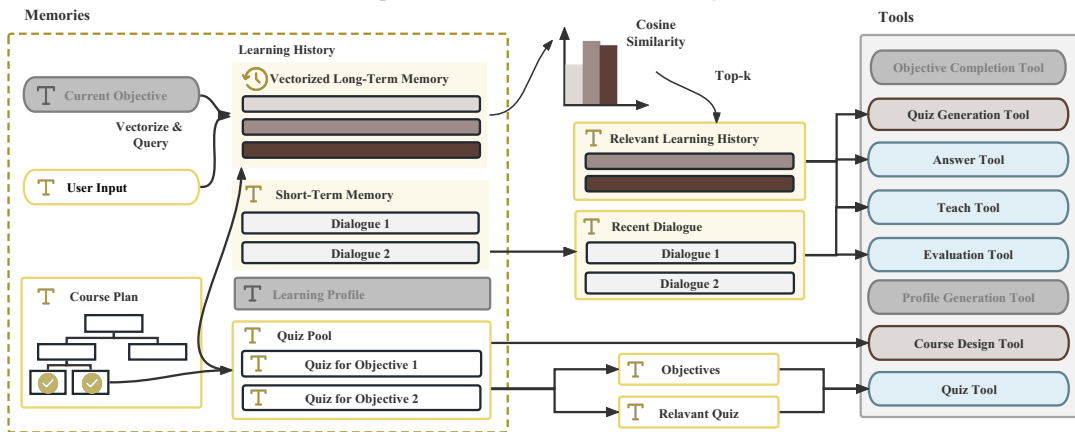
827 For the second ablation system (**w/o Reflection
828 & Reaction**), both reflection and reaction processes
829 are removed, and long-term learning history is also
830 made unavailable. This means only short-term
831 learning history and the initial course plan can be
832 used in all educational activities. Figure 6 is an
833 illustration of the differences between main system
834 and ablation systems.

A.2 Topic collection 835

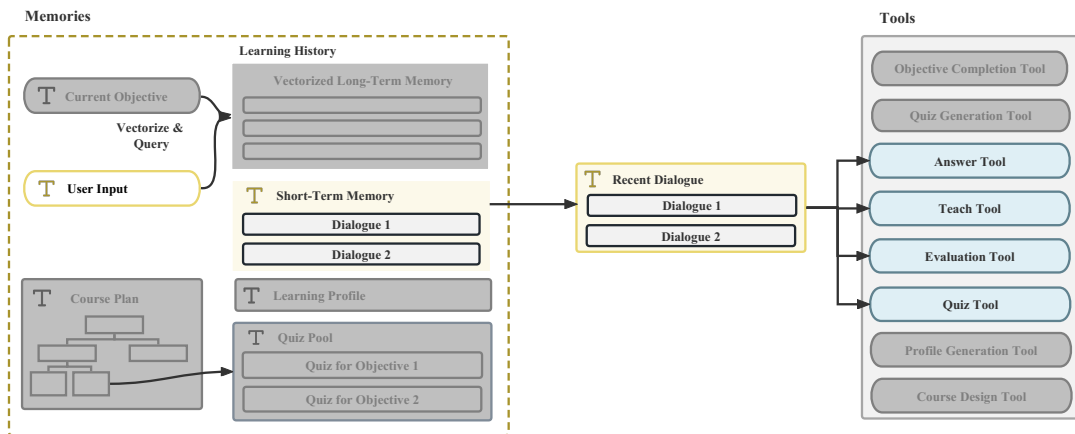
836 We employ GPT-4 to generate the list of main learn-
837 ing objectives used for system evaluation. We first
838 ask GPT-4 to generate a set of general academic
839 subject (e.g. Physics, Biology, Computer Science,
840 etc.). Then we ask for generation of more fine-
841 grained branches of discipline under each subject
842 (e.g. “Developmental psychology” in psychology
843 subject), and related classic concepts (e.g. “New-
844 ton’s first law” in physics subject). Besides, we
845 also include some daily phenomena that may in-
846 spire people’s wondering. We encourage GPT-4
847 to generate a typical list of them in diverse lan-
848 guage style. To demonstrate the system’s robust-
849 ness in dealing with various types of learning ob-
850 jectives, we randomly sample from course-grained
851 subjects, atomic concepts, and daily wondering. To
852 make the learning process more diverse and control-
853 lable, we also design 5 difficulty levels according to
854 Bloom’s taxonomy ([Bloom and Krathwohl, 2020](#)).
855 The course topics are randomly assigned to diffi-
856 culty levels. Specifically, we have 15 courses at
857 difficulty level 1, 25 at level 2, 25 at level 3, 10



(a) Implementation details for the main system.



(b) Implementation details for the first ablation system (w/o Reflection).



(c) Implementation details for the second ablation system (w/o Reflection & Reaction).

Figure 6: Implementation details for ablation systems, in comparison to main system. Modules in gray means the modules are de-activated or non-existent in the system.

858 at level 4, and 5 at level 5. Each course is learned
 859 and evaluated with all three systems, respectively.
 860 Table 10 is a completed list of all selected topics.
 861 In reference to Bloom’s taxonomy, we manually
 862 design the difficulty-specific prompt for course de-
 863 sign tool and teach tool for better adaptiveness and
 864 clarity. Table 5 and Table 6 present the prompts for
 865 course design tool and teach tool.

A.3 User interface

866 We build a frontend interface to support interaction
 867 with the backend system. Figure 7 shows the fron-
 868 t-end design used in real system evaluation. The
 869 main component is a chat window through which
 870 the user interacts with the system in natural lan-
 871 guage. On the right side is the current course plan,
 872 the completed objectives will be shown in blue (as
 873

Course Design Prompt

Level 1: Remembering (Knowledge) At this level, the focus is on providing basic facts. Students are required to remember facts, definitions, and concepts. Design a **very** short course.

Level 2: Understanding (Comprehension) This level involves describing and interpreting facts, definitions, and concepts. Design a course as concise as possible.

Level 3: Applying (Application) At this level, students are expected to use acquired knowledge in new and practical situations. The focus is on applying concepts to solve problems and complete tasks. Try to make the course concise and well-structured.

Level 4: Analyzing (Analysis) In this level, the course should break down information into its components to understand the relationships between parts and the overall structure. Try to make the course concise and well-structured.

Level 5: Evaluating (Evaluation) The course should assess the quality, validity, and relevance of information and arguments. Try to make the course concise and well-structured.

Table 5: Prompts for course design tool at each difficulty level.

Teach Prompt

Level 1: Teach in very simple and accessible language. Keep generated text short within a few sentences.

Level 2: Teach in simple and accessible language. Keep language and wording easy to understand.

Level 3: Teach in well-structured language and paragraph. Make information digestible. Try to use structured format to make it clearer, e.g. bullet point.

Level 4 & 5: Use precise language to explain things in a systematic way. Try to use structured format to make it clearer, e.g. bullet point.

Table 6: Prompts for teach tool at each difficulty level.

Difficulty Level	Update Interval	Max. Dialogue Round	Avg. Time Spent (mins)
1	1	10	17
2	1	15	30
3	2	20	40
4	3	25	60
5	4	30	60

Table 7: Designed update interval of learning profile and the maximum round of dialogue in system evaluation.

874 shown in Figure 7b). Since both ablation systems
875 do not conduct judgment on objective status. We
876 change the next objective status to “completed” at
877 each round of conversation on user interface to
878 make them visually indiscriminate to the main sys-
879 tem to avoid potential influence in evaluation.

880 A.4 Implementation details

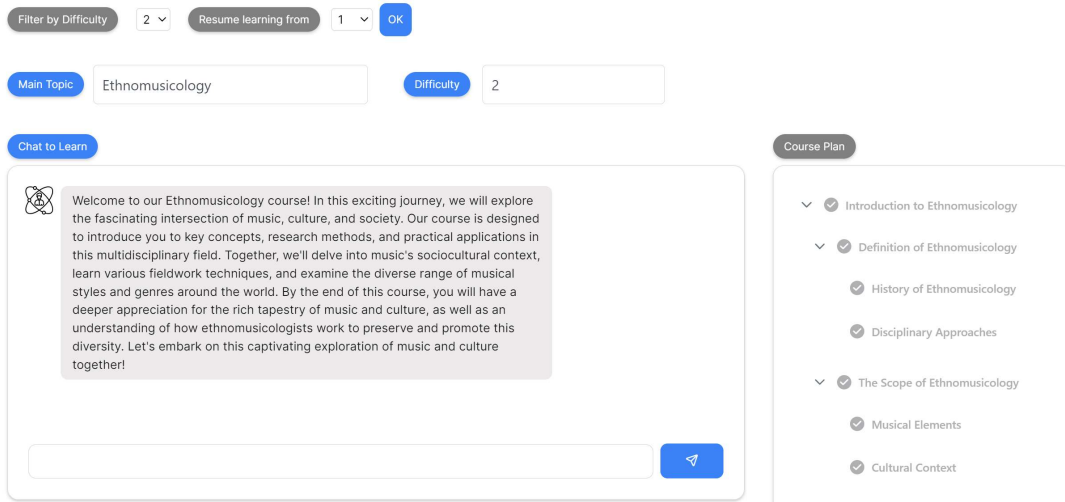
881 To ensure better usability and system stability, we
882 also set different time intervals for learning profile
883 update for each difficulty level, and control for the
884 maximum round of conversation. Table 7 presents
885 the detailed setting and the average time spent on
886 courses at every difficulty level. Altogether, the
887 evaluation for the three systems with 240 courses
888 takes around 150 hours.

In addition to in-course quiz automatically trig- 889
gered by meta agent, when all objectives have been 890
completed or the pre-defined maximum round has 891
been reached, the system explicitly calls for the 892
quiz chain to generate a final quiz for the course. 893
For final quiz generation, our system and ablation 894
system without reflection process use similarity- 895
based search for the most relevant 20 pieces of 896
history as input context, while the ablation sys- 897
tem with only interaction process uses the over- 898
all course plan. As for long-term history storage 899
and query, we use contriever model (Izacard et al., 900
2021) for embedding texts and cosine similarity 901
score for most relevant record ranking and query- 902
ing. The implementation of the whole system is 903
based on LangChain¹. 904

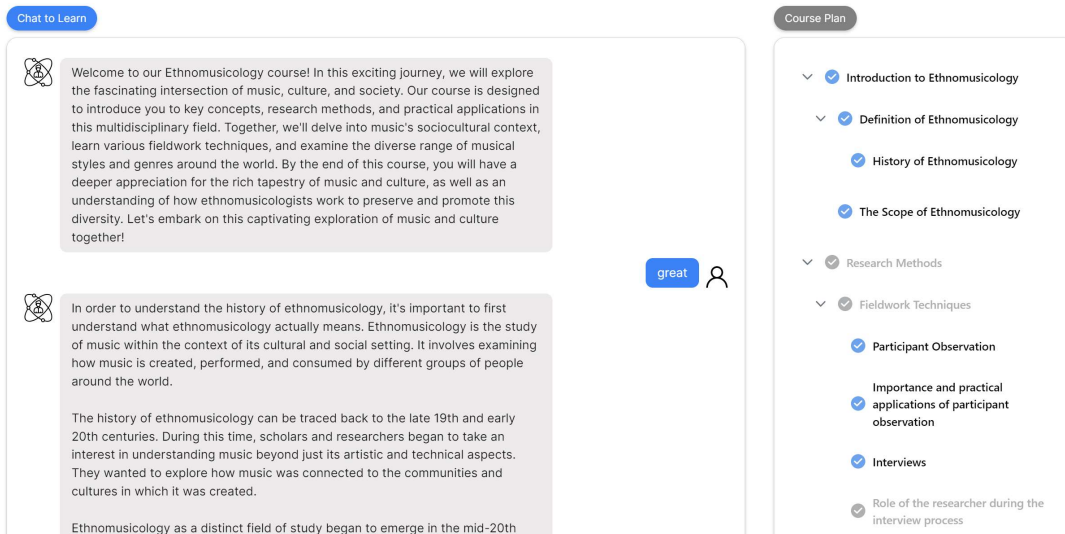
905 B Quiz questions

Table 8 presents sample quiz questions from the 906
main system and the ablation system. It can be 907
seen that the main system generates more detailed 908
questions, while the ablation system largely relies 909
on information available in the course plan and the 910
generated questions are less effective in evaluating 911
student’s ability in mastering what has been taught. 912

¹<https://github.com/langchain-ai/langchain>



(a) User interface for system evaluation at the beginning.



(b) User interface for system evaluation in progress.

Figure 7: User interface for system evaluation. The users select which course to learn and start learning by chatting to the system. The right-hand side shows the real-time course plan.

Topic	Gravity Waves
Course plan	<ul style="list-style-type: none"> 1. Introduction to Gravity Waves <ul style="list-style-type: none"> a. Definition and Key Concepts <ul style="list-style-type: none"> i. Differentiation between Gravity Waves and Gravitational Waves ii. Causes and Influences of Gravity Waves b. Factors Affecting Gravity Wave Formation <ul style="list-style-type: none"> i. Atmospheric Stability ii. Wind Shear iii. Topography and Surface Obstacles 2. Characteristics and Properties of Gravity Waves <ul style="list-style-type: none"> a. Wave Amplitude and Wavelength <ul style="list-style-type: none"> i. criteria for wave breaking ii. energy transfer and propagation b. Wave Speed and Frequency <ul style="list-style-type: none"> i. dispersion relation ii. role of buoyancy and ambient fluid properties 3. Observation and Detection of Gravity Waves <ul style="list-style-type: none"> a. Remote Sensing Techniques <ul style="list-style-type: none"> i. satellite imagery ii. lidar and radar <p>...</p>
Main system	<ul style="list-style-type: none"> 1. What is amplitude in the context of gravity waves? <ul style="list-style-type: none"> a) The distance between successive crests or troughs of a wave b) The maximum displacement of particles within a wave c) The rate at which wave energy is transferred vertically d) The disrupted, turbulent state of a wave 2. How does wavelength influence the behavior of gravity waves? <ul style="list-style-type: none"> a) It determines the spatial scale of the wave b) It represents the distance over which the wave repeats itself c) It influences the rate at which wave energy is transferred vertically d) It determines the amplitude of the wave 3. When does wave breaking occur in gravity waves? <ul style="list-style-type: none"> a) When the wave reaches its highest crest b) When the wave transitions from a regular pattern to a turbulent state c) When the force of gravity restores the disturbed equilibrium d) When the amplitude or wavelength exceeds certain thresholds
Ablation system	<ul style="list-style-type: none"> 1. What is the main focus of the course "Introduction to Gravity Waves"? <ul style="list-style-type: none"> a. Gravitational Waves b. Atmospheric Stability c. Causes and Influences of Gravity Waves d. Topography and Surface Obstacles 2. What are the factors affecting gravity wave formation? <ul style="list-style-type: none"> a. Wave Speed and Frequency b. Wind Shear c. Wave Amplitude and Wavelength d. Dispersion Relation 3. Which remote sensing techniques are used for detecting gravity waves? <ul style="list-style-type: none"> a. Satellite imagery b. In-situ measurements c. Balloon-borne instrument packages d. Lidar and radar








Table 8: Example quiz questions generated by different systems on “Gravity Waves”.

Table 9: A complete list of course topics and difficulty level by category used for system evaluation.

Category	Main Learning Objective	Difficulty Level
Art	Baroque architecture	2
	Installation art	2
	Impressionism	3
Psychology	Cognitive biases	2
	Obedience and conformity	3
	Operant conditioning	1
	Developmental psychology	5
Geology	Paleontology	2
	Earth's mantle	2
	Igneous rocks	2
	Mineralogy	2
	Volcanoes	2
	Geologic formations	3
	Geothermal energy	3
	Erosion	1
Computer Science	Sorting algorithms (e.g., bubble sort, merge sort)	2
	Linked lists	3
	Web development	4
	Computer architecture	4
Chemistry	pH scale	2
	Lewis dot structure	3
Daily Wondering	How do bees communicate and find their way back to the hive?	2
	Biological Circadian Rhythms (sleep-wake cycles)	2
	Gravity Waves (not to be confused with gravitational waves)	2
	Volcanic Eruptions	2
	Predator-prey relationships	2
	Formation of Coral Reefs	2
	How do magnets work, and what makes some materials magnetic?	2
	Change of Seasons	3
	Boom-and-bust cycles	3
	Reflection and Refraction of Light	3
	Formation of Crystals	3
	How does our brain store and retrieve memories?	3
	Hibernation	3
	Bioluminescence in Deep-Sea Creatures	3
	How do plants "know" which way to grow toward light?	3
	Why do some people have allergies, and how do they develop?	3
	Formation of Rainbows	3
	Cyberbullying	3
	How do rainbows form and why do they have different colors?	3
	Bioluminescence (e.g., glowing organisms in the ocean)	1
	Entanglement	1
	Aurora Borealis (Northern Lights)	1
	How do airplanes stay up in the air and maintain flight?	1
	Formation of Rain and Snow	1
	Why do some people have left-handed dominance while others are right-handed?	1
	Creation and Dissipation of Fog	1
	The uncertainty principle	4
	Hurricane formation	4
Personalized ads	4	
Umami taste	4	
Jet lag	5	
Fight-or-flight response	5	
Music	Ethnomusicology	2
	Rhythm and meter	2
	Baroque composers (e.g., Johann Sebastian Bach, Antonio Vivaldi)	2
	Polyphony	1
Physics	Quantum physics	2
	Newton's second law of motion	1
	Newton's first law of motion	1
	Snell's law of refraction	5
History	Medieval history	2

Category	Main Learning Objective	Difficulty Level
	Mongol Empire	3
	Ancient Egyptian civilization	3
	American history	4
Philosophy	Aesthetics	2
	Pragmatism	1
	Syllogism	5
Mathematics	Permutations and combinations	2
	Mathematical modeling	4
	Differential equations	4
Biology	Food chains and food webs	2
	Nitrogen fixation	3
	Plant biology	3
Sociology	Deviance and social control	3
	Social networks	3
Literature	Postcolonial literature and writers	3
	Literary theory	1
	Stream of consciousness	1
	Symbolism in poetry	4

Table 10: An example of dialogue and corresponding real-time course plan.

Dialogue	Course Plan
<p> Welcome to our Ethnomusicology course! In this exciting journey, we will explore the fascinating intersection of music, culture, and society. Our course is designed to introduce you to key concepts, research methods, and practical applications in this multidisciplinary field. Together, we'll delve into music's sociocultural context, learn various fieldwork techniques, and examine the diverse range of musical styles and genres around the world. By the end of this course, you will have a deeper appreciation for the rich tapestry of music and culture, as well as an understanding of how ethnomusicologists work to preserve and promote this diversity. Let's embark on this captivating exploration of music and culture together!</p>	<p>1. Introduction to Ethnomusicology a. Definition of Ethnomusicology i. History of Ethnomusicology ii. Disciplinary Approaches b. The Scope of Ethnomusicology i. Musical Elements ii. Cultural Context</p>
<p> In order to understand the history of ethnomusicology, it's important to first understand what ethnomusicology actually means. Ethnomusicology is the study of music within the context of its cultural and social setting. It involves examining how music is created, performed, and consumed by different groups of people around the world.</p> <p>The history of ethnomusicology can be traced back to the late 19th and early 20th centuries. During this time, scholars and researchers began to take an interest in understanding music beyond just its artistic and technical aspects. They wanted to explore how music was connected to the communities and cultures in which it was created.</p> <p>Ethnomusicology as a distinct field of study began to emerge in the mid-20th century. It was influenced by various disciplines such as anthropology, folklore studies, and musicology. Ethnomusicologists started conducting field research, traveling to different parts of the world to study and document traditional music practices.</p> <p>This approach differed from the traditional Western musicology, which largely focused on the study of classical music. Ethnomusicologists sought to understand and appreciate the diversity of musical traditions found globally. They recognized that music was not just a universal language, but also a reflection of the values, beliefs, and identities of the communities that created it.</p> <p>Ethnomusicologists use a wide range of research methods to study music in its cultural context. They may conduct interviews with musicians and community members, observe musical performances, and analyze the historical and social factors that shape musical practices. Through their work, ethnomusicologists aim to preserve and promote the rich tapestry of musical traditions from around the world.</p> <p>By studying the history of ethnomusicology, we gain a deeper understanding of how our appreciation and understanding of music has evolved over time. It helps us recognize the importance of cultural diversity and the role music plays in shaping our collective human experience.</p>	<p> </p> <p>2. Research Methods a. Fieldwork Techniques i. Participant Observation ii. Interviewing b. Analysis and Interpretation i. Transcription ii. Contextual Analysis</p> <p>3. Key Concepts a. Musical Diversity i. Cultural Differences ii. Music Styles and Genres b. Music, Identity, and Representation i. Social Identity ii. Music as Cultural Expression</p> <p>4. Ethnomusicology in Practice a. Applied Ethnomusicology i. Community-based Projects ii. Advocacy and Activism b. Teaching and Dissemination i. Academic Institutions ii. Public Outreach</p>
<p> Now that we have a better understanding of the history of ethnomusicology, let's explore the scope of this fascinating field. Ethnomusicology encompasses a wide range of topics and approaches, allowing us to delve into various aspects of music and culture.</p> <p>One important aspect of ethnomusicology is the study of musical instruments. Different cultures have unique instruments, each with its own distinct sounds, techniques, and meanings. Ethnomusicologists examine how these instruments are made, how they are played, and how they contribute to the overall musical experience.</p> <p>Another area of interest in ethnomusicology is the study of musical genres. Just as different cultures have their own instruments, they also have their own musical styles and genres. From traditional folk songs to contemporary pop music, ethnomusicologists analyze the characteristics, cultural significance, and historical development of these genres.</p> <p>Ethnomusicology also involves studying the role of music in different social contexts. Music has the power to bring people together, express emotions, and communicate cultural identities. Ethnomusicologists explore how music is used in rituals, celebrations, ceremonies, and everyday life, shedding light on the social functions and meanings behind musical practices.</p> <p>In addition to studying music within specific cultures, ethnomusicology also examines the interactions between different cultures and musical traditions. With globalization and migration, musical styles and practices often intersect and blend. Ethnomusicologists investigate these cross-cultural exchanges, exploring how music travels and adapts to new environments while retaining its original cultural and artistic elements.</p> <p>Lastly, ethnomusicology explores the significance of music in identity formation and cultural preservation. Music plays a crucial role in shaping individual and collective identities, reflecting the values, beliefs, and aspirations of communities. Ethnomusicologists study how music is used as a form of cultural expression and as a means of preserving heritage and traditions.</p> <p>By exploring the diverse scope of ethnomusicology, we can gain a deeper appreciation for the richness and complexity of music and its connection to culture. This multidimensional approach allows us to understand how music serves as a bridge between different societies and as a powerful tool for social change and understanding.</p>	<p> </p>



Now, let's discuss another important aspect of ethnomusicology: fieldwork techniques or research methods. Ethnomusicologists often engage in fieldwork, which involves immersing themselves in a particular culture or community to observe and document musical practices firsthand. Through fieldwork, ethnomusicologists are able to gather valuable information and insights about the music and its cultural context.

One fieldwork technique commonly used in ethnomusicology is participant observation. This method involves actively participating in the musical activities and events of a particular community while also observing and taking notes. By being a part of the musical experience, ethnomusicologists can gain a deeper understanding of the music's cultural significance and how it is performed within its social context.

In addition to participant observation, ethnomusicologists also conduct interviews as part of their fieldwork. They engage in dialogue with musicians, community members, and experts to gather information about the music, its origins, and its role in society. These interviews provide valuable insights into the cultural meanings and practices associated with the music, as well as the personal experiences and perspectives of those involved.

Ethnomusicologists may also employ audio and video recording techniques during their fieldwork. Recording performances, rituals, or everyday musical encounters allows for a more detailed analysis and documentation of the music. These recordings capture the nuances of the music, including its melodies, rhythms, and vocal expressions, which can be analyzed later in the research process.

In addition to these techniques, ethnomusicologists often utilize archival research. This involves studying historical records, manuscripts, recordings, and other existing materials related to the music and culture of interest. Archival research helps ethnomusicologists establish historical context, trace the evolution of musical practices, and understand how music has been shaped by social, political, and cultural forces over time.

Another important tool in ethnomusicological fieldwork is ethnographic writing. Ethnography involves writing detailed descriptions and narratives based on the researcher's observations and interactions in the field. Ethnomusicologists use ethnographic writing to convey the rich sensory experiences of music and to provide cultural context for their findings. This form of writing allows ethnomusicologists to share their research with others and contribute to the broader understanding of music and culture.

By utilizing these fieldwork techniques and research methods, ethnomusicologists are able to explore the complexities and intricacies of music in its cultural context. Through participant observation, interviews, recordings, archival research, and ethnographic writing, they gain a comprehensive understanding of the music's meaning, function, and impact on society. Fieldwork techniques are integral to ethnomusicology as they enable researchers to engage with music and culture in a holistic and immersive manner.

Got it



1. Introduction to Ethnomusicology
 - a. Definition of Ethnomusicology
 - i. History of Ethnomusicology
 - b. The Scope of Ethnomusicology
2. Research Methods
 - a. Fieldwork Techniques
 - b. Analysis and Interpretation
3. Key Concepts
 - a. Musical Diversity
 - b. Music, Identity, and Representation
4. Ethnomusicology in Practice
 - a. Applied Ethnomusicology
 - b. Teaching and Dissemination

ok so why is this subject important? Why can't we just appreciate music?



That's a great question! Ethnomusicology is an important subject because it allows us to go beyond simply appreciating music and understand it in a deeper and more meaningful way. Here are a few reasons why studying ethnomusicology is important:

1. **Cultural Understanding:** Ethnomusicology helps us understand how music is connected to a culture's history, values, beliefs, and social practices. By studying different musical traditions, we gain insight into the diverse ways in which people express themselves artistically and culturally. This understanding promotes cultural appreciation and fosters empathy and respect for different societies and their musical traditions.
2. **Preservation of Cultural Heritage:** Ethnomusicology plays a crucial role in preserving and documenting musical traditions that may be endangered or at risk of being lost. By studying and documenting these traditions, ethnomusicologists contribute to the preservation of cultural heritage and history. This ensures that future generations can learn from and appreciate these rich musical traditions.
3. **Cross-Cultural Dialogue:** Music has the power to bridge gaps and foster understanding between different cultures. Studying ethnomusicology allows us to explore the similarities and differences in musical practices across cultures. By engaging in cross-cultural dialogue and understanding, we promote global awareness and create opportunities for collaboration and exchange.
4. **Shedding Light on Social Issues:** Music often reflects the social, political, and cultural realities of a society. Ethnomusicology allows us to analyze and understand how music can be used to express and address social issues such as identity, inequality, and resistance. By examining the ways in which music is used as a form of cultural expression and activism, we can gain insight into the experiences and struggles of different communities.
5. **Personal Growth and Enrichment:** Studying ethnomusicology can also be personally enriching. It deepens our appreciation for the diversity of human creativity and expression, and allows us to develop a broader perspective on music and its cultural significance. It enhances our ability to critically analyze and interpret different musical traditions, and can also inspire creativity and innovation within our own musical pursuits.

Overall, studying ethnomusicology is important because it provides us with a deeper understanding and appreciation of music's role in shaping our world. It helps us recognize the power of music as a cultural and social force, and allows us to engage with diverse musical traditions in a meaningful and respectful manner.

alright, talk about some practical application then



Sure! Let's talk about a research method called participant observation, which is commonly used in ethnomusicology fieldwork.

Participant observation is a qualitative research technique that involves immersing oneself in a specific musical community or cultural context. Ethnomusicologists who use this method actively participate in the musical activities and events of the community they are studying, while also observing and taking notes on their experiences.

By actively participating in the musical practices of a community, researchers gain a firsthand understanding of the cultural and social significance of music within that particular context. This method allows ethnomusicologists to not only observe but also experience the music in its intended cultural and social setting.

Through participant observation, researchers can develop close relationships with community members, which can provide valuable insights into their perspectives, beliefs, and experiences. By engaging with the community as a participant, researchers are able to build trust and rapport, which leads to more authentic and meaningful data collection.

In addition to observing and participating, ethnomusicologists also document their experiences through fieldnotes. These fieldnotes capture their observations, thoughts, and reflections on the musical practices, performances, and interactions within the community. Fieldnotes provide a detailed record of the researcher's experiences, which can be analyzed and interpreted to gain a deeper understanding of the music and its cultural significance.

Participant observation allows researchers to uncover the nuances and intricacies of a musical tradition. It provides rich and detailed data that goes beyond mere descriptions, allowing for a more holistic analysis of the music and its cultural context.

This method also enables researchers to gain a more nuanced understanding of how music is learned, transmitted, and practiced within a specific community. By actively participating, researchers can witness firsthand the teaching methods, learning processes, and social dynamics that shape the musical practices of the community.

Overall, participant observation is a valuable research method in ethnomusicology because it allows researchers to immerse themselves in a musical community, experience the music firsthand, build connections with community members, and gain a deep understanding of the music's cultural and social significance.



Alright! Let's delve further into the importance and practical applications of participant observation in the field of ethnomusicology.

One of the key benefits of participant observation is its ability to capture the dynamic nature of music within a cultural context. By actively immersing themselves in the musical community, ethnomusicologists can witness the evolving nature of music and how it adapts to changes in society. This method allows researchers to identify and understand the cultural and historical factors that influence the development and transformation of musical traditions.

Another practical application of participant observation is its role in debunking stereotypes and misconceptions surrounding different musical communities. By directly engaging with the community as a participant, researchers can challenge preconceived notions and stereotypes, gaining a more accurate and nuanced understanding of the musical practices and their social significance. This can help in breaking down cultural barriers and fostering cross-cultural understanding and appreciation.

Participant observation also enables researchers to explore the social dynamics and power structures within a musical community. By immersing themselves in the community, ethnomusicologists can observe how music is used as a means of social interaction, communication, and identity formation. This method allows researchers to uncover the hierarchical relationships, gender dynamics, and other social dimensions that shape the musical practices and their broader cultural context.

Furthermore, participating in musical events and activities provides researchers with valuable opportunities for data collection and analysis. By actively engaging in the music-making process, researchers can gather detailed information about musical techniques, performance styles, and the underlying meanings behind the music. This firsthand experience enhances the authenticity and depth of research findings.

Participant observation can also contribute to the preservation and documentation of musical traditions. By actively participating in the community's musical activities, researchers can help to safeguard cultural heritage by documenting and recording performances, rituals, and traditions that might otherwise be lost or forgotten. This documentation serves as a valuable resource for future generations, ensuring the continuity and appreciation of diverse musical practices.

Lastly, participant observation can be a catalyst for meaningful collaborations and partnerships between researchers and community members. By immersing themselves in the community, ethnomusicologists can foster trust and mutual respect, leading to collaborative research projects that address the concerns and interests of the community. This participatory approach allows for a more ethical and inclusive research process, where the voices and perspectives of the community are represented and respected.

In conclusion, participant observation is a vital research method in ethnomusicology that provides researchers with a unique and in-depth understanding of musical traditions and their cultural significance. Through active participation, observation, and documentation, this method allows for a holistic analysis of music within its cultural context, while also promoting cross-cultural understanding, breaking down stereotypes, and ensuring the preservation of diverse musical practices.

good



1. Introduction to Ethnomusicology
 - a. Definition of Ethnomusicology
 - i. History of Ethnomusicology
 - b. The Scope of Ethnomusicology
2. Research Methods
 - a. Fieldwork Techniques
 - i. Participant Observation
 - ii. Importance and practical applications of participant observation
 - iii. Interviews
 - iv. Importance and practical applications of interviews
 - v. Audio and Video Recording Techniques
 - vi. Importance and practical applications of audio and video recording techniques
 - vii. Archival Research
 - viii. Importance and practical applications of archival research
 - ix. Ethnographic Writing
 - x. Importance and practical applications of ethnographic writing
 - b. Analysis and Interpretation
3. Key Concepts
 - a. Musical Diversity
 - b. Music, Identity, and Representation
4. Ethnomusicology in Practice
 - a. Applied Ethnomusicology
 - i. Practical applications of ethnomusicology in cultural understanding, preservation of cultural heritage, cross-cultural dialogue, shedding light on social issues, and personal growth and enrichment
 - b. Teaching and Dissemination

thx for your information, now maybe just give a quick overview of each method



Certainly! Here is a quick overview of the different research methods used in ethnomusicology fieldwork:

1. Participant observation: Involves actively participating in the musical activities and events of a community while also observing and taking notes. Provides a firsthand understanding of the cultural and social significance of music within a specific context.
2. Interviews: Engages in dialogue with musicians, community members, and experts to gather information about the music, its origins, and its role in society. Provides valuable insights into cultural meanings and practices.
3. Audio and video recording: Utilizes recording techniques to capture performances, rituals, and everyday musical encounters. Allows for detailed analysis and documentation of the music's nuances and characteristics.
4. Archival research: Involves studying historical records, manuscripts, recordings, and other existing materials related to the music and culture of interest. Establishes historical context and traces the evolution of musical practices.
5. Ethnographic writing: Involves writing detailed descriptions and narratives based on the researcher's observations and interactions in the field. Conveys the richness and cultural context of the music.

These methods, when used in combination, provide a comprehensive understanding of the music's meaning, function, and impact on society. They allow for an immersive and holistic exploration of music and its cultural context within ethnomusicology research.