

AI Tutor Enhanced with Prompt Engineering and Deep Knowledge Tracing

Radhika Makharia
AI R&D
TagHive Inc.
India
radhika@tag-hive.com

Yeoun Chan Kim
AI R&D
TagHive Inc.
Republic of Korea
ychan@tag-hive.com

Su Bin Jo
AI R&D
TagHive Inc.
Republic of Korea
lizzy@tag-hive.com

Min Ah Kim
AI R&D
TagHive Inc.
Republic of Korea
minah@tag-hive.com

Aagam Jain
Development Team
TagHive Inc.
India
aagam@tag-hive.com

Piyush Agarwal
Development Team
TagHive Inc.
India
piyush@tag-hive.com

Anish Srivastava
Planning Team
TagHive Inc.
Republic of Korea
yoloanish@gmail.com

Anant Vikram Agarwal
Planning Team
TagHive Inc.
Republic of Korea
anantvikram10@gmail.com

Pankaj Agarwal
TagHive Inc.
Republic of Korea
pankaj@tag-hive.com

Abstract—The evolving educational landscape necessitates creative solutions to address the demand for immediate and personalized academic support. This study explores the integration of prompt engineering of the OpenAI's Generative Pre-trained Transformer (GPT) and Deep Knowledge Tracing (DKT) to develop an AI tutor capable of shaping responses to students' knowledge levels, promoting a dynamic and adaptive learning experience. By leveraging Large Language Models (LLMs) like GPT-3.5 and integrating DKT, our AI tutor addresses the need for real-time, tailored academic assistance. LLMs serve as virtual instructors, explaining concepts and providing detailed solutions, while DKT ensures responses align with the student's knowledge level, optimizing challenge and engagement. Our research introduces an AI tutor that revolutionizes personalized learning experiences. Students can interact with the AI tutor by shaking their device during quizzes, initiating customized assistance and encouraging a deeper understanding of concepts, ultimately enhancing academic performance through individualized learning experiences.

Index Terms—prompt engineering, deep knowledge tracing, large language models, chatbots, virtual instructors, educational technology

I. INTRODUCTION

Artificial Intelligence (AI) is a multidisciplinary field that focuses on creating intelligent machines capable of performing tasks at a par with human intelligence. Initially, chatbots, which were dependent on rule-based systems and predefined responses, have been revolutionized by AI with advancements in Machine Learning (ML) and Natural Language Processing (NLP).

AI-powered chatbots provide round-the-clock support to learners, addressing queries and providing assistance at any time, overcoming barriers of time and location. AI algorithms can analyze learning patterns, preferences, and performance data to create personalized learning paths, suggesting appropriate resources and activities to enhance understanding and engagement.

The growth in the educational domain using AI-based chatbots, propelled by LLMs and ChatGPT, has revolutionized personalized learning experiences for students [1]. LLMs can serve as virtual instructors, explaining concepts, responding to questions, and providing detailed solutions to problems. Students who struggle with certain concepts or assignments may find this to be of particular value. They can be combined with a knowledge base to enable the model to retrieve precise data from it and produce well-informed responses to questions.

Considering a recent transition observed towards increased self-guided and asynchronous learning among students, an AI tutor is essential to address the pressing need for immediate and personalized academic support. By tailoring responses based on a student's knowledge level obtained using DKT, it can ensure that the assistance is at the appropriate difficulty level for optimal challenge and engagement, ultimately enhancing the learning process. The utilization of DKT along with prompt engineering of GPT-3.5 is a crucial innovation, enabling a dynamic and adaptive learning experience that can significantly boost academic performance.

Our AI tutor provides an immediate and accessible tool to address students' doubts and questions while studying. By shaking the device while attempting quizzes, students can quickly access personalized assistance, ensuring a seamless and engaging learning experience. Its ability to fit responses customized to students' current knowledge level of the topic encourages a deeper grasp of concepts by providing responses appropriate for their understanding.

II. RELATED WORKS

A. ChatGPT and GPT-3.5

The 21st century has witnessed a transformative shift in education driven by advanced AI technologies like generative AI, notably GPT-3.5, enabling sophisticated content creation across various media. GPT-3.5, with 175 billion parameters, has emerged as a pivotal NLP engine, impacting diverse

sectors ranging from education, healthcare, and finance to journalism, engineering, and economics, showcasing remarkable potential and innovative applications. GPT-3.5 leverages transformer architectures and is trained on extensive textual corpora to learn intricate linguistic patterns and relationships [2].

Educational AI chatbots are gaining traction due to their potential for interactive and personalized learning experiences with minimal teacher intervention, which is crucial in underprivileged areas where there are often fewer teachers or an overwhelming number of students per teacher. Recent advancements in LLMs and prompt engineering present a new avenue for developing educational chatbots. Reinforcement Learning and ChatGPT enhance reliable and interactive interactions with students. Although LLM-based chatbots face challenges regarding memory limitations, vulnerabilities, plagiarism, and understanding complex educational prompts, effective prompting strategies and exploring collaborative roles between teachers and chatbots are being investigated to maximize the educational chatbot effectiveness and ethical considerations [1, 3].

B. ChatGPT powered educational platforms

ChatGPT exhibits diverse educational applications. It showcases potential in generating algebraic hints comparable to those from human tutors [4], illustrating its ability to aid in mathematics learning. Additionally, it powers GPTutor [5], a programming tool facilitating code explanation and further enriching the educational experience for learners in the field of programming.

Likewise, Khanmigo is an AI-powered teaching assistant developed by Khan Academy that is powered by GPT-4 technology from OpenAI to provide personalized learning experiences, immediate feedback, and support to students. It utilizes GPT-4 to identify the student's learning goals, generate a learning plan tailored to the student's individual needs, and provide feedback on the student's work by identifying and correcting their mistakes, among other functions.

C. Deep Knowledge Tracing

The primary purpose of knowledge tracing is to model a student's skill acquisition. Initially, knowledge tracing relied on statistical analysis, such as Bayesian statistics. It used the Hidden Markov Model to represent a student's mastery of a skill as a binary variable [6]. However, knowledge tracing based on statistical analysis has limitations. The binary representation of a student's skill mastery does not effectively represent each individual's level of mastery. Furthermore, it requires expertise to monitor a student's growing knowledge.

With advancements in computer-aided education and machine learning, deep learning has assisted in a new trend in knowledge tracing DKT. DKT addresses the limitations of previous models by representing a student's level of understanding with decimal values and modeling a student's knowledge by identifying the best representation using a Recurrent Neural Network (RNN) [7].

Following the introduction of the transformer model, the transformer architecture has been implemented in DKT models, surpassing the performance of RNN-based models [8, 9]. While RNN-based DKT models aimed to predict a student's future performance per topic, transformer-based DKT models predict future performance per question. This allows for a more detailed modeling of an individual's level of understanding.

III. METHODOLOGY

A. GPT API with prompt engineering

The OpenAI API is a critical tool in modern natural language processing, enabling developers to seamlessly integrate cutting-edge language models like GPT-3.5 into a variety of applications, businesses, and services. These models allow users to perform a wide range of language-related tasks, ranging from text generation and question answering to conversation simulation and language translation, all while allowing the ability to customize instructions with metadata. At its core, the API employs two primary model types: completion models and chat completion models. The former, illustrated by models below GPT-3.5, performs as an autoregressive language model, predicting and producing text iteratively based on the input provided. When provided prompts, completion models excel at offering logical and contextually appropriate completions. On the other hand, chat completion models represent a specialized variant, optimized to simulate dynamic conversations by generating text in a conversational manner. These models are designed to process a stream of messages with different senders, like "system," "user," or "assistant," as input, leading to more interactive and interesting interactions. In our study, we used gpt-3.5-turbo, the most capable and cost-effective model from the GPT-3.5 series to build an interactive conversational experience for students.

A critical component of using language models like GPT-3.5 is prompt engineering, which involves creating carefully crafted instructions or queries to direct the generated output in an appropriate manner. One effective approach to prompt engineering within the GPT API framework entails integrating metadata directly into the prompt. This metadata encompasses user-specific preferences, contextual details, or any pertinent information that augments the model's understanding of the desired output. For example, every time a student asks for the AI tutor's assistance, the prompt is strategically structured to embed details about the user's grade, the subject they are studying, the academic curriculum being followed by them, their current understanding level of the topic to create a more user-centric interaction experience. Figure 1 depicts the systematic integration of a student's question, metadata, and knowledge levels derived from DKT to formulate a comprehensive prompt for the GPT API. Such directives can specify response length, desired tone, or even the level of detail required.

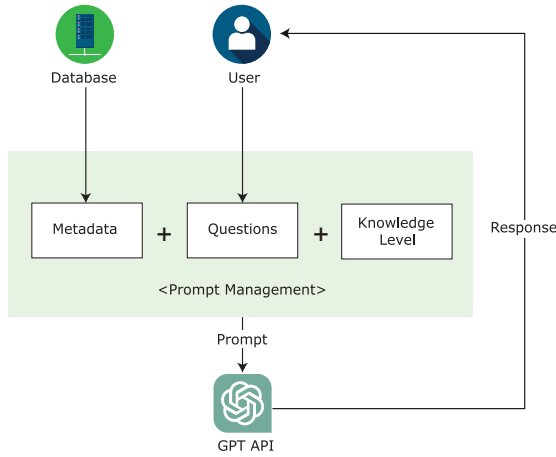


Fig. 1: Prompt management flowchart

B. DKT Model

In this experiment, our research team implemented the SAINT+ model from [9]. The SAINT+ model incorporates two features in encoder embeddings and three features in decoder embeddings. For the encoder embeddings, the model utilizes exercise ID and part features, where exercise ID represents the ID of each question, and "part" represents the subject. In the case of decoder embeddings, the model takes into account correctness, elapsed time, and lag time features, where each feature represents answer correctness, the time taken to solve a problem, and the time interval between interactions in the student's history. It is important to note that encoder embeddings encompass metadata, while decoder embeddings encompass action data. All of these features are embedded with the dimension of $d_{\text{model}} = 128$, and the features in the decoder embeddings include the start token embedding. Each encoder and decoder embedding is then combined with position embeddings.

The encoder layers of SAINT+ consist of multi-head self-attention layers followed by a feedforward layer. In this experiment, we employed 4 identical encoder layers. The decoder layers comprise two multi-head self-attention sub-layers followed by a feedforward layer. Similar to the encoder layers, residual connections are applied to all sub-layers. The output from the encoder layers is fed into each second multi-head self-attention layer of the 4 decoder layers. Both the encoder and decoder layers feature a feedforward sublayer with two linear transformations, incorporating a ReLU activation in the middle. The inner layer of the linear transformation has dimension of $d_{\text{model}} * 4$. After the decoder layers, the output undergoes a linear transformation, followed by the application of a softmax activation function to obtain the final probability of answering each question correctly. As reported in [9], the SAINT+ model was evaluated with the updated version of the EdNet dataset and achieved an accuracy (ACC) of 0.7252 and an area under the receiver operating characteristic curve (AUC) of 0.7914. Since the updated EdNet dataset is not

publicly available, we evaluated our model using the published EdNet dataset [10], and our implementation achieved an ACC of 0.7311 and an AUC of 0.7743.

Since the SAINT+ model predicts a student's future performance for each question, it is possible to obtain probabilities of correctly answering questions in a specific topic. We defined the mean value of these probabilities as the knowledge level of that topic.

IV. EXPERIMENTAL RESULTS

In this study, we conducted comprehensive experiments to assess the efficacy of integrating the GPT API into an AI tutor chatbot designed to assist students in their educational queries.

The stage one experiments were structured to evaluate various facets of the chatbot's performance. First, we assessed the AI tutor's ability to provide precise and correct answers to a diverse range of student queries. We focused on checking coherence and relevance of the responses by analyzing the logical flow and appropriateness of the chatbot's responses in an educational context.

Moving on to the second stage of our experiment, we conducted a rigorous evaluation of the chatbot's response accuracy, deliberately simulating diverse scenarios of student queries. Building upon the findings from stage one, where the AI tutor responded to all student questions regardless of subject or grade alignment, our objective was to confirm that the chatbot only addressed questions directly related to the subject matter and grade level.

Furthermore, we conducted an investigation into the chatbot's response patterns, tailoring them based on varying levels of student proficiency. We carefully analyzed and observed how the chatbot provided personalized assistance according to a student's grasp of a particular topic. The knowledge levels, extracted from the DKT framework, are represented by the variable x where $0 \leq x \leq 1$. When a student demonstrated a low level of understanding ($0 \leq x < \frac{1}{3}$) for a topic, the chatbot furnished comprehensive, step-by-step solutions to aid in better understanding. For those with a moderate level of understanding ($\frac{1}{3} \leq x < \frac{2}{3}$), the chatbot offered a streamlined outline of steps to encourage further progress. Moreover, for students with a high level of understanding ($\frac{2}{3} \leq x \leq 1$), the chatbot provided subtle hints to the problem, refraining from giving a direct answer. Figures 2, 3, and 4, visually represent the research findings, providing clear depictions of responses designed to suit low, medium, and high levels of understanding of individuals for the question "The lengths of diagonals of a rhombus are 24 cm and 18 cm respectively. What will be the length of each side of the rhombus?"

By systematically analyzing these aspects, we gained valuable insights into the capabilities and limitations of the AI tutor chatbot empowered by the GPT API. These findings not only shed light on the potential of this technology in enhancing the educational experience but also offer a foundation for further development of AI-driven educational tools, aiming to optimize learning experiences for students across different knowledge levels.

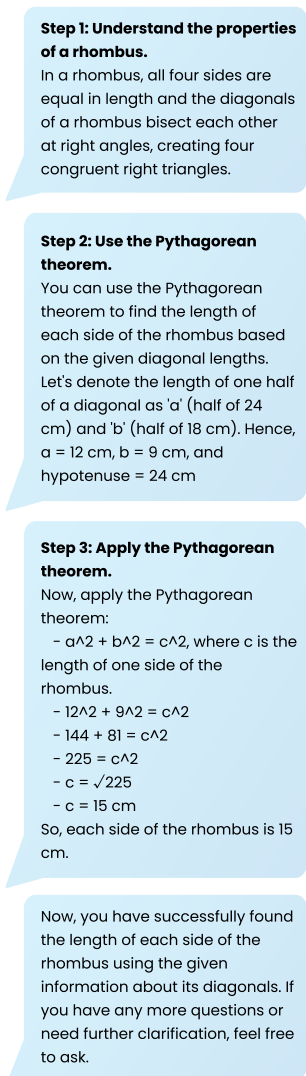


Fig. 2: Illustration of response from AI tutor for a student with a **low understanding level** of the topic, depicting a detailed step by step solution to the asked question.

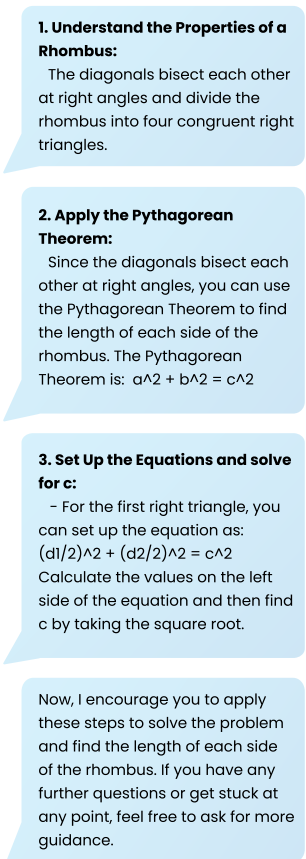


Fig. 3: Illustration of response from AI tutor for a student with a **medium understanding level** of the topic, depicting only an outline of solution to the asked question.

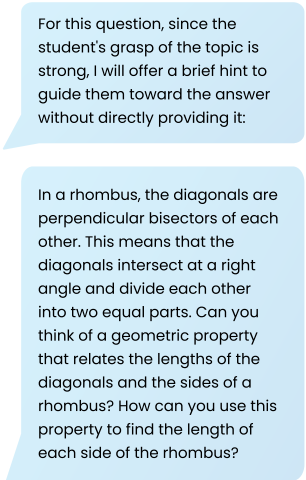


Fig. 4: Illustration of response from AI tutor for a student with a **high understanding level** of the topic, depicting a brief hint without a direct answer to the asked question.

V. IMPLEMENTATION

Class Saathi is an innovative tool that renders formative assessments seamless and data-driven for students by providing an enhanced learning experience. Saathi Tutor, as shown in Fig. 5, a vital component of the Saathi app, serves as a dynamic virtual assistant tailored to students' individual knowledge levels. Built on Flutter, like the Saathi front app, it seamlessly integrates into the platform, ensuring a consistent and user-friendly experience.

Accessing Saathi Tutor is a breeze; it is conveniently located on the Play screen. Users have the option to summon it either by tapping the dedicated Saathi Tutor button or by

shaking their mobile devices. The shake mechanism, powered by the 'shake' package¹, is finely tuned for responsiveness, with parameters like minimumShakeCount, shakeSlopTimeMS, shakeCountResetTime, and shakeThresholdGravity carefully configured for an intuitive user interaction.

Upon entering the Saathi Tutor screen, users are greeted with a personalized message: "Hey [User_Name], I see you are working on [Topic_Name]. How can I assist you with this topic?" This warm welcome sets the stage for a compassionate learning experience. Saathi Tutor relies on our robust API to provide relevant solutions based on user queries. If a student

¹<https://pub.dev/packages/shake>

remains uncertain, our UI provides a "Still Confuse" button. When tapped, it triggers the API to regenerate context based on the user's evolving understanding, ensuring continuous and meaningful assistance.

As the user's journey with Saathi Tutor concludes, and they feel satisfied with the provided solution, we archive the entire conversation. This data collection empowers our developers to gain valuable insights into user query patterns and continuously enhance the solution-providing capabilities of our API. Saathi Tutor is more than an assistant; it's a learning partner dedicated to personal growth and academic success.

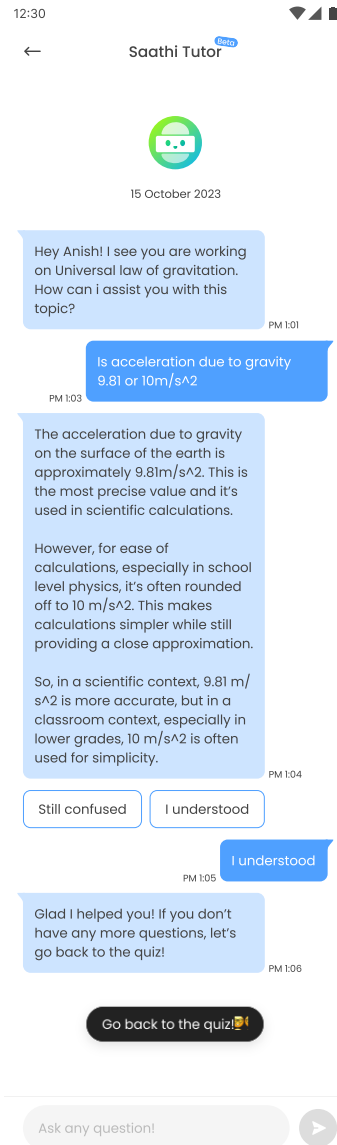


Fig. 5: An overview of implementation of Saathi Tutor in the Class Saathi app

VI. CONCLUSION

In this study, we undertook a comprehensive exploration, rigorously evaluating the integration of the GPT API through prompt engineering and Deep Knowledge Tracing into an AI tutor chatbot tailored to assist students with their educational inquiries. Our research and development unfolded in two proceeded. Initially, we carefully examined the chatbot's performance in terms of accuracy, coherence, and relevance in responding to a diverse array of student queries, aligning it with various educational contexts. Subsequently, we refined our assessment to ensure the chatbot's personalization by simulating diverse knowledge levels of students, emphasizing subject and grade relevance.

Looking ahead, this research lays the groundwork for further advancements in AI-driven educational tools. One such avenue involves addressing out-of-syllabus student queries by creatively linking them to the current subject injecting interest and significance into the conversation. This strategy not only fosters students' natural curiosity but also guarantees a lively and interesting teaching experience. Another promising direction involves the implementation of a robust database to store all questions and responses. By employing ML similarity algorithms, we can efficiently retrieve previously answered questions from the database, minimizing reliance on the GPT API for recurring inquiries. Furthermore, fine-tuning the chatbot's understanding of student preferences and learning patterns will undoubtedly elevate the educational support provided. Overall, the envisioned future enhancements aim to elevate the AI tutor chatbot to offer seamless, personalized, and engaging educational support to the students.

ACKNOWLEDGMENT

This work was supported by the Technological Innovation R&D Program (S3280745) funded by the Ministry of SMEs and Startups (MSS, Korea).

REFERENCES

- [1] S. S. Gill, M. Xu, P. Patros, H. Wu, R. Kaur, K. Kaur, S. Fuller, M. Singh, P. Arora, A. K. Parlikad *et al.*, "Transformative effects of chatgpt on modern education: Emerging era of ai chatbots," *Internet of Things and Cyber-Physical Systems*, vol. 4, pp. 19–23, 2024.
- [2] D. Baidoo-Anu and L. O. Ansah, "Education in the era of generative artificial intelligence (ai): Understanding the potential benefits of chatgpt in promoting teaching and learning," *Journal of AI*, vol. 7, no. 1, pp. 52–62, 2023.
- [3] C. Koyuturk, M. Yavari, E. Theophilou, S. Bursic, G. Donabauer, A. Telari, A. Testa, R. Boiano, A. Gabbiadini, D. Hernandez-Leo *et al.*, "Developing effective educational chatbots with chatgpt prompts: Insights from preliminary tests in a case study on social media literacy," *arXiv preprint arXiv:2306.10645*, 2023.
- [4] Z. A. Pardos and S. Bhandari, "Learning gain differences between chatgpt and human tutor generated algebra hints," *arXiv preprint arXiv:2302.06871*, 2023.

- [5] E. Chen, R. Huang, H.-S. Chen, Y.-H. Tseng, and L.-Y. Li, "Gptutor: a chatgpt-powered programming tool for code explanation," *arXiv preprint arXiv:2305.01863*, 2023.
- [6] A. T. Corbett and J. R. Anderson, "Knowledge tracing: Modeling the acquisition of procedural knowledge," *User modeling and user-adapted interaction*, vol. 4, pp. 253–278, 1994.
- [7] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. J. Guibas, and J. Sohl-Dickstein, "Deep knowledge tracing," *Advances in neural information processing systems*, vol. 28, 2015.
- [8] S. Pandey and G. Karypis, "A self-attentive model for knowledge tracing," *arXiv preprint arXiv:1907.06837*, 2019.
- [9] D. Shin, Y. Shim, H. Yu, S. Lee, B. Kim, and Y. Choi, "Saint+: Integrating temporal features for ednet correctness prediction," in *LAK21: 11th International Learning Analytics and Knowledge Conference*, 2021, pp. 490–496.
- [10] Y. Choi, Y. Lee, D. Shin, J. Cho, S. Park, S. Lee, J. Baek, C. Bae, B. Kim, and J. Heo, "Ednet: A large-scale hierarchical dataset in education," in *Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, July 6–10, 2020, Proceedings, Part II 21*. Springer, 2020, pp. 69–73.