Agentic AI-Driven Multimodal Personalized Tutoring with Dynamic Knowledge Retrieval

Alfonsus Rodriques Rendy (121040014)
School of Data Science
The Chinese University of Hong Kong, Shenzhen
121040014@cuhk.edu.cn

Abstract

The rising demand for interactive, personalized education underscores the limitations of traditional learning tools, driving the need for scalable AI-driven solutions. IntelliTutor is an advanced multimodal AI tutoring system that provides dynamic, step-by-step guidance tailored to individual needs. Users can upload educational materials in PDF format, which are processed into a vector database for efficient retrieval. By combining large language models (LLMs), retrieval-augmented generation (RAG), advanced text-to-speech (TTS), and automatic speech recognition (ASR) technologies, IntelliTutor enables adaptive, engaging learning experiences. The system offers multimodal explanations, including text and voice-guided study steps, while retaining conversational memory for continuity. Designed for accessibility, adaptability, and privacy, IntelliTutor is a versatile tool for self-directed learners, redefining personalized education system.

1 Introduction

Recent advancements in artificial intelligence, particularly in large language models (LLMs), have transformed how machines understand, retrieve, and generate multimodal information. Among these breakthroughs, in-context learning stands out as a mechanism that enables LLMs to dynamically adapt their responses based on the examples and instructions provided in their input prompts [1]. This ability has expanded the versatility of LLMs across diverse domains. Additionally, Retrieval-Augmented Generation (RAG) techniques have addressed critical limitations such as inaccuracies and hallucinations by grounding LLM responses in external data sources [2].

Recent advancements in artificial intelligence, particularly in large language models (LLMs), have revolutionized the way machines retrieve, understand, and generate information. Techniques such as in-context learning enable LLMs to adapt their responses dynamically based on input prompts, significantly expanding their versatility across various domains [1]. Meanwhile, Retrieval-Augmented Generation (RAG) addresses critical challenges like inaccuracies and hallucinations by grounding responses in reliable external data sources [2].

Building on these advancements, the agentic design pattern has emerged as a key innovation. This paradigm empowers LLMs to function as autonomous agents capable of managing complex workflows, including information retrieval, task decomposition, and iterative refinement. By dynamically orchestrating multi-step processes, such systems can deliver precise and adaptive outputs, especially in applications requiring deeper task comprehension and user interaction.

The educational field stands to benefit significantly from such innovations, as interactive automated tutoring systems can enhance student engagement and learning outcomes. To meet this demand, we present IntelliTutor, an AI-driven framework that integrates advanced LLMs, personalized materials, and conversational interfaces. By leveraging agentic design principles, IntelliTutor delivers accurate, context-aware explanations and fosters deeper learner interaction, offering a modern solution for personalized education.

2 Related Works

General LLM as a personalized tutor The application of large language models (LLMs) in education has gained significant attention due to their potential as effective learning companions. Initially, their use in tutoring systems was limited to basic question-and-answer interactions. However, recent research highlights the growing capabilities of LLMs in enhancing learning outcomes [3, 4]. A notable example is NewtBot, a physics education chatbot tailored for secondary school students, which leverages GPT-3.5 [5]. This study demonstrated that NewtBot not only improved students' learning experiences but also left a positive impression on its users.

Equipping LLM-based tutoring system with RAG Large Language Models (LLMs) often produce plausible but factually incorrect outputs, a phenomenon known as hallucination [6]. To address this, recent efforts integrate Retrieval-Augmented Generation (RAG) for better information retrieval in knowledge-intensive applications like automated tutoring systems. For instance, Dong Chenxi's AI Tutor, built using the OpenAI Assistants API, employs RAG to access course-specific materials, generating accurate, cited answers and achieving high satisfaction rates in tests with 50 diverse questions [7]. Similarly, Horia Modran et al. use the LlamaIndex framework to manage a vector database of educational content, enabling personalized assistance and efficient retrieval [8]. These approaches demonstrate RAG's potential to enhance the accuracy and relevance of LLM-based tutoring, fostering personalized and effective learning experiences.

Fostering Engagement Through Agentic AI Engagement is essential for effective learning but is often missing in LLM-based tutoring systems. Lieb et al. emphasize behaviors that foster engagement, such as using an enthusiastic tone, providing clear explanations, offering examples and analogies, and delivering constructive feedback [5]. These strategies can be implemented through an agentic AI framework, as proposed in this report, enabling LLMs to operate autonomously, challenge users proactively, and adapt to their needs. Unlike traditional tutoring system, IntelliTutor not only formulates follow-up questions but also incorporates read-aloud features to enhance system interactivity and user engagement.

3 Approach and Implementation

We introduce IntelliTutor which combines state-of-the-art LLM techniques including efficient prompting, retrieval augmented generation, combined with agentic mechanism. The whole system is divided into four module: tutor, chat, knowledge, and speech module.

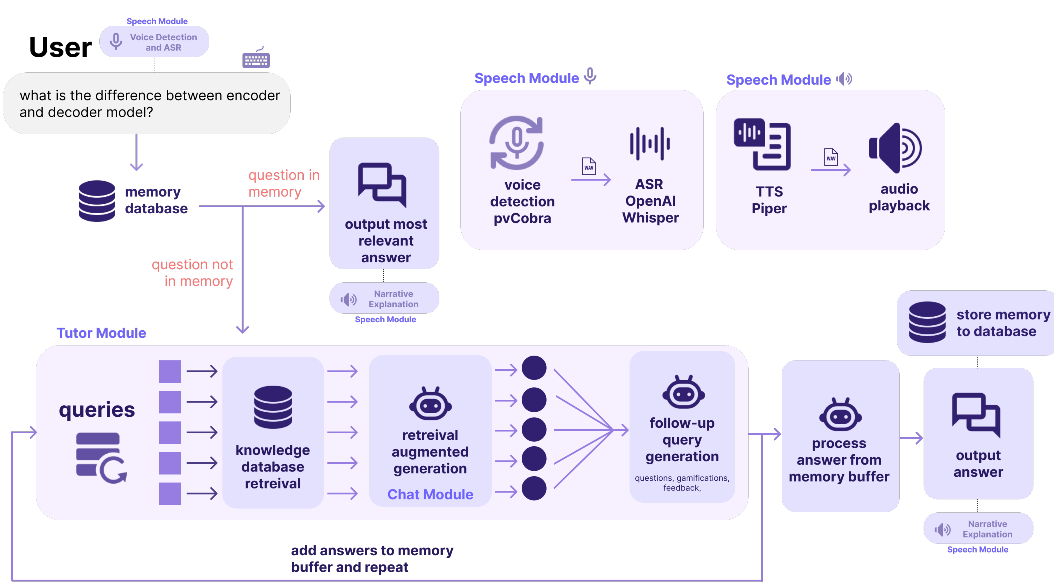

Figure 1: Detailed Pipeline of IntelliTutor

3.1 Module Implementation

The knowledge module manages knowledge and memory retrieval using two databases: a memory database for user questions and answers, and a knowledge database with curated materials. The chat module enables interactions with the LLM, focusing on follow-up queries like additional questions or user feedback to enhance learning. The current prototype supports follow-up questions as a proof of concept, retrieves relevant database information for accurate answers, and summarizes multi-turn interactions.

The speech module handles voice input and output using pvCobra and Whisper API for transcription and includes text-to-speech for accessibility. The tutor module integrates all components, employing a multi-turn self-conversation approach where the agent generates follow-up queries iteratively, processed via Retrieval-Augmented Generation (RAG), to ensure comprehensive responses.

3.2 NLP Techniques: RAG, Efficient Prompting, and Agentic

Several advanced techniques are employed to enhance the quality of the tutoring system. As noted in prior research, Retrieval-Augmented Generation (RAG) improves the accuracy of information generated by the LLM. Additionally, instruction-based prompting guides the system to provide credible sources for the information it generates. A newly introduced approach, the agentic pipeline, enables the system to focus on more specific questions, helping users gain a deeper understanding of the topic while fostering curiosity to explore further.

4 Experiment

4.1 Data and Evaluation Metric

Several curated PDF notes, sourced from various SLP course materials on NLP applications, were compiled to serve as a knowledge base for the evaluation. The evaluation was conducted quantitatively to assess three quality metrics of the answers: (1) correctness and credibility, (2) engagement, and (3) depth of explanation.

4.2 Evaluation and Analysis

Several answer samples were collected to compare the performance of IntelliTutor with other system approaches, revealing three key strengths of the implemented system. First, IntelliTutor excels at retrieving personalized, credible answers from a curated knowledge base. This capability enhances specificity, particularly for multi-domain concepts that require deeper contextual understanding.

Second, IntelliTutor effectively explores specific topics raised by users through its multi-turn self-conversation approach. This enables it to generate more detailed subtopics and clarify previously generated answers, providing a richer and more comprehensive response. Finally, IntelliTutor assists users in navigating source materials by citing information from the uploaded knowledge documents. These citations not only enhance answer credibility but also guide users to the original source, which may offer richer information, including textual and visual content, to deepen understanding.

5 Conclusion and Future Works

This report introduces IntelliTutor, an AI-driven tutoring framework that leverages advanced techniques like Retrieval-Augmented Generation (RAG), efficient prompting, and agentic design to deliver personalized and interactive educational experiences. Its modular architecture enables precise, context-aware explanations while fostering engagement and deeper understanding through multi-turn self-conversation and citation-based navigation of knowledge sources.

Future work will enhance interactivity by incorporating progress monitoring, quizzes, and games, as well as improving the question database with topic mapping and clustering to help users visualize and explore knowledge more effectively. These advancements aim to solidify IntelliTutor as a dynamic, scalable solution for modern personalized learning.

References

[1] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners, 2020.

[2] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks, 2021.

[3] Alessandro Vanzo, Sankalan Pal Chowdhury, and Mrinmaya Sachan. Gpt-4 as a homework tutor can improve student engagement and learning outcomes, 2024.

[4] Fernando Flores Limo, David Tiza, Maribel Roque, Edward Herrera, José Muñoz Murillo, Jorge Jinchuña Huallpa, Victor Ariza Flores, Alejandro Castillo, Percy Peña, and Christian Carranza. Personalized tutoring: Chatgpt as a virtual tutor for personalized learning experiences. *Przestrzen Spoleczna*, 23:293–312, 06 2023.

[5] Anna Lieb and Toshali Goel. Student interaction with newtbot: An llm-as-tutor chatbot for secondary physics education. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, CHI EA '24, New York, NY, USA, 2024. Association for Computing Machinery.

[6] Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, and Ting Liu. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, November 2024.

[7] Chenxi Dong. How to build an ai tutor that can adapt to any course and provide accurate answers using large language model and retrieval-augmented generation, 2024.

[8] Horia Modran, Ioana Corina Bogdan, Doru Ursutiu, Cornel Samoila, and Paul-Livius Modran. Llm intelligent agent tutoring in higher education courses using a rag approach, 07 2024.