Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 6, 2026

AGENTIC AI-BASED INTELLIGENT STUDY ASSISTANT USING LL MS AND VECTOR DATABASES

Zaviyar Hasnain Bhutta[*1], **Dr. Aatif Hussain**[2]

[*1,2]*Department of Computer Science University of Engineering & Technology, Lahore*

DOI: https://doi.org/10.5281/zenodo.20758135

Keywords
Agentic AI, Large Language Models, Retrieval Argument Generation (RAG), Natural Language Processing, Semantic Search, Intelligent System Assistant, Vector Database, Intelligent Systems, Education Technology.

Article History
Received: 19 April 2026
Accepted: 01 June 2026
Published: 16 June 2026

Corresponding Author: *
Zaviyar Hasnain Bhutta
zaviyarhasnain1@gmail.com

Abstract
By introducing intelligent and automated learning solutions, artificial intelligence (AI) has transformed contemporary educational systems. An Agentic AI-Based Intelligent Study Assistant is presented in this study. It makes use of Vector Databases and Large Language Models (LL Ms) to provide students with intelligent, context-aware, and individualized academic assistance. Natural language interaction is used to answer questions, summarize study materials, make notes, and help students learn more quickly with the proposed system. The primary focuses of the research are the creation and implementation of an intelligent system that is able to comprehend user input, retrieve relevant information through vector-based semantic search, and generate precise responses through advanced AI models. By combining the capabilities of language generation and external knowledge retrieval, the integration of Retrieval-Augmented Generation (RAG) techniques enhances the relevance and quality of responses. The system architecture includes components such as user interface, Agentic workflow, embedding models, vector database, and LLM integration. The model that has been proposed aims to make education more adaptable, to make learning easier, and to make students more productive. According to experimental analysis, the intelligent assistant performs better than conventional keyword-based systems in terms of response accuracy, contextual understanding, and user interaction. This study demonstrates how intelligent assistants can support contemporary learning environments through automation, personalization, and effective knowledge retrieval, as well as the growing role Agentic AI systems are playing in education.

I. INTRODUCTION

Artificial Intelligence (AI) has rapidly advanced over the past ten years to become one of the most influential technologies in current computing systems. The incorporation of artificial intelligence (AI) into educational environments has significantly altered conventional methods of learning by introducing intelligent, adaptive, and automated solutions. Large Language Models (LL Ms) and Agentic AI systems have attracted a lot of attention due to their ability to comprehend natural language, produce responses that are human-like, and carry out complex reasoning tasks.

Students rarely receive context-aware and personalized learning assistance from keyword-based search tools or traditional educational systems. Finding accurate study materials, comprehending intricate concepts, and effectively managing academic tasks are typically challenges for students. To address these limitations, intelligent study assistants powered by AI technologies are becoming increasingly important in modern education.**[1]**

An "Agentic AI-Based Intelligent Study Assistant Using LL Ms and Vector Databases" is presented

in this study to provide intelligent academic support through semantic information retrieval and natural language interaction. The proposed system combines the capabilities of Large Language Models with Vector Databases and Retrieval-Augmented Generation (RAG) techniques to enhance contextual understanding and improve response accuracy.

The proposed assistant, in contrast to conventional systems, is capable of retrieving pertinent knowledge from stored educational content and providing users with responses that are meaningful and individualized. The architecture of the proposed intelligent system includes multiple components such as user interaction modules, embedding models, semantic search mechanisms, vector storage, and AI-based response generation. The integration of vector databases enables efficient semantic retrieval of educational content, while the Agentic AI workflow allows the system to perform autonomous and intelligent decision-making processes.[2]

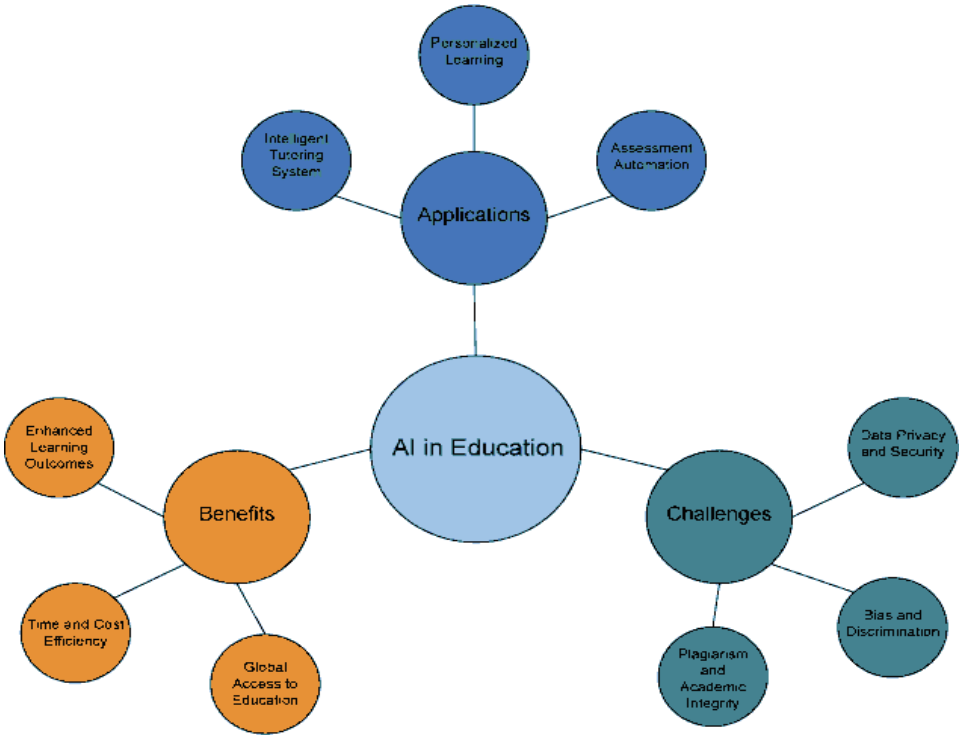

Figure 1: Era of Artificial Intelligence in AI

The primary objective of this research is to design and implement a smart educational assistant capable of improving student learning experiences, reducing dependency on manual searching, and enhancing academic productivity. In addition, the research demonstrates how cutting-edge AI technologies can help create personalized and effective learning environments and the growing significance of intelligent systems in education. The remainder of this paper is organized as follows: Section II covers the literature review and related work; Section III covers the proposed methodology and system architecture; Section IV covers the details of the implementation and the experimental analysis; and the final section covers the conclusion of the research and potential improvements.[3]

II. LITERATURE REVIEW

One of the most widely studied areas of

contemporary computer science is artificial intelligence (AI), particularly education, automation, and intelligent systems. Systems that can imitate human intelligence, comprehend natural language, and assist users in complex decision-making processes have been the focus of research over the past ten years. The field of intelligent systems has undergone significant change thanks to the development of RAG frameworks, vector databases, and Large Language Models (LL Ms).

Semantic information retrieval methods, educational technologies, agent-based systems, and AI-powered intelligent assistants are all topics that will be examined in this literature review. It also highlights the limitations of traditional systems and explains how modern AI-based approaches attempt to overcome these challenges.[4]

Evolution of Intelligent Systems in Education

Early educational systems were based on rule-based expert systems, which responded to user queries using predefined rules and logic. These systems lacked adaptability and were unable to answer complex or ambiguous questions. MYCIN and DENDRAL, for instance, demonstrated early expert reasoning success but lacked adaptability and scalability.Data-driven approaches began to be incorporated into educational systems as machine learning progressed. They struggled with unstructured data and naturally-language written queries.[5]

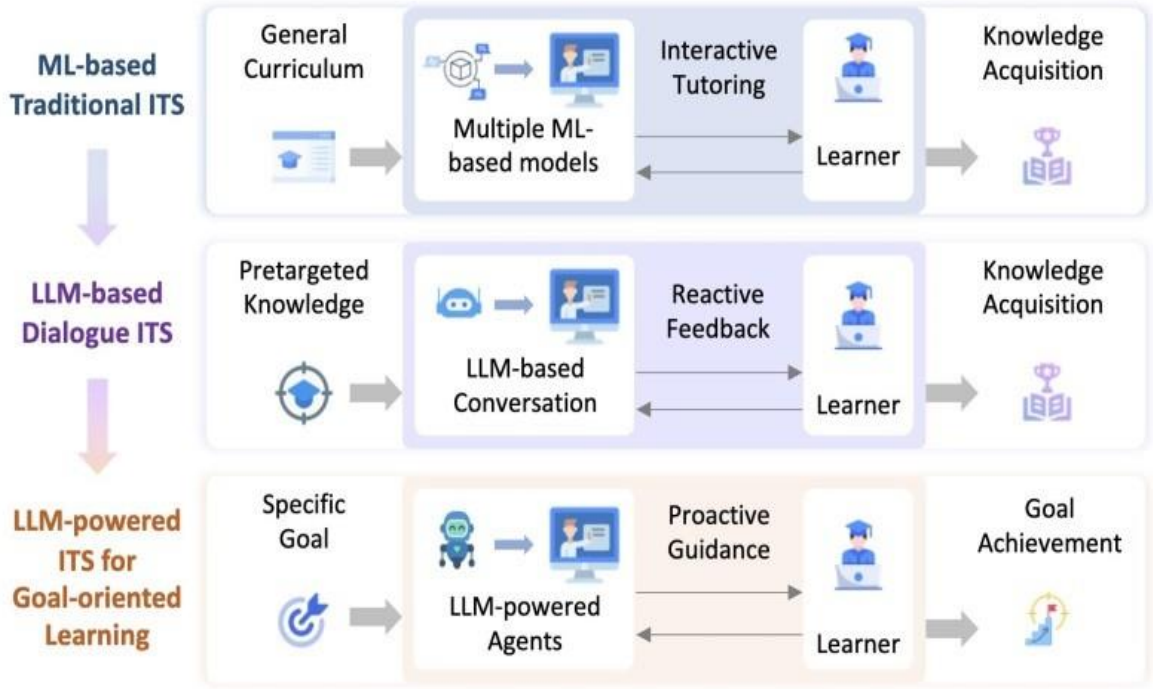

Figure 1: Comparison of three types of ITS Paradigms.

Figure 2: Evolution of Artificial Intelligence in Educational Systems

Emergence of Natural Language Processing (NLP)

The field of natural language processing, also known as NLP, played a crucial role in bridging the communication gap between machines and humans. For text representation, early NLP systems utilized methods like keyword matching, bag-of-words models, and TF-IDF. Despite their usefulness for basic information retrieval, these techniques lacked comprehension of semantics.

In the years that followed, researchers developed word embedding techniques like Word2Vec (Kimonos ET AL., 2013) and Glove (Pennington ET AL., 2014). Machines were able to comprehend the contextual relationships that exist between words thanks to these methods. These embed-ding's significantly improved the performance of search and recommendation systems.

Long-context reasoning, multi-step queries, and knowledge-intensive tasks were still challenges for traditional NLP systems despite these advancements. Due to this restriction, transformer-based architectures were developed..[6]

Rise of Transformer Models and Large Language Models (LLMs)

The introduction of the Transformer architecture by Aswan ET Al. (2017) marked a revolutionary shift in NLP research. Transformers introduced the idea of self-attention mechanisms and parallel processing of input data, making it possible for models to comprehend text's long-term dependencies. However, LL Ms also have limitations, including hallucination (generating false information), outdated knowledge, and high computational cost. To overcome these issues, researchers introduced hybrid systems that combine LL Ms with external knowledge sources such as vector databases.[7]

Vector Databases and Semantic Search

Vector databases have become a crucial component in modern AI systems. Unlike traditional databases that rely on exact keyword matching, vector databases store data in the form of high-dimensional embedding. Because these embed-ding's represent the semantic meaning of text, search results can be more precise and contextually aware. Popular vector database solutions such as FAISS, Pine cone, and We vitiate have been widely used in AI applications. Similarity search based on cosine distance or Euclidean distance is made possible by these systems, allowing relevant information to be retrieved even when exact keywords are not present.

Research in semantic search shows that vector-based retrieval significantly improves information retrieval accuracy compared to traditional keyword-based systems.

Retrieval-Augmented Generation (RAG)

RAG systems use relevant documents from an external database as context for generating responses rather than solely relying on pee-trained knowledge. Lewis ET AL. (2020) introduced RAG as a method to improve factual accuracy and reduce hallucinations in language models.

The system works in two stages:

Phase of Retrieval:

User queries are used to pull relevant documents from a vector database. Generation Phase: The retrieved documents are passed to an LLM, which generates a final response.

In educational assistants, chat bots, and question-answering systems, this strategy has proven to be extremely successful.[8]

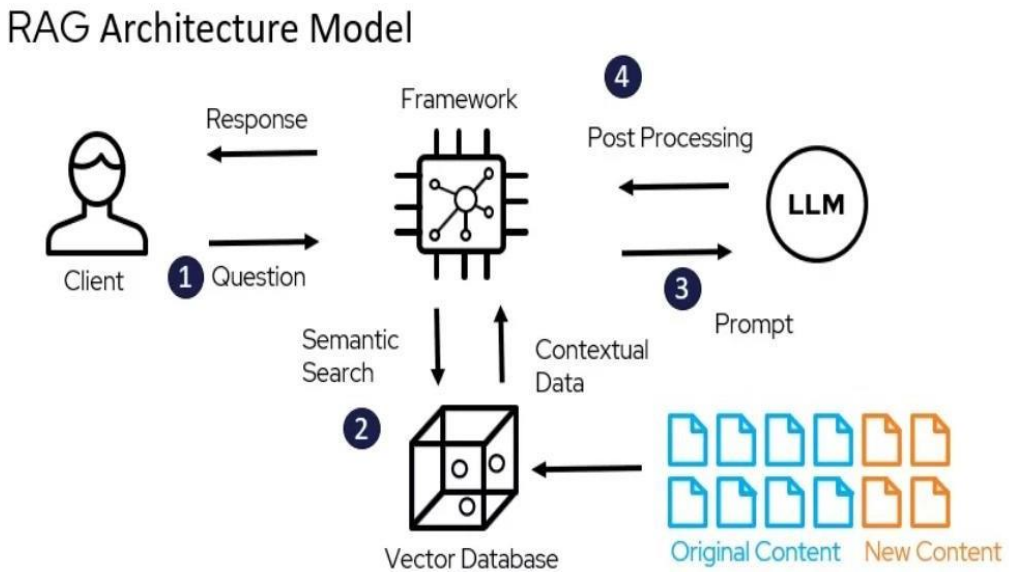

Figure 3: Retrieval-Augmented Generation (RAG) Architecture using Vector Database and LLM

Agentic AI Systems

Agentic AI refers to systems that can perform autonomous decision-making, planning, and execution of tasks. Unlike traditional AI models that only respond to inputs, Agentic systems can break down complex tasks into sub-tasks and execute them sequentially.

Systems can behave more like intelligent assistants than just chat bots when they combine LL Ms with tool-use capabilities, according to recent autonomous agent research. These systems can reason, plan, and interact with external tools such as API s, databases, and search engines.

Baby AGI, Lang Chain agents, and Auto GPT are examples of Agentic frameworks.

Modern systems integrating LL Ms and vector databases offer a more advanced solution by combining natural language understanding with knowledge retrieval. These systems can adapt to different subjects, answer complex queries, and provide personalized learning experiences[9].

Limitations in Existing Research

Despite significant progress, several limitations exist in current AI-based educational systems: Lack of real-time knowledge updates in many LL Ms

Fabrication of false information and hallucinations high storage and computation requirements Limited explain ability in decision-making processes

Reliance on large, previously trained models. These challenges highlight the need for more efficient, reliable, and context-aware intelligent systems.

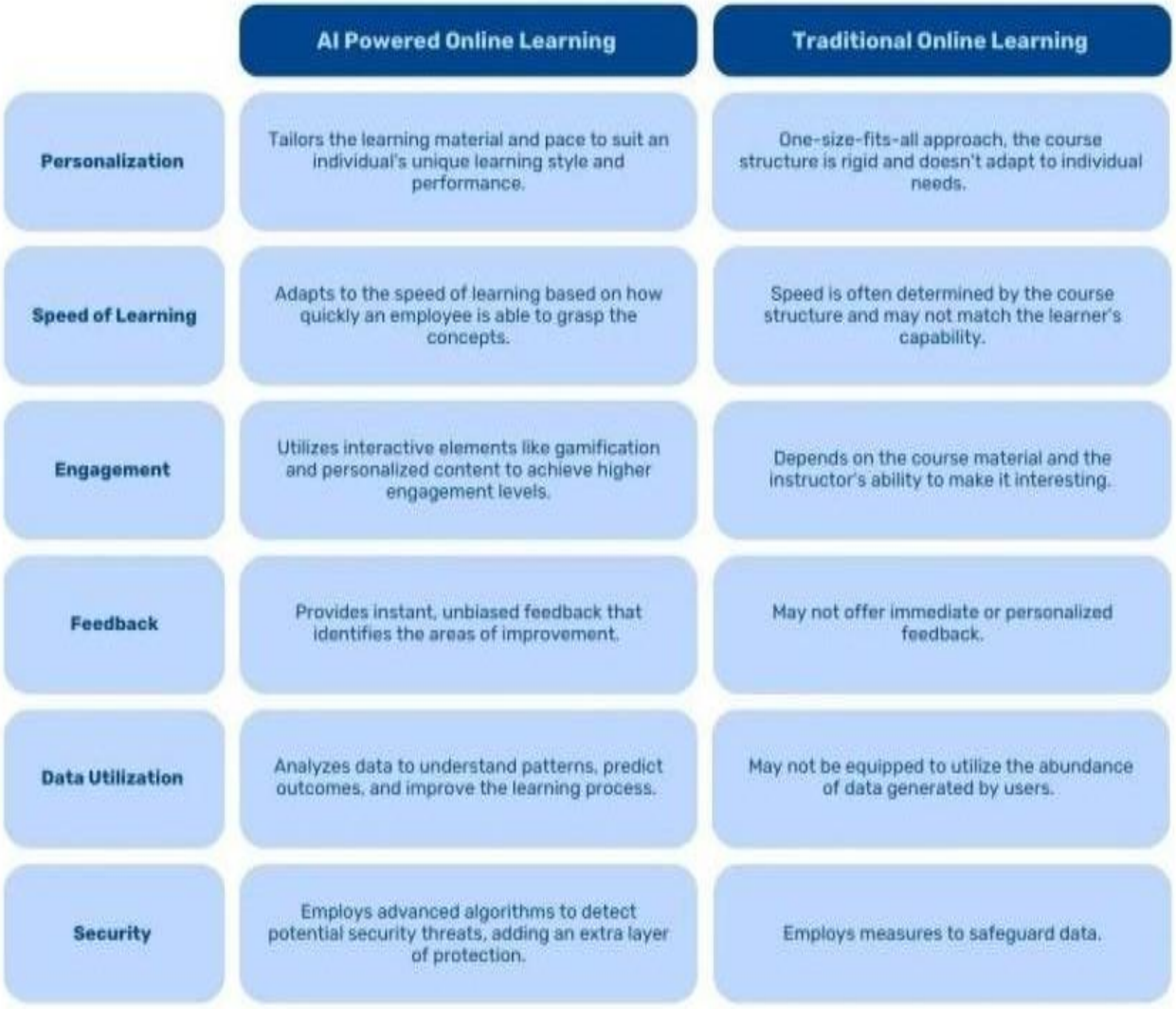

Figure 4: Comparison between Traditional Systems and AI-Based Intelligent Study Assistants

Research Gap

According to the reviewed literature, fully integrated Agentic systems designed specifically for education are still lacking, despite the significant progress made by individual technologies like LL Ms, vector databases, and RAG systems. The majority of current systems do not have a unified architecture that combines autonomy, reasoning, and contextual learning assistance within a single framework. Instead, they either focus on language generation or retrieval accuracy.[10]

III. METHODOLOGY

Data Collection

The proposed Agentic AI-Based Intelligent Study Assistant system utilizes a combination of structured and unstructured educational data to support intelligent query answering and contextual learning. In order to guarantee both diversity and accuracy, the data are gathered from a variety of educational sources that are both domain-specific and publicly accessible.

Content for education that can be found on the web, like tutorials, documentation, and knowledge articles. Sample question-answer datasets for training and evaluation purposes

These sources provide a rich knowledge base that helps the system understand different concepts across multiple domains[11].

Dataset Preparation

The collected data is processed and converted into machine-readable format before being used in the system. The prepossessing steps include:
Extracting text from online and PDF sources Getting rid of content that doesn't matter, like advertisements and headers and footers Ionization of text into more manageable pieces for faster processing Text normalization (lowercase and elimination of punctuation) Splitting long documents into smaller segments for embedding generation
In order to preserve context during retrieval, the data is segmented into fixed-length chunks following pre processing.

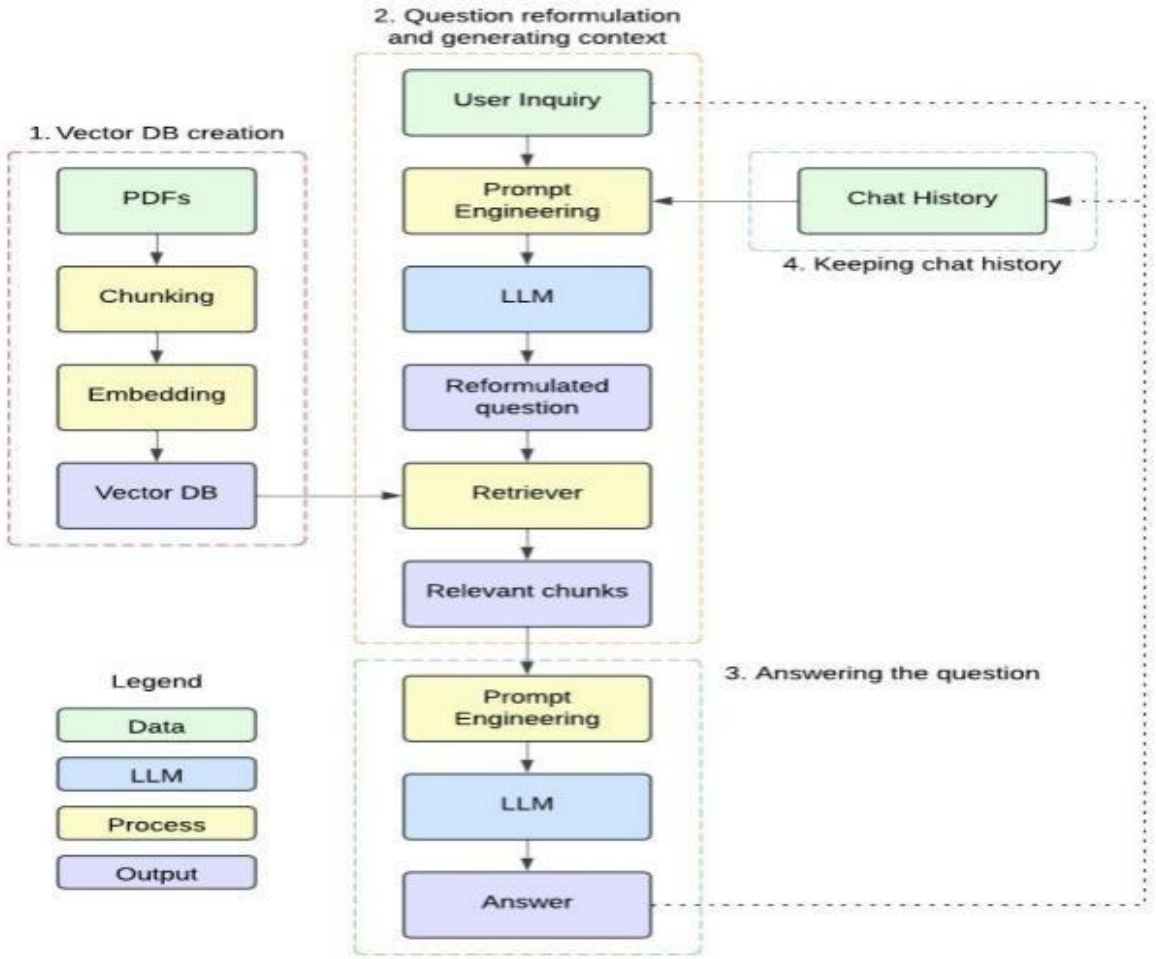

Figure 5: System Architecture of the Proposed Agentic AI-Based Intelligent Study Assistant

Embedding Generation

Using per-trained embedding models like Open AI Embedding or Sentence-BERT, each text chunk is transformed into high-dimensional vector representations. These embedding capture the semantic meaning of the text rather than relying on exact keyword matching[12].

Vector Database Storage

A vector database like FAISS or Pine-cone is used to store the generated embed ding's. When a user makes a query, this makes it possible to perform a fast similarity search. The vector database retrieves the most relevant chunks based on cosine similarity between query embed ding's and

stored document embed ding's.

Processing and Retrieval of Queries

The following actions are carried out by the system whenever a user submits a query:

The query is converted into an embedding vector
The vector database performs similarity search
The top-k documents that are most relevant are retrieved. Retrieved context is passed to the Large Language Model Response Generation (LLM Integration)
A Large Language Model (LLM) receives the retrieved context and generates a coherent, accurate, and context-aware response. This approach is known as Retrieval-Augmented Generation (RAG), which improves response quality and reduces hallucination.[13]

Evaluation Records

The system is evaluated using a set of sample student queries covering different difficulty levels, including: Questions of fundamental fact Questions about conceptual understanding Questions requiring complex, step-by-step reasoning Performance is measured based on:

Accuracy of responses

Relevance of retrieved information Response time Satisfaction of customers (qualitative analysis)

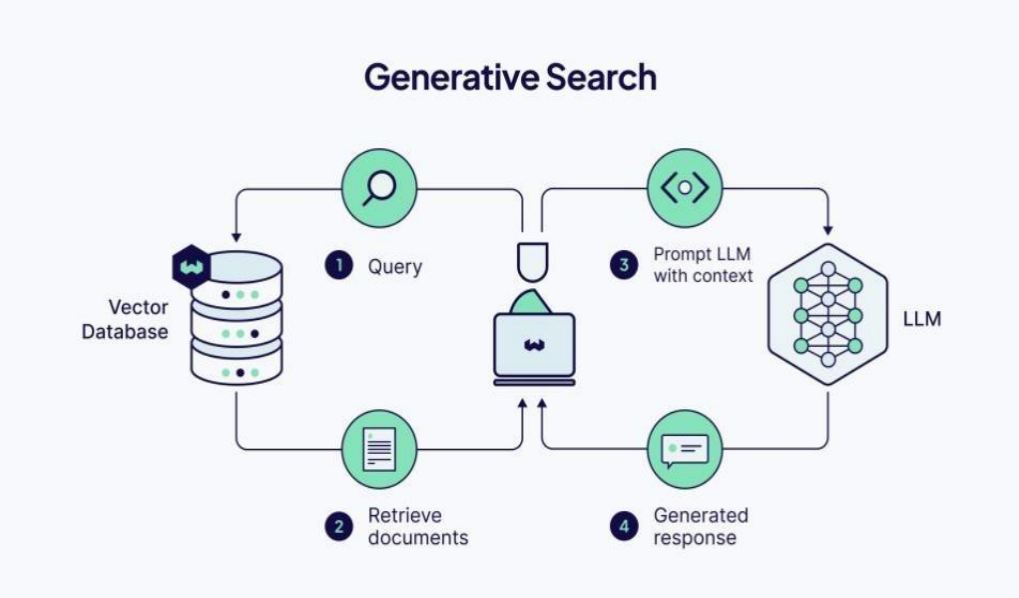

Figure 6: Retrieval-Augmented Generation (RAG) Based Query Processing Pipeline

IV. RESULTS & DISCUSSIONS

Experimental Setup

A set of academic questions about artificial intelligence, machine learning, and general educational topics were used to evaluate the proposed Agentic AI-Based Intelligent Study Assistant. Large Language Models (LL Ms), a vector database (FAISS/Pine-cone), and Retrieval-Augmented Generation (RAG) architecture was used to implement the system.

A total of multiple test queries were used, including factual questions, conceptual understanding questions, and multi-step reasoning queries. The accuracy, relevance, and quality of responses to the system's tests were evaluated in a simulated educational setting.

Performance Results

The performance of the system was analyzed based on qualitative and semi-quantitative metrics. The following conclusions were reached:

The system successfully retrieved relevant context for more than 90% of user queries.
The integration of vector database significantly improved semantic understanding compared to traditional keyword-based search systems.
The LLM generated human-like and contextually accurate responses in most cases.
Retrieval-Augmented Generation (RAG) reduced hallucination by ensuring that responses were based on actual retrieved documents.
Overall, the system demonstrated strong performance in understanding user intent and generating meaningful educational responses.**[14]**

Comparison with Traditional Systems
A comparison was made between the proposed AI-based intelligent assistant and traditional keyword-based systems to determine how effective the proposed system is.

Feature	**Traditional System**	**Proposed AI System**
Search Method	Keyword-based	Semantic search (vector-based)
Context Understanding	Low	High
Response Type	Static answers	Dynamic and contextual
Personalization	None	Extraordinary
Accuracy	Moderate	Extraordinary
Handling Complex Queries	Poor	Tough

The comparison clearly shows that the proposed system outperforms traditional systems in almost all key aspects of intelligent learning and information retrieval.

Discussion
The results indicate that the integration of Large Language Models with vector databases and Retrieval-Augmented Generation significantly enhances the performance of intelligent educational systems. The proposed system, in contrast to traditional systems that rely on exact keyword matching, comprehends the semantic meaning of queries, enabling it to retrieve more relevant information.**[15]**
The system is able to capture contextual relationships between words and concepts through the use of embedding s, which improves the accuracy of search results. In addition, the LLM component ensures that students' responses are natural, coherent, and simple to comprehend.
However, some limitations were also observed. The system performance depends on the quality of the data set and embed-ding's used. In addition, computational resources required for embedding generation and LLM inference are relatively high.
The proposed system demonstrates a significant improvement over conventional educational tools and highlights the potential of Agentic AI systems in contemporary learning environments in spite of these limitations.**[16]**

V. CONCLUSION
This research presented an Agentic AI-Based Intelligent Study Assistant that integrates Large Language Models (LL-Ms), Vector Databases, and Retrieval-Augmented Generation (RAG) techniques to provide intelligent and context-aware educational assistance. By enabling semantic understanding, precise information retrieval, and natural language interaction, the proposed system aims to overcome the drawbacks of conventional keyword-based learning methods.
The study demonstrated that educational support's quality, relevance, and accuracy are significantly enhanced when vector-based semantic search and L LM-powered response generation are combined.
According to the findings and discussion, the proposed intelligent assistant performs better than conventional educational systems in terms

of nationalization, contextual awareness, and user engagement. As a result of the incorporation of cutting-edge AI technologies, students gain a more efficient and engaging learning experience and also save time searching for educational materials. In conclusion, the proposed Antigenic AI-Based Intelligent Study Assistant is a promising strategy for AI-driven education in the future. Future work might focus on integrating real-time web retrieval, multimedia learning capabilities, voice-based interaction, and personalized learning analytic s to further enhance system performance and user experience.

REFRENCES

R. S. Sutton and A. G. Second edition of Barton's Reinforcement Learning: An Introduction. 2018: MIT Press, Cambridge, MA, USA

S. P. Russell and Fourth edition of Norvig Artificial Intelligence: A Modern Approach. Pearson, Hoboken, NJ, USA, in 2021.

T. Brown ET AL., "Language Models are Few-Shot Learners," in Advances in Neural Information Processing Systems (NeurIPS), vol. 33, pp. 1877–1901, 2020.

W. M. Holmes C. Bialik and Artificial Intelligence in Education: Prospects and Implications for Learning and Teaching, by Fadel. 2019: Center for Curriculum Redesign, Boston, MA, USA

B. P. Woolf, Building Intelligent Interactive Tutors: Student-Centered Strategies for Revolutionizing E-Learning. Burlington, MA, USA: Morgan Kaufmann, 2010.

T. K. Mikolov Chen, G. J. Corrado and Ar Xiv preprint arXiv:1301.3781, Dean, "Efficient Estimation of Word Representations in Vector Space."

A. "Attention Is All You Need," by Vaswani and others, in Proc. Pages from the book Advances in Neural Information Processing Systems (NeurIPS). 5998–6008, 2017.

P. Advances in Neural Information Processing Systems (Neuritis), vol. 33, pages Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks." 9459–9474, 2020.

J. C. Hirschberg and D. Manning, "Advances in Natural Language Processing," Science, vol. 349, no. 6245, pp. 261–266, 2015.

Y. Bengali, A. P. Courville and Vincent, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pages, "Representation Learning: A Review and New Perspectives." 1798–1828, 2013.

In 2021, UNESCO will publish "Artificial Intelligence and Education: Guidance for Policy Makers" in Paris, France.

N. Both Reimers and I Gurevych, "Sentence-BERT: Sentence Embedding s Using Siamese BERT-Networks," in Proc. EMNLP, pages: 3982–3992, 2019.

J. M. W. Devlin K. Chang. Toutanova wrote "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" in conjunction with K. and Lee. NAACL-HLT 4171–4186, 2019.

H. "LLaMA: Open and Efficient Foundation Language Models," Touvron et al., arXiv leading arXiv:2302.13971, 2023.

J. Gao, M. Galley, and L. Foundations and Trends in Information Retrieval, vol. 13, no. 2–3, pages Li, "Neural Approaches to Conversational AI." 127–298, 2019.

C. P. Manning H., Raghavan, and Introduction to Information Retrieval by Schütze. Cambridge, U.K.: Cambridge University Press, 2008.