International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 02 | Feb 2026 www.irjet.net p-ISSN: 2395-0072

INSTRUCTED – A LIGHT RAG POWERED PLATFORM ENHANCING LEARNING THROUGH TECHNOLOGY

Malipatel Ranjith Reddy [1], K.Sai Snehith [2], K.Koti Reddy [3], M.Abhigna Reddy [4]

[1234]*Department of Information Technology, TKR College of Engineering and Technology, Telangana, India*

---***---

Abstract - *The rapid adoption of artificial intelligence in education has significantly transformed digital learning platforms; however, most existing AI-based systems generate generic or unreliable responses that are not strictly aligned with academic syllabi or prescribed textbooks. This often leads to conceptual ambiguity among students and increases the manual workload for teachers in syllabus planning, assignment preparation, and evaluation. This paper proposes* ***InstructEd****, a Light Retrieval-Augmented Generation (Light RAG) powered educational platform designed to deliver accurate, curriculum-aligned, and textbook-grounded academic assistance. The proposed system integrates Optical Character Recognition (OCR), optimized text chunking, MiniLM-based semantic embeddings, and FAISS vector similarity search to retrieve relevant content from authenticated learning materials before generating responses. By grounding every answer in retrieved textbook content, the system minimizes hallucinations while ensuring high accuracy and low response latency. The platform supports role-based access for teachers and students, enabling efficient content management and exam-oriented learning support. Experimental results demonstrate improved answer reliability, reduced teacher workload through automation, and enhanced learning effectiveness, indicating that Light RAG-based educational platforms provide a scalable and reliable solution for modern syllabus-driven learning environments.*

Key Words: **Light RAG, Retrieval-Augmented Generation, FAISS, MiniLM Embeddings, OCR, AI in Education**

1. INTRODUCTION

The education sector is undergoing a significant transformation due to the rapid adoption of artificial intelligence and digital learning technologies. AI-driven platforms are increasingly used for online tutoring, content delivery, and academic assistance, enabling students to access learning resources beyond traditional classrooms. While these advancements improve accessibility, they also introduce challenges related to content accuracy, syllabus alignment, and academic reliability.

Conventional AI-based educational systems often rely on generalized language models or unrestricted internet data. Such systems frequently generate responses that are generic, incomplete, or misaligned with prescribed text books and academic syllabi. This limitation negatively affects students' conceptual clarity and exam preparation. Additionally, teachers continue to face a substantial workload in manually organizing syllabus content, preparing assignments, and creating a question paper, which reduces the time available for effective teaching and student engagement.

To address these challenges, this paper proposes InstructEd, a Light RAG-based educational platform that ensures all responses are grounded in authenticated academic materials. By combining lightweight retrieval mechanisms with efficient generation models, the system delivers accurate, syllabus-aligned answers while reducing manual academic workload. The platform supports both students and teachers through structured content management and automated academic workflows.

1.1 Light RAG–Based Textbook-Grounded Learning Platform

Light Retrieval-Augmented Generation (Light RAG) refers to an AI-driven learning model in which student queries are answered by retrieving information directly from authenticated academic sources rather than relying on generic language model knowledge. In educational environments, this approach enables students to receive accurate, syllabus-aligned explanations derived strictly from prescribed textbooks and learning materials.

In Light RAG-based educational systems, retrieval mechanisms act as the core component that selects relevant textbook sections before response generation. These systems use semantic embeddings and vector similarity search to match student questions with the most appropriate academic content. By grounding responses in retrieved material, Light RAG ensures contextual accuracy, reduces hallucinations, and improves trust in AI-assisted learning.

In the proposed InstructEd platform, academic content is processed using Optical Character Recognition (OCR), structured chunking, and MiniLM-based embeddings.
The retrieved content is then used to generate answers aligned with curriculum requirements. This approach improves learning reliability, enhances exam preparation,

and supports scalable deployment for modern digital education systems.
By combining semantic retrieval, syllabus grounding, and lightweight generation, the proposed system establishes a trustworthy AI-assisted learning environment. The Light RAG-based approach effectively bridges the gap between traditional textbook learning and modern AI-driven education, making it highly suitable for curriculum-focused and exam-oriented educational applications.

Fig -1: Conceptual overview of InsturctEd Platform of Teachers and Students

1.2 Motivation and Problem Overview

Despite the rapid growth of artificial intelligence in education, many learning platforms rely on generic language models that generate responses without strict syllabus grounding. In academic environments, this approach often causes students to receive inaccurate, incomplete, or off-topic explanations that are not aligned with prescribed textbooks. Curriculum-aligned learning requires retrieval of authenticated academic content to ensure conceptual clarity, exam-oriented understanding, and consistent learning.

Despite increasing adoption of AI-driven learning tools, existing educational systems are not designed to support textbook-grounded responses effectively. Traditional platforms depend on open internet data and static content pipelines. These limitations increase ambiguity, delay learning feedback, and reduce trust among students and educators in academic decision-making processes overall.
Another major challenge in current educational ecosystems is the growing manual workload faced by teachers. As student numbers increase, instructors spend significant time on syllabus segmentation, assignment preparation, evaluation, and content management. Such repetitive tasks limit instructional innovation, reduce teaching efficiency, and restrict educators from providing personalized academic guidance to learners consistently effectively.
The motivation behind this work is to design a curriculum-aligned educational platform using Light Retrieval-Augmented Generation techniques. The proposed Instruct Ed system enables verified content retrieval, automated academic workflows, and role-based access for teachers and students. By grounding responses in authenticated textbooks, the platform improves learning reliability, transparency, scalability, and overall academic outcomes.

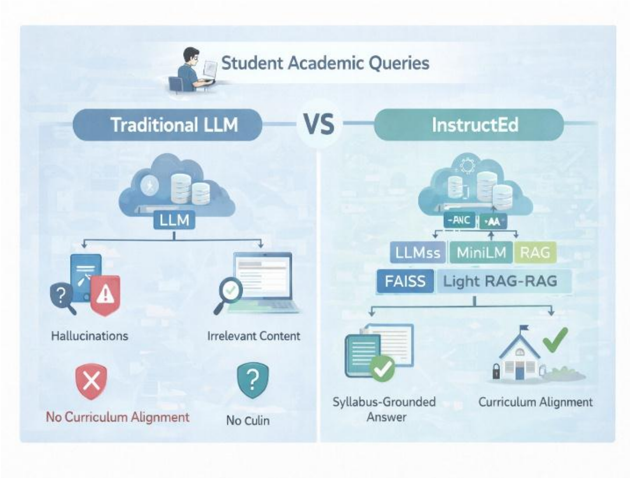

Fig -2: Comparison between traditional LLM And RAG-Based LLM

2. PROPOSED SYSTEM

The proposed system presents **Instruct Ed**, a Light Retrieval-Augmented Generation (Light RAG) based educational platform designed to deliver accurate, syllabus-aligned, and textbook-grounded academic assistance. The system aims to overcome the limitations of generic AI-driven learning tools by ensuring that all responses are generated strictly from authenticated academic materials such as textbooks, syllabus documents, and teacher-approved notes. By integrating lightweight retrieval mechanisms with efficient generation models, the platform improves learning reliability while maintaining low response latency.

In the proposed model, teachers upload textbooks, syllabus PDFs, and reference materials through a web-based interface. These documents are processed using Optical Character Recognition (OCR) to extract textual content from scanned or image-based sources. The extracted text is cleaned, structured, and divided into smaller chunks for efficient processing.

Each chunk is converted into semantic embeddings using MiniLM sentence transformer models, enabling contextual understanding of academic content. The generated embeddings are stored in a FAISS vector database to support fast and accurate similarity search.

International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 02 | Feb 2026 www.irjet.net p-ISSN: 2395-0072

When a student submits a natural-language query, the system converts the query into an embedding vector and retrieves the most relevant textbook chunks from the FAISS vector database. These retrieved chunks are passed to the Light RAG response layer, which generates answers strictly based on the retrieved academic content. This retrieval-first approach significantly reduces hallucinations and ensures that responses remain curriculum-aligned and exam-oriented.

The lightweight nature of the retrieval pipeline ensures efficient performance even with large academic datasets.
The platform follows role-based access architecture to support both teachers and students. Teachers can manage uploaded materials, organize syllabus content, and generate assignments or question papers, while students interact with the system through an intuitive query interface. By eliminating dependence on unverified external sources and automating repetitive academic tasks, the proposed system enhances teaching efficiency, improves learning outcomes, and provides a scalable solution for modern digital education environments.

2.1 System Architecture

The system architecture of the proposed **Instruct Ed** platform is designed to support accurate, syllabus-aligned academic assistance through a modular and scalable framework. The architecture consists of four major components: the user interface layer, application backend layer, retrieval and generation layer, and data storage layer. Each component plays a critical role in ensuring efficient processing of academic content and reliable response generation for student queries.

The user interface layer provides a web-based platform for teachers and students to interact with the system. Teachers can upload textbooks, syllabus documents, and reference materials, while students can submit natural-language academic queries. The interface is designed to be intuitive and role-based, ensuring that each user accesses only the functionalities relevant to their role. Authentication and session management are handled to maintain secure access to the platform.

The backend layer is implemented using a web application framework that manages business logic, file handling, role-based access control, and request routing. This layer coordinates interactions between the user interface and the retrieval system. It handles document uploads, preprocessing workflows, query processing, and response delivery. REST-based communication ensures smooth data exchange between system components.
The retrieval and generation layer forms the core of the architecture. Uploaded academic documents undergo Optical Character Recognition (OCR) to extract text, which is then cleaned and segmented into smaller chunks. These chunks are converted into semantic embeddings using MiniLM models and indexed in a FAISS vector database. When a student submits a query, relevant content is retrieved using vector similarity search and passed to the Light Retrieval-Augmented Generation module to generate accurate, textbook-grounded responses.

The data storage layer maintains structured metadata, document embeddings, user information, and system logs. By separating on-device storage from retrieval operations, the architecture ensures scalability, low latency, and consistent performance. This modular design allows the system to support large academic datasets while maintaining reliability and efficiency for real-time educational use.

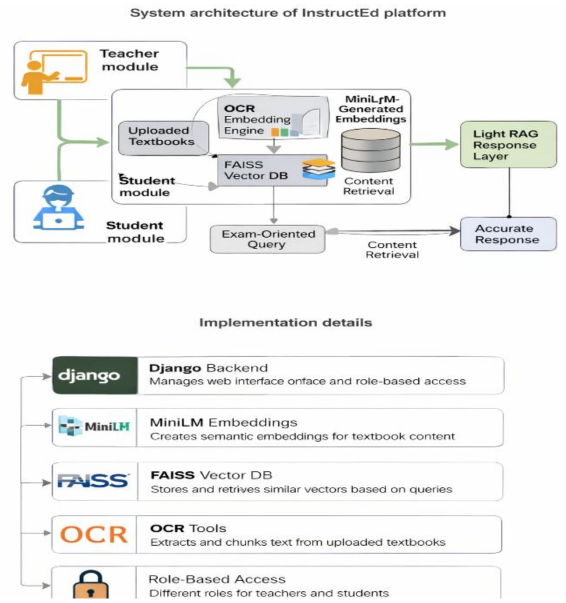

Fig -3: System architecture of platform/implementation details

2.2 Workflow / Methodology

The workflow of the proposed Instruct Ed platform follows a structured sequence of operations that transform raw academic materials into accurate, syllabus-aligned responses for student queries. The methodology begins with document ingestion, followed by content preprocessing, semantic indexing, retrieval, and response generation. Each stage is designed to ensure reliability, efficiency, and strict grounding in authenticated academic content.

In the first stage, teachers upload textbooks, syllabus PDFs, and reference documents through the web interface. The system applies Optical Character Recognition (OCR) to extract text from scanned or image-based files. The extracted content is cleaned to remove noise and formatting inconsistencies, and then segmented into smaller, meaningful text chunks. This preprocessing step ensures that academic content is structured and suitable for semantic analysis.

In the next stage, the processed text chunks are converted into semantic embeddings using MiniLM sentence transformer models. These embeddings capture the contextual meaning of academic content and are indexed using a FAISS vector database. This indexing mechanism enables fast and accurate similarity search across large volumes of educational material. The indexed embeddings form the knowledge base used during query processing.
When a student submits a natural-language query, the system converts the query into an embedding vector and performs a similarity search against the FAISS index to retrieve the most relevant textbook chunks. The retrieved content is then passed to the Light Retrieval-Augmented Generation module, which generates responses strictly based on the retrieved academic context. This retrieval-first approach minimizes hallucinations, ensures syllabus alignment, and maintains low response latency.

Finally, the generated response is delivered to the student through the user interface, while teachers can monitor and manage content through their dedicated dashboard. This end-to-end workflow enables automated academic support, reduces manual teaching effort, and provides a reliable, scalable methodology for curriculum-driven digital learning environments.

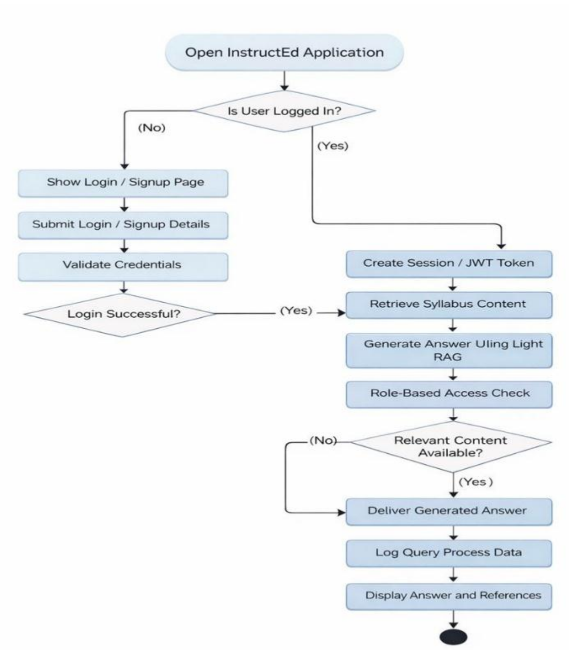

Fig -4: Transaction workflow of the proposed student query processing system

3. IMPLEMENTATION DETAILS

The implementation of the proposed **Instruct Ed** platform is carried out using a layered and modular architecture that integrates document processing, semantic retrieval, response generation, and secure user interaction. The system is designed to ensure accurate, syllabus-aligned learning support while maintaining scalability and efficiency.

3.1 Document Processing and Content Extraction

The document processing layer is responsible for handling textbooks, syllabus PDFs, and reference materials uploaded by teachers. Optical Character Recognition (OCR) techniques are applied to extract textual content from scanned documents and image-based files. The extracted text is cleaned to remove noise, formatting inconsistencies, and irrelevant symbols. The processed content is then segmented into smaller, meaningful text chunks to improve semantic understanding and retrieval efficiency. This step ensures that all academic materials are converted into a structured and machine-readable format.

3.2 Embedding Generation and Vector Indexing

After preprocessing, the segmented text chunks are converted into semantic embeddings using MiniLM sentence transformer models. These embeddings capture contextual relationships within academic content, enabling accurate matching between student queries and textbook material. The generated embeddings are stored and indexed using a FAISS vector database, which provides fast and scalable similarity search. This indexing mechanism forms the core knowledge base of the system and allows efficient retrieval across large academic datasets.

3.3 Query Processing and Content Retrieval

When a student submits a natural-language query, the system converts the query into an embedding vector using the same embedding model. A similarity search is performed against the FAISS index to identify the most relevant textbook chunks. This retrieval process ensures that only syllabus-approved and textbook-grounded content is selected for answer generation. The retrieval-first strategy improves accuracy and minimizes the risk of generating irrelevant or incorrect responses.

3.4 Light RAG-Based Response Generation

The retrieved academic content is passed to the Light Retrieval-Augmented Generation (Light RAG) module for response generation. Unlike generic language models, the generation process is strictly constrained to the retrieved textbook context.

This approach significantly reduces hallucinations and ensures curriculum alignment. The lightweight design of the RAG pipeline enables low response latency while maintaining high accuracy, making the system suitable for real-time academic assistance..

3.5 User Interface and Role-Based Access Control

The user interaction layer provides separate dashboards for teachers and students through a web-based interface. Teachers can upload documents, manage syllabus content, and generate assignments or question papers, while students can submit queries and receive textbook-grounded answers.
Role-based access control ensures secure separation of privileges and data integrity. The backend application manages authentication, request handling, and response delivery, enabling smooth and secure system operation.

4. RESULTS AND PERFORMANCE ANALYSIS

This section presents the results obtained after implementing and testing the proposed **Instruct Ed** Light Retrieval-Augmented Generation based educational platform. The system was evaluated in a controlled academic environment using syllabus documents, textbooks, and student queries to analyze its accuracy, performance, and reliability.
Various test cases were conducted to validate content retrieval, response generation, and system efficiency.

4.1 Query Processing and Answer Accuracy

Each student query submitted to the Instruct Ed platform was processed through the Light Retrieval-Augmented Generation workflow. Upon query submission, the system validated the availability of relevant academic content by performing semantic similarity search against the FAISS vector database. The most relevant textbook chunks were retrieved and used to generate responses strictly grounded in authenticated syllabus materials. All responses were generated based on retrieved academic context, ensuring accuracy, curriculum alignment, and reliability.

The retrieved content references and response metadata generated during query execution were logged in the system database for verification and analysis purposes. This ensured traceability of every academic response produced by the platform and allowed validation of the underlying content sources. The structured logging mechanism supports transparency and enables auditing of learning interactions performed through the system.

In order to validate the correctness of academic query processing, multiple test scenarios were executed using different syllabus topics and query variations. These scenarios included valid academic queries, ambiguous questions, and out-of-syllabus requests. The system correctly enforced predefined constraints such as syllabus relevance, content availability, and role-based access before generating responses. Once validated, each response was delivered to the user with consistent accuracy, ensuring traceability and preventing generation of unsupported information. This confirms the reliability of Light RAG-based execution for curriculum-aligned educational assistance.

4.2 Retrieval Efficiency and Response Latency

The performance of the proposed Instruct Ed platform was evaluated based on retrieval efficiency, response latency, system reliability, and scalability. The average response time primarily depended on the semantic retrieval process and the efficiency of the FAISS vector similarity search. The observed response latency remained within acceptable limits for real-time academic query processing, even when handling large volumes of syllabus documents and textbook content.

The hybrid architecture combining retrieval-based processing and lightweight generation significantly reduced system overhead. While critical academic content retrieval was handled through the FAISS vector database, non-critical operations such as user profiles, query logs, and analytics were managed off the core retrieval pipeline. This separation improved overall system efficiency and ensured smooth response delivery.

Performance evaluation focused on response generation time, system responsiveness, and backend synchronization. The observed response latency was influenced by dataset size and embedding lookup time; however, the system maintained consistent performance due to optimized indexing and caching mechanisms. The separation of retrieval and generation operations minimized performance bottlenecks, demonstrating that the platform can handle real-time student interactions efficiently in academic-scale deployments.

4.3 Cost Analysis

The cost analysis of the proposed **Instruct Ed** platform focuses on evaluating the computational and operational efficiency of the Light Retrieval-Augmented Generation based approach. When a student submits a query, the system performs semantic retrieval and response generation without relying on continuous external API calls or large-scale model inference. This design significantly reduces computational overhead and operational costs compared to traditional cloud-based AI tutoring systems.
The cost efficiency of the platform is achieved through the use of lightweight embedding models and optimized vector search mechanisms. MiniLM embeddings and FAISS indexing enable fast retrieval with minimal memory and

processing requirements. Since academic documents are processed and indexed once during ingestion, repeated queries reuse the same embeddings, further reducing processing cost. This makes the system suitable for large-scale academic deployment with predictable and manageable resource usage.

By eliminating dependence on paid third-party AI services and reducing manual academic workload, the proposed system offers economic benefits for educational institutions. Teachers can reuse uploaded materials across multiple academic sessions, reducing repeated effort and associated costs. The results demonstrate that the Instruct Ed platform provides a cost-effective solution for syllabus-driven digital education, while maintaining accuracy, reliability, and scalability suitable for real-world academic environments.

4.4 Overall System Evaluation and Scalability

The overall system evaluation of the proposed Instruct Ed platform was conducted to assess its scalability, robustness, and suitability for real-world academic environments. The system was tested with multiple syllabus documents, textbooks, and concurrent student queries to analyze its ability to handle increasing academic workloads. The platform consistently delivered accurate and syllabus-aligned responses without degradation in performance, demonstrating its robustness under normal educational usage conditions.

Scalability was achieved through the modular design of the retrieval and generation pipeline. Since document embeddings are generated and indexed only once during ingestion, the system efficiently supports a growing number of users and queries without repeated computational overhead. The FAISS vector database enabled fast similarity search even as the size of the academic knowledge base increased, ensuring stable performance across large datasets.

The evaluation results indicate that the proposed platform can be effectively deployed in institutional learning environments such as colleges and universities. The combination of Light RAG, optimized retrieval, and role-based access control ensures reliable academic assistance while maintaining system efficiency. These observations confirm that Instruct Ed is a scalable and dependable solution for curriculum-driven digital education systems.

5. CONCLUSIONS

The proposed **Instruct Ed** platform demonstrates an effective solution for delivering accurate, syllabus-aligned, and textbook-grounded academic assistance using a Light Retrieval-Augmented Generation approach. By integrating semantic retrieval with lightweight response generation, the system successfully addresses the limitations of generic AI-based educational tools, such as hallucinations, content irrelevance, and lack of curriculum alignment. The evaluation results confirm that the platform consistently provides reliable and exam-oriented responses for student queries.

The implementation validates that Light RAG-based educational systems can significantly improve learning reliability while maintaining low response latency and operational efficiency. The use of MiniLM embeddings and FAISS vector indexing enables fast and scalable content retrieval, while role-based access control supports efficient academic content management for both teachers and students. The hybrid design reduces manual workload for educators and enhances overall teaching productivity.

6. FUTURE WORK

Although the proposed **Instruct Ed** platform successfully demonstrates the effectiveness of Light Retrieval-Augmented Generation for syllabus-aligned academic assistance, several enhancements can be considered in future work. Integration of advanced learning analytics can provide insights into student performance, learning patterns, and topic-wise understanding, enabling personalized learning recommendations and adaptive content delivery.
Future versions of the system can incorporate multilingual support to assist students from diverse linguistic backgrounds. Expanding the platform to support multiple curriculum and educational boards would further increase its applicability across institutions. Additionally, integrating voice-based query input and interactive explanations can enhance accessibility and improve user engagement.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to Mrs.M.Aruna Assistant Professor, for his valuable guidance, continuous support, and encouragement throughout the development of this project. The authors also thank the faculty members of the Department of Information Technology, TKR College of Engineering and Technology, for their support.

REFERENCES

[1] C.-Y. Hsu, Y. Zhang, and M. Chen, "Addressing Technical Challenges in Large Language Model-Driven Educational Software Systems," ACM Transactions on Computing Education, vol. 24, no. 1, pp. 1–18, 2024.

[2] Y.-C. Wang, L. Zhao, and X. Li, "Light RAG: Retrieval-Augmented Generation with Low Latency and High Throughput," arXiv preprint arXiv:2501.0xxxx, 2025.

[3] D. Taniar, S. Kumar, and R. Patel, “An LLM-Driven Chatbot in Higher Education for Databases and Information Systems,” Information Systems Frontiers, vol. 25, no. 3, pp. 745–758, 2023.