Generative AI Based Teaching Assistant

G. Santhoshi[1], Mamilla Rithika[2], Manda Anjani[3], Berelli Preetham Rao[4]
[1,2,3,4]*Department Of Computer Science and Engineering, Geethanjali College Of Engineering and Technology, Hyderabad, India*

***Abstract*— Our project proposes a method for building GenAI-Based Teaching Assistant, an AI-powered educational tool designed to assist teachers by generating content and Bloom's Taxonomy-based questions from either text input or uploaded documents. The system is built using OpenAI's GPT models and integrated with LangChain for orchestrating prompt management and document handling workflows. To ensure efficient semantic retrieval, we utilized Sentence Transformers for embedding the text and stored the vectorized content in ChromaDB. Document inputs, including PDFs, DOCX files, are processed using custom-built tools to clean, chunk, and summarize data before question generation. A Flask-based backend facilitates communication between the model and the user, while the web interface is designed using HTML, CSS, and JavaScript for intuitive usability. This assistant aims to reduce teachers' workload by automatically generating level-based questions, making educational content creation faster, structured, and more adaptive.**

***Index Terms*— Bloom's Taxonomy, ChromaDB, GPT-4.0, Generative AI, LangChain, Large Language Models (LLMs), Natural Language Processing, OpenAI API, Question Generation, Sentence Transformers.**

I. INTRODUCTION

As artificial intelligence (AI) continues to advance, its role in transforming education becomes increasingly evident. In particular, breakthroughs in natural language processing (NLP) and generative AI are opening up new possibilities for making teaching and learning more efficient, personalized, and engaging. This paper presents the GenAI-Based Teaching Assistant, a smart system designed to help educators by automatically generating educational content from topics and creating Bloom's Taxonomy-based questions from uploaded documents.

At the heart of the system is OpenAI's GPT-4.0, a powerful language model that works alongside the LangChain framework to handle everything from prompt generation to document processing. When users input a topic, the assistant generates clear, structured content related to that subject. For document-based use, educators can upload files such as PDFs or Word documents which are then processed through a set of internal tools. These tools handle tasks like parsing the content, summarizing it, breaking it into chunks, and converting it into vector form using Sentence Transformers. The resulting vectors are stored in ChromaDB, allowing the system to retrieve the most relevant pieces of information for question generation.

The backend is developed using Flask-API, ensuring seamless communication between the AI model and the user interface. The frontend, built with HTML, CSS, and JavaScript, is designed to be intuitive and easy to use, giving educators a smooth experience from input to output. By clearly separating the content and question generation workflows, the assistant ensures both processes are handled accurately and efficiently. This work highlights how generative AI can be applied thoughtfully in the classroom helping educators save time, improve content quality, and support a more adaptive, student-centered learning experience.

II. BACKGROUND

The recent speedup in the growth of Artificial Intelligence (AI) in education is mainly attributed to the development of Natural Language Processing (NLP) and Generative AI. As several AI technologies have come up, the latest Large Language Models (LLMs) such as OpenAI's GPT Models, these advancements have also shown great capabilities in understanding and generating what seems to be human speech. Among many other applications in education, such models are getting increasingly used for the automation and enhancement of quite a few activities, such as content creation and personalized learning.

Traditionally, the process of content creation and question-making for education has been manual and

time-consuming for educators. With the ever-increasing complexity and diversity of learning materials, creating assessments that are truly relevant and meaningful as well as cognitively appropriate proves to be a real challenge. One of the major pedagogic foundations upon which educational content is built is Bloom's Taxonomy, which breaks learning objectives down into levels of cognitive hierarchy. This helps educators orient their contents and assessments toward deep learning and critical thinking.

Despite its widespread use, the manual application of Bloom's Taxonomy to generate varied and high-quality questions is often labor-intensive. Furthermore, existing educational AI tools do not possess sufficient capabilities to modify their outputs through such a lens to create truly meaningful learning experiences. Thus, there is a growing demand for AI-driven systems with the capacity to generate educational content and design assessments that meet the cognitive learning objective. The other side of the coin involves increasing digitization of educational materials, such as PDFs, DOCX files, or other digital resources, that together lead to a large reservoir of textual data for feeding back into the learning-teaching process. However, a tool that can understand, digest, and quickly extract meaningful knowledge from those documents might be a more complex challenge.

The GenAI-Based Teaching Assistant is seeking to address these problems by using generative AI to automate both content generation and question creation. It utilizes GPT-4.0 for creating content based on topics inputted by the user and uses NLP tools to automatically process documents and generate questions based on Bloom's Taxonomy. Educators can upload files to the system, which are then parsed, summarized, and converted into structured content from which relevant high-quality questions are generated.

This research highlights how generative AI can enhance educational practices by alleviating manual workloads, improving the quality of content and assessments, and fostering an adaptive learning environment for both educators and students. The GenAI-Based Teaching Assistant represents a significant advancement in the application of AI in education, offering a practical approach to bridging the gap between traditional educational models and modern AI-driven methodologies.

III. LITERATURE SURVEY

The integration of AI into education has been an important area of research that has given great promise for developing content generation, automated assessment, and personalized learning. The early works in educational AI relied on rule-based expert systems, which were limited in terms of adaptability and comprehension of context. However, the advent of large language models like those of GPT-3 and GPT-4 has changed the patterns in which intelligent systems are used in education.

Tom B. Brown et al. [1] demonstrated the ability of GPT-3 to generate human-like text, sparking interest in their educational potential. Subsequent research by OpenAI and others has validated the effectiveness of these models in generating coherent, context-aware educational content. GPT-4.0, the most recent iteration, offers even greater precision and contextual understanding, making it suitable for adaptive learning environments.

Zishan Ahmed et al. [2] explored the potentiality of Generative AI tools in enhancing higher education through platforms like ChatGPT. Their study found that such tools can serve as supplementary teaching aids, particularly in technical subjects such as programming, improving learner engagement and comprehension. However, the authors also caution against over-reliance due to possible inaccuracies and ethical concerns related to bias and misinformation.

Uday Mittal et al. [3] presented a comprehensive review of Generative AI applications in education. The paper categorized use cases into content generation, adaptive learning, and automated assessment. It emphasized the transformative capabilities of GenAI while also highlighting critical issues such as data privacy, the need for ethical governance, and the challenge of maintaining human oversight in AI-assisted learning environments.

Ghader Kurdi et al. [4] conducted a systematic review of Automatic Question Generation (AQG) methods for educational purposes. They examined rule-based, template-based, and neural-network-driven techniques, discussing their respective strengths and limitations. The study highlighted the importance of aligning

generated questions with cognitive frameworks like Bloom's Taxonomy to enhance pedagogical effectiveness.

Subhankar Maity and Aniket Deroy [5] investigated the impact of large language models on the future of learning. The research examined GenAI's role in automating assessments and feedback mechanisms, promoting self-directed learning, and enabling real-time tutoring. While recognizing the pedagogical benefits, the study warned against ethical risks such as content hallucination and dependency.

Kevin Hwang et al. [6] proposed an AI-assisted system for generating multiple-choice questions at scale, aligned with Bloom's Taxonomy. Utilizing GPT-3.5 and zero-shot prompting techniques, the system was evaluated using automated scoring models such as RoBERTa and IWF detection. Results demonstrated the model's ability to generate high-quality and cognitively diverse questions, establishing a foundation for scalable assessment tools.

Matthew Nyaaba et al. [7] studied the adoption of Generative AI by pre-service teachers (PSTs), who used these tools for quiz creation, lesson planning, and personalized learning. Their findings revealed a general optimism toward AI's support role in teaching, though concerns regarding reliability, creativity erosion, and ethical use persisted. The study called for targeted AI literacy training in teacher education programs.

Ethan Dickey and Andres Bejarano [8] proposed a conceptual model for integrating GenAI into curriculum design through the GenAI Content Generation Framework. Their model aligns with backward design principles and supports learning objectives using prompt engineering and taxonomy-based content alignment. The study emphasized collaborative AI-human workflows to maintain pedagogical integrity.

Collectively, the project is grounded in these studies, emphasizing the importance of ethical design, domain-specific adaptation, and deeper insights into AI's interpretative capabilities in education. Building on the insights from these investigations, the project explores the potential of AI-assisted teaching and learning.

IV. IMPLEMENTATION

A. System Architecture Overview

The system is built with a clean, web-based interface using Flask, making it easy for users to either enter a topic for content generation or upload documents for question generation. When a user submits a topic, the backend uses GPT-powered models via LangChain to create rich, educational content. For uploaded files, the system summarizes the content first and then generates Bloom's Taxonomy-based questions especially targeting higher-order thinking skills. ChromaDB helps manage embeddings efficiently during this process. The modular design ensures that each component works together smoothly, allowing for easy future upgrades and integration with more AI tools.

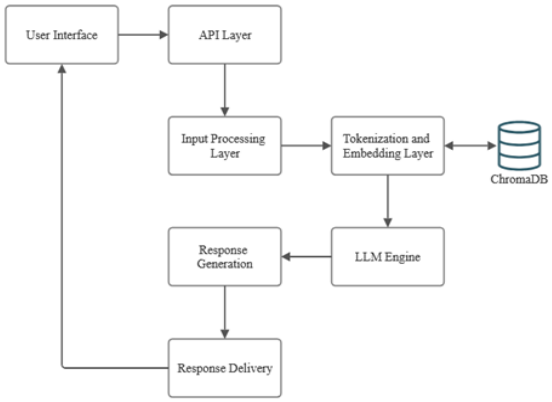

Fig.1 System Design

B. Model Selection

After extensive research and experimentation with various language models, GPT-4 was selected as the backbone of the system. Its ability to understand context, summarize text accurately, and generate high-quality responses made it well-suited for educational tasks. Integrated with LangChain, GPT-4 enabled smooth orchestration of both content and question generation workflows.

C. Documents Preprocessing

The uploaded documents in formats such as PDF or DOCX are parsed and chunked using LangChain tools. Each chunk is transformed into semantic embeddings using Sentence Transformers, enabling efficient similarity search and retrieval. These embeddings are stored in ChromaDB, which serves as a fast, persistent vector database for subsequent retrieval tasks.

D. Summarization Layer

A summarization step is embedded into the question generation pipeline to ensure that the generated questions are focused, relevant, and aligned with the core content. This module uses GPT-4 via LangChain to generate coherent summaries, which serve as the base input for Bloom's-aligned question prompts.

E. Content Generation

Users can input any academic topic through the frontend. This input is routed to the backend where LangChain formats it into a structured prompt for GPT-4. The model returns rich, contextual content related to the topic, which is then displayed on the interface with an option for the user to download the content in text format for further use.

F. Questions Generation

Once a document is uploaded and summarized by GPT-4, key concepts are extracted and passed to LangChain. LangChain then orchestrates Bloom-specific prompts to generate categorized questions, ranging from lower-order cognitive skills to higher-order cognitive skills. The final output is displayed on the interface, along with an option to download the generated question set.

G. User Interface and Integration

The system presents a clean and responsive web interface developed using HTML, CSS, and JavaScript, ensuring smooth navigation for users of all backgrounds. A simple registration and login mechanism is implemented using MongoDB, providing secure access to the application's dashboard. On the backend, the application is powered by Flask, which acts as the bridge between the frontend and the LLM. Flask handles all user requests, manages file uploads, and routes them to the appropriate modules for processing. For content generation, the user's topic is sent to GPT-4 via LangChain, which returns a well-formatted summary. For question generation, uploaded documents are parsed and chunked using LangChain, converted into embeddings using Sentence Transformers, and stored in ChromaDB for efficient retrieval. Summarization and question generation are then carried out using GPT-4 based on Bloom's-aligned prompts.

Together, the frontend and backend are tightly integrated to ensure a seamless experience. The responsive design, combined with efficient model orchestration, makes the platform accessible, scalable, and ideal for educational use.

V. RESULTS AND DISCUSSION

The findings revolve around the assessment of the user interface, functionality, and performance of the AI educational assistant. Evaluation metrics included the accuracy and relevance of the generated content and Bloom's Taxonomy-aligned questions, system responsiveness, and overall user experience. Both the Content Generation and Question Generation modules were ensured to maintain consistency and educational value.

The integration of LangChain, GPT-4, Sentence Transformers, and ChromaDB seamlessly supported document handling while enabling real-time semantic understanding. The Flask-based frontend offered a smooth and intuitive user experience, allowing users to upload documents or enter topics with ease.

Performance tests demonstrated that the system delivered high-quality outputs with low latency in practical use cases. User feedback highlighted the platform's user-friendly design, its effectiveness in the learning process, and the cognitive depth of the generated questions, thereby affirming its value in AI-enhanced education.

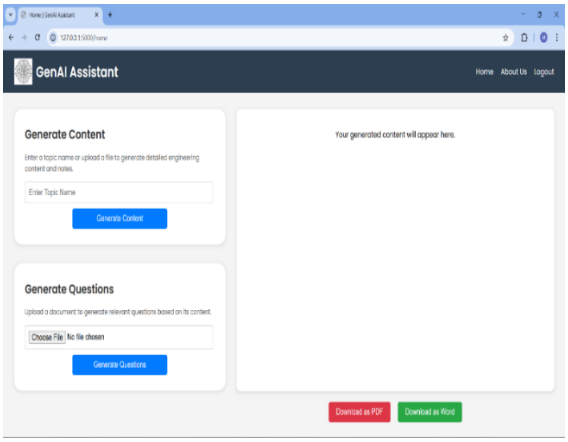

Fig.2 Home Page

The application's home page features a clean and user-friendly interface, guiding users through its two core functionalities: Content Generation and Question

Generation. It incorporates a simple registration and login system that manages access to the application. The layout is designed for easy navigation, making it accessible even to educators with limited technical background.

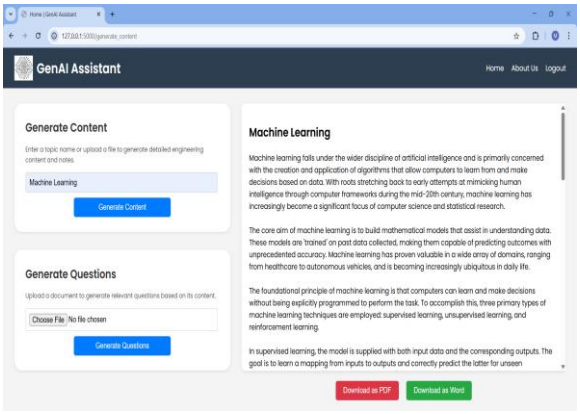

Fig.3 Content Generation

After logging in, users are directed to the dashboard where they can access the application's features. In the Content Generation section, users can enter any topic of their choice, and the system generates coherent and informative educational content based on that input. This feature is designed to assist learners and educators in quickly obtaining relevant content on demand. Users also have the option to download the generated material for offline use, enhancing accessibility and convenience.

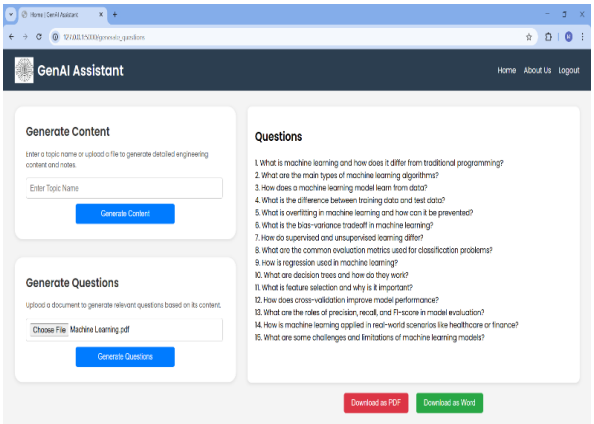

Fig.4 Questions Generation

The Question Generation feature allows users to upload documents in formats such as PDF, DOCX, or TXT. The system automatically processes the file, performs summarization, and generates questions aligned with various levels of Bloom's Taxonomy, supporting a range of cognitive skills from basic recall to critical thinking. These questions are also made available for download, enabling educators to easily incorporate them into assessments or study guides. The intuitive layout ensures a smooth user experience throughout the process.

VI. CONCLUSION

This project marks a significant achievement in the development of an AI-powered educational assistant capable of generating Bloom's Taxonomy-aligned questions and content from user-uploaded documents or defined topics. By integrating state-of-the-art technologies such as LangChain, GPT-4, Sentence Transformers, and ChromaDB, the system effectively demonstrates how generative AI can support personalized learning and automated assessment in educational settings. The architecture emphasizes modularity and efficiency, while the use of Flask ensures a user-friendly interface for seamless interaction and accessibility.

The question generation pipeline, combined with document summarization and semantic embedding, provides a robust mechanism for extracting pedagogically meaningful content from educational materials. This enables users to effortlessly generate questions across cognitive levels, thereby enhancing comprehension and critical thinking.

This work serves as compelling evidence of the transformative potential of generative AI in education, offering a novel tool for both educators and learners. Future enhancements may include support for multilingual content, integration of speech-based interactions, and dynamic alignment with diverse curriculum standards, further advancing the development of AI-driven educational tools.

ACKNOWLEDGMENT

The authors wish to express their heartfelt appreciation to Geethanjali College of Engineering and Technology for offering the essential infrastructure and resources that facilitated the successful execution of this research. Our sincere gratitude extends to the mentors and department staff whose expert guidance, unwavering support, and insightful feedback were pivotal throughout the course of the project.

Additionally, the authors acknowledge the valuable open-source tools, libraries, and research literature that significantly contributed to the design, implementation, and evaluation stages of the project. Lastly, the authors extend their thanks to all individuals whose contributions, both direct and indirect, have been instrumental in the completion of this work.

REFERENCE

[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, Dario Amodei. (2020). Language Models are Few-Shot Learners. arXiv preprint arXiv:2005.14165.

[2] Zishan Ahmed, Shakib Sadat Shanto and Akinul Islam Jony (2024). Potentiality of generative AI tools in higher education: Evaluating ChatGPT's viability as a teaching assistant for introductory programming courses. STEM Education, 4(3), 165–182.

[3] Uday Mittal, Siva Sai, Vinay Chamola and Devika Sangwan (2024). A Comprehensive Review on Generative AI for Education. IEEE Access.

[4] Ghader Kurdi, Jared Leo, Bijan Parsia, Uli Sattler and Salam Al-Emari (2020). A Systematic Review of Automatic Question Generation for Educational Purposes. International Journal of Artificial Intelligence in Education, 30:121–204.

[5] Subhankar Maity & Aniket Deroy (2024). The Future of Learning in the Age of Generative AI: Automated Question Generation and Assessment with Large Language Models. arXiv preprint arXiv:2410.09576.

[6] Kevin Hwang, Sai Challagundla, Maryam M. Alomair, Lujie Karen Chen, & Fow-Sen Choa (2023). Towards AI-Assisted Multiple Choice Question Generation and Quality Evaluation at Scale: Aligning with Bloom's Taxonomy. In Workshop on Generative AI for Education (GAIED).

[7] Matthew Nyaaba, Lehong Shi, Macharious Nabang, Xiaoming Zhai, Patrick Kyeremeh, Samuel Arthur Ayoberd, and Bismark Nyaaba Akanzire (2024). Generative AI as a Learning Buddy and Teaching Assistant: Pre-service Teachers' Uses and Attitudes. ResearchGate. arXiv preprint arXiv:2407.11983.

[8] Ethan Dickey, & Andres Bejarano (2023). A Model for Integrating Generative AI into Course Content Development. arXiv preprint arXiv:2308.12276.

[9] Changyoon Lee, Junho Myung, Jieun Han, Jiho Jin, and Alice Oh (2023). Learning from Teaching Assistants to Program with Subgoals: Exploring the Potential for AI Teaching Assistants. arXiv preprint arXiv:2309.10419.