WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal Homepage: www.ijrpr.com ISSN 2582-7421

Adaptive AI Tutors: Enhancing Student Support through Dynamic Intent Parsing with Open-Source Models

Bharti Garje[1], Chaudhari V. S[2]*

[1] Student, Vishwbharti Academy College of Engineering, Ahilyanagar 414201, India
[2] Associate Professor, Vishwbharti Academy College of Engineering, Ahilyanagar 414201, India

A B S T R A C T

The rapid advancement of Artificial Intelligence (AI) and Large Language Models (LLMs) has created new opportunities for improving the quality of education through personalized and interactive learning experiences. However, traditional learning environments often face challenges in providing immediate assistance, individualized guidance, and continuous academic support to students. To address these challenges, this paper proposes AutoTA, a dynamic intent-based virtual teaching assistant that utilizes open-source Large Language Models to provide intelligent and context-aware academic assistance. The proposed system is designed to understand student queries, identify their underlying intent, and generate relevant responses based on the specific learning requirements. It categorizes student queries into different academic intents, such as conceptual explanations, homework assistance, syllabus-related questions, and lecture discussions. Based on the identified intent, the system dynamically selects an appropriate response generation strategy. Furthermore, the integration of Retrieval-Augmented Generation (RAG) enables the assistant to retrieve relevant information from uploaded academic materials, including lecture notes, textbooks, and course documents, thereby improving the relevance and reliability of generated responses. The use of open-source LLMs provides flexibility, scalability, and reduced dependency on proprietary AI services. AutoTA aims to enhance student engagement, support self-paced learning, and provide accessible academic assistance beyond traditional classroom hours. The proposed framework demonstrates the potential of combining intent classification, document retrieval, and open-source language models to develop an intelligent and adaptive educational support system.

Keywords: Artificial Intelligence, Large Language Models, Virtual Teaching Assistant, Intent Classification, Open-Source LLMs, Retrieval-Augmented Generation, Personalized Learning, Educational Technology.

Introduction

The rapid development of Artificial Intelligence (AI) has significantly transformed the education sector by introducing innovative approaches to teaching, learning, and academic support. Traditional educational systems primarily depend on classroom-based instruction, where students receive guidance from teachers during scheduled sessions. However, students often encounter difficulties in understanding complex concepts, obtaining immediate answers to academic queries, and accessing personalized learning resources outside classroom hours. The increasing demand for flexible, accessible, and personalized education has created the need for intelligent systems that can provide continuous academic assistance. In this context, AI-powered virtual teaching assistants have emerged as a promising solution for enhancing student learning experiences.

Recent advancements in Natural Language Processing (NLP) and Large Language Models (LLMs) have enabled machines to understand and generate human-like responses to natural language queries. LLMs can explain complex concepts, summarize educational materials, answer questions, and support interactive learning. Unlike traditional rule-based chatbots, which rely on predefined responses and fixed conversational patterns, LLM-based systems can interpret diverse user queries and generate contextually relevant responses. These capabilities make LLMs suitable for developing intelligent educational assistants that can support students across different subjects and academic levels.

Despite these advancements, existing educational chatbots often face several limitations. Many systems provide generic responses without considering the specific intent of the student's query or the context of the learning material. In addition, some AI-based educational platforms depend on proprietary language models, which may introduce operational costs, limited customization, and concerns regarding data privacy. Furthermore, generating responses without referring to verified academic resources can result in inaccurate or irrelevant information. Therefore, there is a need for an intelligent virtual teaching assistant that can understand student intent, utilize relevant academic resources, and provide accurate and personalized educational support.

To address these challenges, this paper proposes AutoTA, a Dynamic Intent-Based Virtual Teaching Assistant for Students Using Open Source LLMs. The proposed system integrates intent classification, natural language understanding, and open-source Large Language Models to provide context-aware academic assistance. When a student submits a query, the system first analyzes the input to identify the underlying intent, such as conceptual clarification, subject-related questions, assignment assistance, or document-based queries. Based on the identified intent, the system dynamically selects an appropriate response generation approach. This intent-based mechanism helps improve the relevance and effectiveness of the responses provided to students.

The proposed AutoTA framework can further incorporate Retrieval-Augmented Generation (RAG) to retrieve relevant information from academic resources, including lecture notes, textbooks, syllabus documents, and study materials. The retrieved information is provided as contextual input to the language model, enabling it to generate responses that are more closely aligned with the student's learning requirements. The use of open-source LLMs offers greater flexibility in model selection, customization, and deployment while reducing dependency on external proprietary AI services. The system is designed to support interactive learning by allowing students to ask questions in natural language and receive explanations according to their academic needs.

The primary objective of AutoTA is to develop an intelligent and adaptive educational assistant that improves the accessibility and effectiveness of student learning. By combining intent-based query processing with open-source language models and academic knowledge retrieval, the proposed system aims to provide personalized assistance, encourage self-paced learning, and reduce the dependency on continuous human intervention for routine academic queries. The system can serve as a supportive educational tool for students and educators by making learning resources more accessible and interactive.

The remainder of this paper is organized as follows. Section II presents the literature survey of existing AI-based educational assistance systems and LLM-driven learning applications. Section III describes the proposed AutoTA system and its major components. The subsequent sections discuss the implementation methodology, experimental evaluation, results, and conclusions with future scope.

Literature Survey

The development of Artificial Intelligence (AI) and Natural Language Processing (NLP) has introduced new possibilities for intelligent educational systems. Researchers have explored various approaches, including intent-based chatbots, transformer-based models, Large Language Models (LLMs), and Retrieval-Augmented Generation (RAG), to improve the quality of academic assistance. The following literature survey reviews relevant research studies and identifies the limitations of existing approaches.

A. Intent-Based Educational Chatbots

Baha et al. (2022) proposed an adaptive educational chatbot using the Xatkit framework. The system employed a transformer-based encoder using CamemBERT for understanding student queries and an intent classification decoder for identifying student requirements. The chatbot was designed to support educational activities by simplifying communication between students and instructors. However, the system relied on predefined intents and responses, which restricted its ability to handle unexpected or complex student queries. The study highlighted the importance of developing more adaptive conversational systems for educational environments.

B. Intent-Based and LLM-Powered Teaching Assistants

Alsafari et al. (2024) presented a study titled “Towards Effective Teaching Assistants: From Intent-Based Chatbots to LLM-Powered Teaching Assistants.” The researchers investigated the development of educational chatbots for Data Mining and Text Analytics courses at the University of Leeds. Initially, an intent-based chatbot was developed using Amazon Alexa. The study later explored a Retrieval-Augmented Generation approach using LangChain and GPT-3.5 Turbo. The findings demonstrated the limitations of traditional intent-based chatbots in handling diverse student queries and highlighted the potential of LLM-based systems for improving adaptability and response generation. However, the use of proprietary LLMs can introduce dependency on external services and associated operational costs.

C. Large Language Models in Education

Yan et al. (2024) conducted a systematic scoping review titled “Practical and Ethical Challenges of Large Language Models in Education.” The study reviewed 118 peer-reviewed research papers and identified several applications of LLMs in education, including teaching support, feedback generation, assessment, content creation, and personalized learning. The researchers also discussed practical and ethical challenges associated with the adoption of LLMs in educational environments. The review emphasized the importance of addressing reliability, privacy, and responsible usage when integrating AI into educational systems. These findings provide a foundation for developing educational assistants that offer flexible and reliable academic support.

D. Retrieval-Augmented Generation for Question Answering

Lewis et al. (2020) introduced the concept of Retrieval-Augmented Generation in their research paper titled “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks.” The proposed approach combines a pre-trained language model with an external knowledge retrieval mechanism. Instead of relying entirely on the information stored within the model parameters, the system retrieves relevant documents and uses them as contextual information during response generation. The study demonstrated that RAG can generate more specific and factual responses than traditional generation models in knowledge-intensive tasks. This approach is highly relevant to educational assistants, as it enables responses to be generated using academic materials such as lecture notes, textbooks, and course documents.

E. Dense Passage Retrieval for Academic Question Answering

Karpukhin et al. (2020) proposed “Dense Passage Retrieval for Open-Domain Question Answering.” The study introduced a dual-encoder architecture that converts questions and documents into dense vector representations. These representations enable the system to retrieve relevant passages based on semantic similarity rather than relying only on keyword matching. The proposed method demonstrated improved retrieval performance compared with

traditional sparse retrieval approaches on several question-answering benchmarks. This research provides a foundation for implementing semantic search in AutoTA, allowing students to retrieve relevant academic information even when their queries use different wording from the original study materials.

F. Open-Source LLM-Based Educational Assistance

Recent research on generative AI in education has explored the use of locally deployed language models and retrieval-based architectures to improve academic support while maintaining greater control over educational data. A 2025 study on a local generative AI teaching assistant integrated a language model with RAG, an embedding model, and a vector database. The system was designed to support campus learning through intelligent question answering and academic assistance. The reported findings indicated improvements in response accuracy and semantic relevance when retrieval was integrated with the language model. This approach supports the development of educational systems that can operate within controlled environments and use institution-specific academic resources

G. Research Gap

The existing literature demonstrates significant progress in educational chatbots, intent classification, and LLM-based question-answering systems. However, several limitations remain:

1. Traditional intent-based chatbots depend on predefined intents and responses, limiting their ability to handle complex and unexpected student queries.
2. General-purpose LLMs may generate inaccurate or irrelevant responses when they lack access to verified academic resources.
3. Some educational assistants depend on proprietary AI services, which can introduce operational costs and reduce customization flexibility.
4. Existing systems may not dynamically adapt their response generation strategies according to the specific intent of a student query.
5. There is a need for an integrated framework that combines intent classification, academic knowledge retrieval, and open-source LLMs to provide context-aware educational assistance.

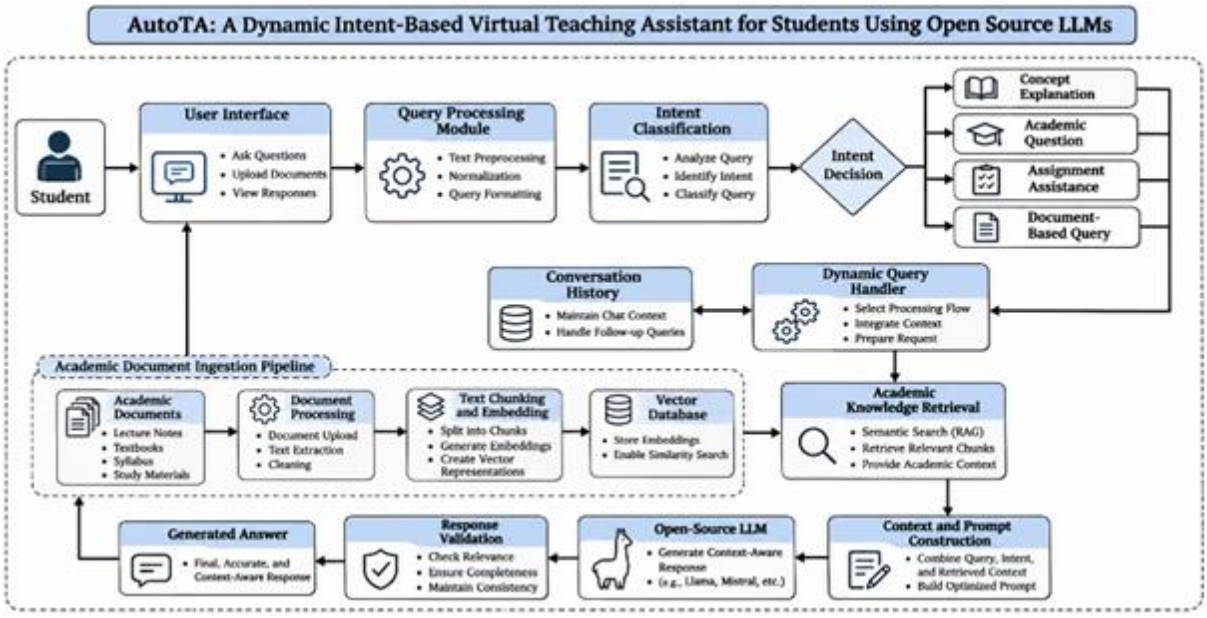
System Architecture

The proposed AutoTA system follows a modular architecture that integrates a user interface, query processing module, intent classification module, academic knowledge retrieval system, and open-source Large Language Model (LLM). The primary objective of the architecture is to provide students with accurate, context-aware, and personalized academic assistance through natural language interaction.

The system accepts student queries through a web-based interface and processes them using Natural Language Processing (NLP) techniques. The intent classification module identifies the purpose of the query, while the retrieval module searches relevant academic resources when required. The retrieved information and identified intent are then passed to the open-source LLM, which generates a suitable response. Finally, the response is displayed to the student through the interface.

The architecture also includes a document management module that allows academic materials such as lecture notes, textbooks, and syllabus documents to be uploaded, processed, and stored for semantic retrieval.

B. Architecture Components



1. User Interface

The user interface provides an interactive environment through which students can communicate with the virtual teaching assistant. Students can enter academic questions, upload learning materials, and receive responses in natural language. The interface is designed to support simple and accessible interaction.

2. Query Processing Module

The query processing module receives the student's input and performs preprocessing operations such as text normalization, removal of unnecessary characters, and query formatting. This module prepares the input for further analysis and ensures that the query is processed in a consistent format.

3. Intent Classification Module

The intent classification module identifies the underlying purpose of the student's query. The system categorizes queries into different intents, including:

- Conceptual explanation
- Academic question answering
- Assignment assistance
- Document-based question answering
- General educational queries

The identified intent determines the appropriate processing strategy for generating the response.

4. Dynamic Query Handler

The dynamic query handler selects the appropriate processing workflow based on the identified intent. For example, a conceptual question may require a simplified explanation, whereas a document-based query may require retrieving information from uploaded academic materials.

This component enables AutoTA to adapt its response generation process according to the student's requirements.

5. Academic Knowledge Retrieval Module

The academic knowledge retrieval module uses Retrieval-Augmented Generation (RAG) to retrieve relevant information from academic resources.

The document processing pipeline includes:

- Document uploading and extraction
- Text preprocessing
- Text chunking
- Embedding generation
- Vector database storage

When a student submits a document-related query, the system converts the query into an embedding and performs semantic similarity search against the stored academic materials. The most relevant document chunks are retrieved and passed to the language model as contextual information.

6. Open-Source LLM

The open-source LLM acts as the primary response generation component. The retrieved academic context, student query, and identified intent are combined to construct a suitable prompt.

The model generates a response that may include explanations, examples, summaries, or step-by-step academic guidance.

Models such as Llama and Mistral can be considered for implementation, depending on the available computational resources.

7. Response Validation Module

The response validation module checks the generated output for relevance, completeness, and consistency with the retrieved academic context. It can also apply formatting rules to ensure that the response is clear and understandable.

1. This module aims to reduce irrelevant or unsupported responses before they are displayed to the student.

8. Conversation History Module

The conversation history module maintains relevant information from previous interactions. It allows AutoTA to understand follow-up questions and maintain continuity during a learning session.

For example, when a student asks, "Explain neural networks," followed by "Give me an example," the system can use the previous conversation to understand the context of the second query.

C. Working of the Proposed System

The proposed AutoTA system operates through the following sequence:

1. The student submits an academic query through the user interface.
2. The query processing module preprocesses the input.
3. The intent classification module identifies the student's intent.
4. The dynamic query handler selects the appropriate processing workflow.
5. If academic context is required, the retrieval module searches the vector database for relevant information.
6. The retrieved context, identified intent, and student query are combined to construct a prompt.
7. The open-source LLM generates a context-aware response.
8. The response validation module checks the generated answer.
9. The final response is displayed to the student through the user interface.
10. Relevant conversation information is maintained to support follow-up queries.

D. Advantages of the Architecture

The proposed architecture provides the following advantages:

- Dynamic intent handling: Adapts the response generation process according to the student's query.
- Context-aware responses: Uses academic resources to improve the relevance of generated answers.
- Open-source integration: Supports flexible model selection and customization.
- Modular design: Individual components can be updated or replaced independently.
- Personalized learning: Supports different types of academic questions and learning requirements.
- Scalability: Additional academic subjects, documents, and intents can be incorporated into the system.

4. Future Work

The proposed AutoTA system provides an intelligent and interactive platform for delivering personalized academic assistance using intent classification and open-source Large Language Models (LLMs). Although the system aims to improve the accessibility and effectiveness of learning, several enhancements can be explored in future development.

In the future, the system can be extended by incorporating advanced intent classification techniques to improve the understanding of complex, ambiguous, and context-dependent student queries. The integration of multilingual support can enable students to interact with the assistant in different regional and international languages, making the system more accessible to diverse learners. Furthermore, the implementation of voice-based interaction can allow students to ask questions and receive responses through speech, creating a more natural and interactive learning experience.

The proposed framework can also be enhanced through the integration of multimodal learning capabilities. This would enable AutoTA to process not only text-based academic materials but also images, diagrams, mathematical equations, and educational videos. The use of advanced Retrieval-Augmented Generation techniques and domain-specific fine-tuning can further improve the accuracy, relevance, and reliability of generated responses. Additionally, adaptive learning mechanisms can be incorporated to analyze student interaction patterns and provide personalized explanations according to individual learning requirements.

Future research can also focus on developing an automated evaluation framework to assess the performance of AutoTA using metrics such as response accuracy, retrieval precision, response time, and student satisfaction. The system can be deployed on cloud infrastructure to support a larger number of concurrent users while maintaining scalability and availability. Finally, integration with Learning Management Systems (LMS) such as Moodle and Google Classroom can provide a unified platform for academic content management, student interaction, and learning progress monitoring.

These enhancements can transform AutoTA into a more comprehensive, intelligent, and adaptive virtual teaching assistant capable of supporting students across different academic disciplines and learning environments.

Future Enhancements

- Multilingual Support: Enable interaction in English, Hindi, Marathi, and other languages.
- Voice-Based Interaction: Integrate speech-to-text and text-to-speech capabilities.
- Multimodal Learning: Support academic images, diagrams, equations, and video-based content.
- Adaptive Learning: Personalize explanations based on student learning needs.
- Advanced RAG: Improve retrieval accuracy and reduce unsupported responses.
- Performance Evaluation: Measure response accuracy, retrieval quality, and system efficiency.
- LMS Integration: Connect AutoTA with educational platforms such as Moodle and Google Classroom.
- Scalable Deployment: Support multiple students through optimized cloud or local model deployment.

5. Conclusion

The The proposed AutoTA system presents an intelligent and dynamic approach to providing academic assistance through the integration of intent classification, Natural Language Processing (NLP), and open-source Large Language Models (LLMs). The system is designed to overcome the limitations of traditional educational chatbots by identifying the underlying intent of student queries and selecting an appropriate response generation strategy. By incorporating Retrieval-Augmented Generation (RAG), AutoTA can retrieve relevant information from academic resources and use it to generate context-aware responses.

The proposed architecture combines query processing, intent identification, academic knowledge retrieval, and LLM-based response generation into a unified framework. The use of open-source language models provides flexibility, customization, and reduced dependency on proprietary AI services. Furthermore, the integration of academic document retrieval enables the system to support students with subject-specific explanations, assignment-related guidance, and document-based question answering.

AutoTA aims to improve the accessibility of educational resources, encourage self-paced learning, and provide continuous academic support beyond traditional classroom environments. The modular architecture also allows the system to be extended with additional educational features and learning resources. Although the proposed framework provides a foundation for intelligent educational assistance, its effectiveness depends on the quality of intent classification, retrieved information, and generated responses. Future enhancements, including multilingual interaction, voice-based assistance, adaptive learning, and multimodal support, can further improve its capabilities.

In conclusion, AutoTA demonstrates the potential of combining dynamic intent-based processing with open-source LLMs to develop an intelligent, flexible, and accessible virtual teaching assistant. The proposed system can serve as a supportive learning platform that enhances student engagement and provides personalized academic assistance in modern educational environments.

References

[1] R. Dahal, G. Murray, R. Chataut, M. Hefeida, A. K. Srivastava, and P. K. Gyawali, “AutoTA: A dynamic intent-based virtual teaching assistant for students using open source LLMs,” IEEE Access, vol. 13, pp. 118122–118134, 2025, doi: 10.1109/ACCESS.2025.3576329.

[2] B. Alsafari, E. Atwell, A. Walker, and M. Callaghan, “Towards effective teaching assistants: From intent-based chatbots to LLM-powered teaching assistants,” Natural Language Processing Journal, vol. 8, Art. no. 100101, 2024, doi: 10.1016/j.nlp.2024.100101.

[3] “Towards highly adaptive Edu-Chatbot,” Procedia Computer Science, vol. 198, pp. 397–403, 2022, doi: 10.1016/j.procs.2021.12.260.

[4] P. Lewis et al., “Retrieval-augmented generation for knowledge-intensive NLP tasks,” in Advances in Neural Information Processing Systems, vol. 33, 2020, pp. 9459–9474. [Online]. Available: https://arxiv.org/abs/2005.11401

[5] V. Karpukhin et al., “Dense passage retrieval for open-domain question answering,” in Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP), 2020, pp. 6769–6781, doi: 10.18653/v1/2020.emnlp-main.550.

[6] J. Swacha and M. Gracel, “Retrieval-augmented generation (RAG) chatbots for education: A survey of applications,” Applied Sciences, vol. 15, no. 8, Art. no. 4234, 2025, doi: 10.3390/app15084234.

[7] Y. Yan et al., “Practical and ethical challenges of large language models in education: A systematic scoping review,” British Journal of Educational Technology, vol. 55, no. 1, pp. 90–112, 2024.

[8] S. Kweon, S. Nam, H. Lim, H. Hong, and E. Choi, “A large-scale real-world evaluation of an LLM-based virtual teaching assistant,” in Proc. 63rd Annu. Meeting Assoc. Comput. Linguistics (ACL), Industry Track, 2025, pp. 850–864, doi: 10.18653/v1/2025.acl-industry.60.

[9] D. Alexander and A. P. de Vries, “In a few words: Comparing weak supervision and LLMs for short query intent classification,” in Proc. 48th Int. ACM SIGIR Conf. Research and Development in Information Retrieval, 2025, pp. 2977–2981, doi: 10.1145/3726302.3730213