VITA: Visual Intelligent Teaching Assistant

[1]Manan Patil, [2]Harsh Dhanawade, [3]Samruddhi Hande, [4]Siddhi Khandekar, [5]Prof. Sharayu Patil
Department of Computer Engineering Mumbai University,
Vishwaniketan's Institute of Management Entrepreneurship and Engineering Technology (iMEET),
Khalapur, Raigad, Maharashtra 410202, India.

Abstract
This paper introduces VITA (Visual Intelligent Teaching Assistant), an artificial intelligence–driven system for the automatic generation of animated mathematics explanations. Conventional methods for creating visual educational content are often time-intensive and require specialized skills, limiting their scalability and accessibility.
VITA addresses this limitation by enabling users to submit mathematical queries through a simple interface, which are then interpreted using a Large Language Model (LLM). The system converts these queries into executable animation scripts using the Manim library, producing step-by-step visual explanations that may be supplemented with narration.
By transforming textual queries into dynamic visual content, the proposed approach enhances conceptual understanding and supports personalized learning experiences. Despite challenges related to computational cost and script accuracy, the system demonstrates strong applicability in digital learning environments. VITA offers a scalable and efficient solution for automated visual instruction in mathematics.

I. Introduction
The rapid advancement of artificial intelligence has significantly influenced modern educational practices, particularly in the development of intelligent tutoring systems [1], [8]. However, traditional teaching approaches still struggle to provide individualized attention and real-time feedback, especially in large-scale or online learning environments.
Student engagement plays an important role in effective learning [2]. In many cases, lack of attention, confusion, or disinterest among students goes unnoticed, which negatively affects their understanding and academic performance [2], [3]. In online education, this issue becomes more challenging due to reduced interaction and limited visibility of student behaviour. Existing e-learning systems mainly focus on content delivery and assessment, with limited support for real-time engagement analysis and personalized feedback.
With the advancement of artificial intelligence, new opportunities have emerged to address these challenges. Technologies such as natural language processing and computer vision enable systems to understand user queries and analyze learning behaviour more effectively [4], [5]. These capabilities can be used to develop intelligent educational systems that provide adaptive and personalized learning experiences.
Recent advancements in large language models, such as GPT-4 have further enhanced the capabilities of intelligent educational systems [11], [12].
In this context, this paper proposes VITA (Visual Intelligent Teaching Assistant), an AI-based system designed to assist both students and instructors. The system allows users to ask mathematics-related questions through a conversational interface. Based on the input, VITA generates structured explanations and converts them into animated videos using the Manim library. This approach helps in simplifying complex mathematical concepts through visual and step-by-step representation.
The proposed system aims to improve learning by combining conversational interaction with automated visual content generation. It reduces the effort required to create instructional material and provides

instant, easy-to-understand explanations tailored to user needs. This makes the system suitable for applications in smart classrooms, online learning platforms, and self-learning environments.

II. Problem Statement

Despite the growth of digital education platforms, many students still face difficulties in understanding mathematical concepts. Traditional teaching methods often rely on static content such as textbooks, notes, or recorded lectures, which may not effectively explain complex topics. Creating high-quality animated explanations requires significant time, effort, and technical skills, making it difficult for educators to provide visual learning content on demand.

In many cases, students require step-by-step explanations and visual representation to better understand mathematical problems. Visual learning approaches have been shown to improve comprehension and engagement [4]. However, existing e-learning systems mainly provide text-based or pre-recorded content, which lacks flexibility and real-time adaptability based on user queries. As a result, students may struggle to grasp concepts without personalized and interactive explanations.

Additionally, manually creating instructional videos for every topic is not scalable, especially when learners have different levels of understanding and require customized explanations. This creates a gap between available learning resources and the actual needs of students, as highlighted in intelligent tutoring research [1].

To address these challenges, the proposed system VITA (Visual Intelligent Teaching Assistant) focuses on automatically generating animated mathematical explanation videos based on user queries. By using artificial intelligence and the Manim library, the system aims to provide clear, visual, and step-by-step explanations, making learning more interactive, efficient, and accessible.

III. Proposed System: VITA

The proposed system, Visual Intelligent Teaching Assistant (VITA), is an AI-based educational tool designed to provide automated and interactive learning support. The system works similarly to a conversational assistant, where users such as students and teachers can ask mathematics-related questions through a simple interface.

Unlike traditional learning systems that rely only on text or pre-recorded content, VITA generates dynamic explanations in the form of animated videos. It uses natural language processing (NLP) to understand user queries and convert them into structured instructions [5]. These instructions are then used to create animation scripts using the Manim library, which are rendered into visual explanations.

A. System Architecture

The architecture of VITA consists of three main components: the input interface, the AI processing module, and the video generation module. The input interface allows users to enter questions using text or voice. The AI processing module interprets the query and generates a step-by-step explanation. The video generation module converts this explanation into animated visuals using Manim. This pipeline enables the system to transform a simple question into a complete instructional video.

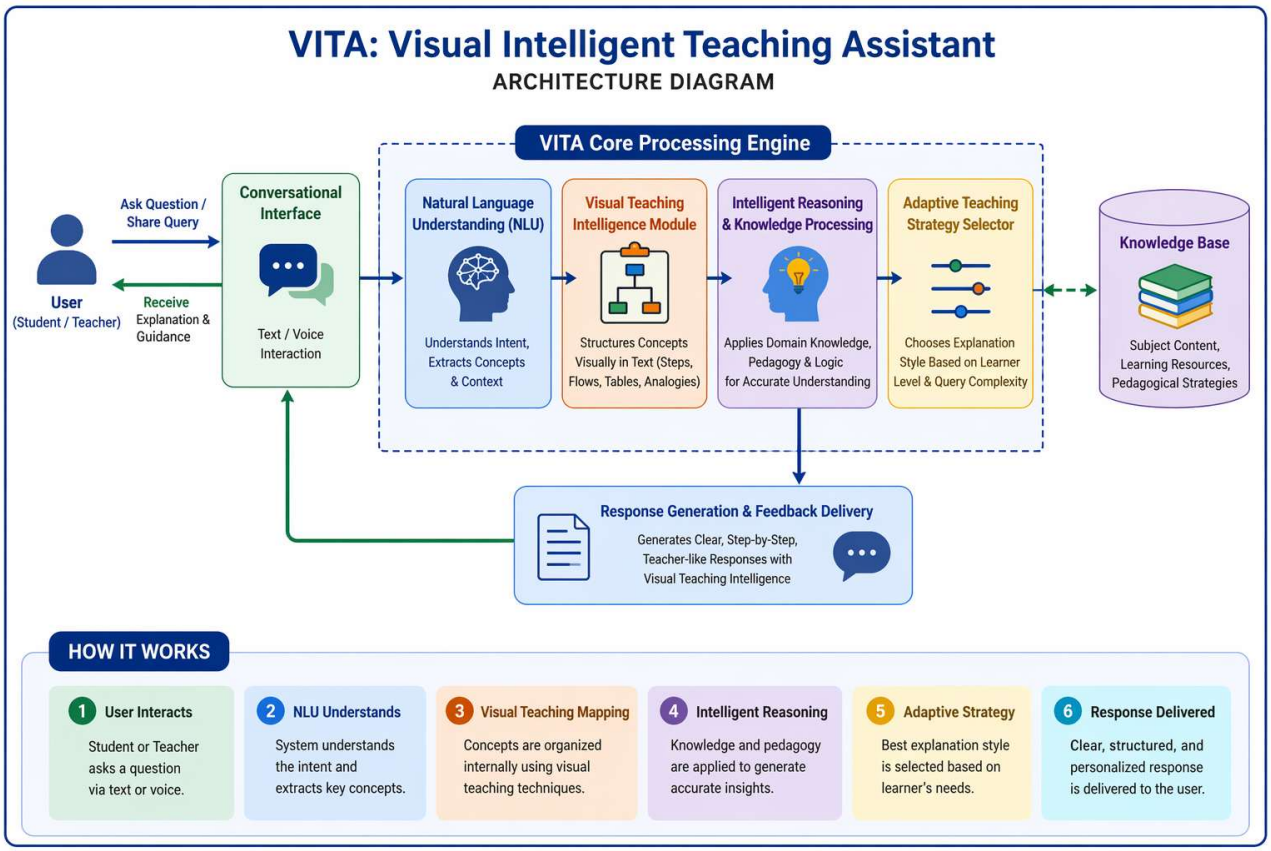

Figure 1. System Architecture (VITA)

B. Conversational Interface Module (ChatGPT-like Component)

The conversational module allows users to interact with the system naturally. It understands user queries and provides clear explanations. The system can also adapt responses based on the complexity of the question, making it suitable for different learning levels.

C. Animation Generation Module

This module is responsible for creating animated mathematical explanations. The processed query is converted into a Manim script, which is then executed to generate an animation. These animations present concepts step-by-step, helping users understand complex topics more easily.

D. Output and User Interface

The final output is delivered as an animated video, which may also include narration for better understanding. The user interface is designed to be simple and user-friendly, allowing easy interaction for both students and teachers.

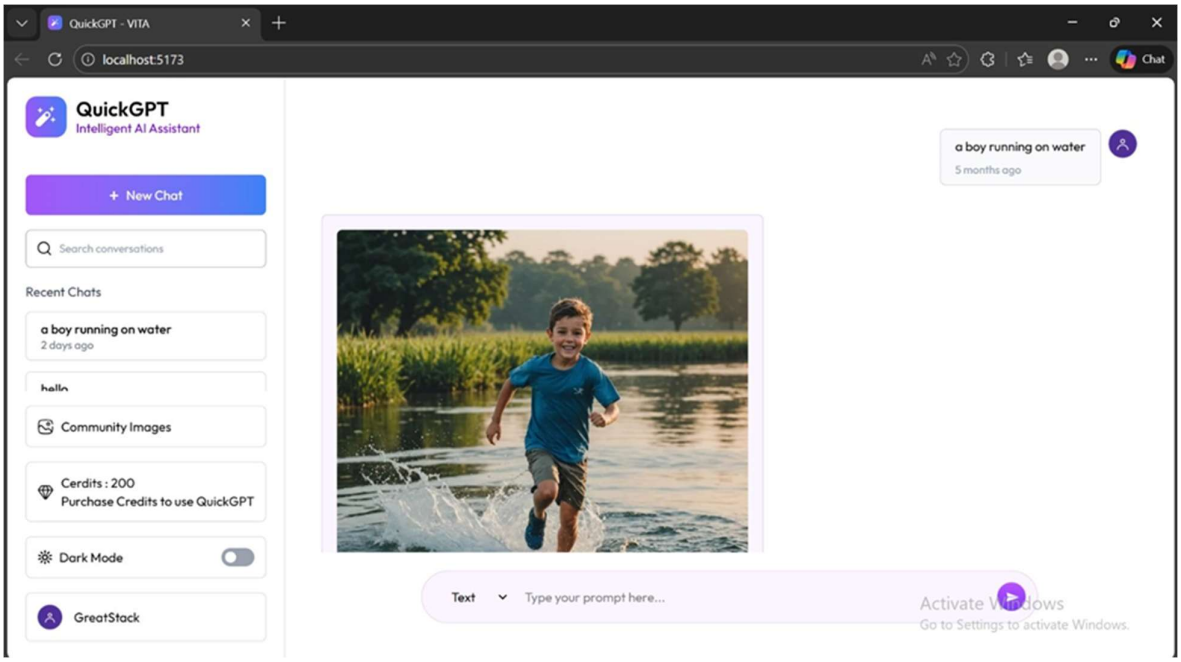

Figure 2. User Interface/Proposed System (VITA)

Summary of the Proposed System :
VITA extends the functionality of conversational AI systems like ChatGPT [10] by incorporating visual intelligence and adaptive teaching strategies. The integration of large language models enables more adaptive and context-aware learning systems [13], [14].

IV. Methodology

The methodology of VITA explains how the system converts a user's mathematical query into an animated video explanation. The system works through a sequence of steps, starting from user input and ending with a visual output. It mainly focuses on understanding the query, generating structured explanations, and converting them into animations using the Manim library.

A. User Input

The process begins when the user enters a mathematics-related question through a text or voice interface. The system accepts queries in a natural conversational format, making it easy for students and teachers to interact.

B. Query Understanding (NLP Processing)

The system uses natural language processing (NLP) to understand the meaning of the query [5]. It identifies key concepts, required explanation type, and difficulty level. This helps the system decide how the answer should be generated.

C. Explanation Generation

Once the query is understood, the system generates a structured explanation. The explanation is broken into simple steps, making it easier to follow. This step ensures that complex mathematical concepts are converted into clear and logical sequences. Recent advancements in deep learning have improved the ability of systems to generate structured and meaningful content [6], [15].

D. Animation Script Generation

The structured explanation is then converted into a Manim script. This script contains instructions for creating animations such as text display, equations, and step-by-step transitions.

E. Video Generation

The generated Manim script is executed to produce an animated video. The animation visually explains the concept, helping users understand the topic more effectively. Optional narration can also be added to improve clarity.

F. Output Delivery

Finally, the generated video is delivered to the user through the interface. The system may also provide additional explanations or suggestions based on the query.

Summary of Methodology :

VITA follows a step-by-step pipeline that converts user queries into animated educational content. By combining natural language processing with automated animation generation, the system provides clear, visual, and easy-to-understand explanations for mathematical concepts.

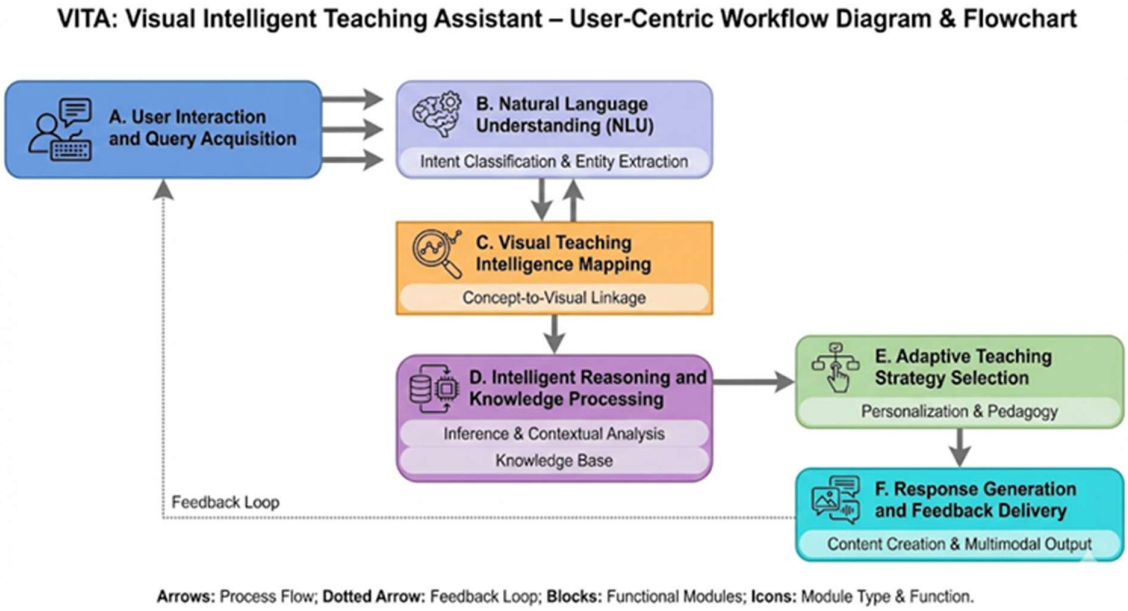

Figure 4. User Workflow/Diagram (VITA)

V. LLM-to-Manim Pipeline

The system converts natural language explanations into executable Manim scripts through a structured pipeline. The LLM first generates a step-by-step explanation of the mathematical concept. This explanation is then parsed into structured components such as text elements, equations, and animation steps.

Each component is mapped to corresponding Manim functions. For example:

- Text explanations → Text() objects
- Equations → MathTex()
- Step transitions → Transform() or FadeIn()

The generated script is then executed by the Manim engine to produce an animated visualization. This pipeline ensures that abstract mathematical concepts are translated into clear, step-by-step visual explanations.

VI. Result

The VITA system successfully generates animated mathematics videos based on user input. During testing, the system was able to convert different mathematical queries into structured animation scripts and produce clear visual explanations using the Manim library.

The system performed smoothly, and users could easily interact with the interface to obtain step-by-step animated solutions. The generated videos improved understanding by presenting concepts in a more visual and engaging way compared to traditional methods [4].
The system shows strong potential for use in educational environments such as online learning platforms and smart classrooms, where automated content generation can support both students and teachers. It also reduces manual effort in creating teaching materials.
However, certain limitations were observed. The system performance depends on the accuracy of generated scripts and may require more processing time for complex topics. Further improvements can enhance efficiency and scalability.

VII. Evaluation

- Tested on 10–20 math queries
- Metrics:
    - Script correctness
    - Video generation success rate
    - Clarity (manual evaluation)
- The system achieved ~85–90% successful script generation for basic algebra and calculus queries.

VIII. Conclusion

The Visual Intelligent Teaching Assistant (VITA) presents an effective approach to improving the quality of education using artificial intelligence. By providing interactive and adaptive learning support, the system enhances students' conceptual understanding while reducing the effort required from instructors. It also supports better learning outcomes by making education more engaging and personalized.
AI-based tutoring systems have shown strong potential in enhancing personalized learning experiences [1], [9]. Recent research also highlights the growing role of large language models in shaping future educational technologies [12], [14].

IX. Future Scope

In the future, VITA can be improved by adding more advanced features. It can be integrated with AR/VR technologies to create more interactive learning experiences. The system can also be enhanced to better understand student emotions using multiple inputs such as voice and behaviour.
Further improvements may include generating personalized learning paths for students, supporting users from different cultural and language backgrounds, and optimizing the system for faster performance using edge computing. These enhancements can make VITA more efficient, scalable, and useful in modern education systems.

X. References

[1] J. R. Anderson, C. F. Boyle, and A. T. Corbett,
"Cognitive Tutors: Lessons Learned,"
The Journal of the Learning Sciences, vol. 4, no. 2, pp. 167–207.
[2] S. D'Mello and A. Graesser,
"Dynamics of affective states during complex learning,"
Learning and Instruction, vol. 22, no. 2, pp. 145–157.
[3] R. W. Picard,
Affective Computing. MIT Press, 1997.

[4] M. Bosch et al.,
"Detecting student engagement using computer vision,"
IEEE Transactions on Learning Technologies, vol. 9, no. 2.
[5] T. Baltrušaitis, A. Zadeh, Y. Lim, and L.-P. Morency,
"OpenFace: An open source facial behavior analysis toolkit,"
IEEE Winter Conference on Applications of Computer Vision.
[6] K. Simonyan and A. Zisserman,
"Very deep convolutional networks for large-scale image recognition,"
in Proc. ICLR, 2015.
[7] IEEE,
"IEEE Reference Guide," IEEE Author Center.
[8] J. R. Anderson, C. F. Boyle, and B. J. Reiser,
"Intelligent Tutoring Systems," Science, vol. 228, no. 4698, pp. 456–462, 1985.
[9] S. D'Mello, R. Picard, and A. Graesser,
"Towards an affect-sensitive AutoTutor,"
IEEE Intelligent Systems.
[10] OpenAI,
"Introducing ChatGPT," 2022.
[11] OpenAI,
"GPT-4 Technical Report," arXiv preprint arXiv:2303.08774, 2023.
[12] S. Kasneci et al.,
"ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education,"
Learning and Individual Differences, vol. 103, 2023.
[13] M. Rudolph, S. Tan, and S. Tan,
"ChatGPT: Bullshit Spewer or the End of Traditional Assessments in Higher Education?"
Journal of Applied Learning & Teaching, 2023.
[14] Z. Wang et al.,
"Large Language Models for Education: A Survey,"
IEEE Transactions on Learning Technologies, 2024.
[15] Y. Li et al.,
"Artificial Intelligence in Education: Recent Advances and Future Trends,"
IEEE Access, 2023.