International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0548
e ISSN: 2584-2854
Volume: 03 Issue: 12
December 2025
Page No: 3452 - 3457

Youtalk: Interactive Q&A System for YouTube Videos

Prekshitha S S[1], Mahalakshmi B[2], Pragathi N[3], Sanketh S[4], Prajwal H[5]
1–5 Department of Computer Science and Engineering, AMC Engineering College, Bangalore-560083, Karnataka, India.
Emails:*prekshithaprekshitha695@gmail.com[1],mahalakshmi.balakrishnakumar@amceducation.in[2], pragathi123n@gmail.com[3], sanketh.appu524@gmail.com[4],prajwalwanitha22@gmail.com[5]*

Abstract
The way we learn has undergone a significant transformation with the rise of educational content on platforms like YouTube. However, despite the vast amount of information available, viewers often interact with videos passively, which can hinder deep understanding and retention of concepts. To address this challenge, we introduce YouTalk, an innovative interactive Question Answering (QA) system designed to enhance user engagement with educational and informational YouTube videos. Our system transforms passive video consumption into an active and participatory learning experience. It enables learners to ask questions in real time, addressing confusion or curiosity as it arises. Leveraging Artificial Intelligence (AI) and Natural Language Processing (NLP), the system interprets user queries and delivers relevant answers. A key feature of YouTalk is its ability to generate context-specific Q&A pairs linked to precise video timestamps, allowing learners to revisit important moments and improve retention and comprehension. By enhancing interactivity and personalization in video-based learning, You Talk empowers students, educators, and lifelong learners by creating a tailored and engaging educational experience.

Keywords: *Artificial intelligence; Interactive learning; Natural language processing; Question answering system; Video-based learning*

1. Introduction

The way we consume educational and informational videos is on the verge of a revolution. Traditional video watching has become a passive experience, leaving viewers to absorb information with-out truly engaging with it. This can lead to shallow understanding, limited knowledge retention, and a missed opportunity for deeper exploration. Our project seeks to change this by introducing an in-novative system that transforms passive viewing into an active learning process. At the heart of our system is the ability for users to ask questions directly within the video playback interface at spe-cific timestamps. When a concept is unclear or curiosity is piqued, viewers can pause the video and type their query. This context-sensitive in-teraction is crucial, as it allows for timely clari-fication and deeper exploration of topics as they unfold. Our system emulates the benefits of direct interaction with an instructor or participation in a study group, encouraging critical thinking and active knowledge seeking. Our system leverages advanced Artificial Intelligence (AI) and Natural Language Processing (NLP) techniques to power the interactive Q&A experience. When a user sub-mist a question, our AI algorithms analyze and interpret the query to discern its intent, identify key informational needs, and understand nuances of natural language. The AI then employs a dual strategy for answer retrieval, searching both the video's intrinsic content and extensive external knowledge bases to provide accurate and relevant answers. What sets our system apart is its ability to persistently store user-generated questions and AI-powered answers, creating a personalized layer of contextual annotations throughout the video. Upon subsequent viewings, users can revisit these annotations, review previous questions, and refresh their understanding of the answers provided. This reinforces learning, creates a unique educational artifact, and helps users track their progress. Our project aims to create a more engaging, effective, and personalized learning environment within the familiar context of online video. By

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0548
e ISSN: 2584-2854
Volume: 03 Issue: 12
December 2025
Page No: 3452 - 3457

fostering ac-tie inquiry, providing instant feedback, and creating a durable record of learning interactions, our system empowers users to deepen them understand-in, retain information longer, and cultivate a more proactive approach to education. This innovation has the ability to significantly improve the educational utility of online video platforms for various users, including students, lifetime learners and ma-trial creators.

2. Objective

The primary goal of our project is to design, grow and implement an innovative" interactive Q&A" system that revolutionizes the way users engage with videos. To achieve this, we have identified major objectives, including enabling timestamped questions, answering the AI-managed question, us-ing multi-source information recovery, ensuring data firmness, and timestamp-based recover and performance facilities. Additionally, we aim to adapt NLP models and information recover pro-cases for high relevance and accuracy, develop a user -friendly interface, handle data alignment, de-sign a scalable system, and implement user authentication (optional). By achieving these objectives, we aim to create a system that provides a spontaneous and spontaneous experience, able to ask users to ask questions, get accurate answers and maintain more information effectively, ultimately fostering a deeper understanding of complex topics and promoting a more effective learning experience.

3. Purpose

Imagine to be able to ask that moment a concept confuses you or conspiracy. Our system makes this possible, providing immediate and context-specific answers that help solidify your understand-ing. No more videos stop, score the Internet for clarification, or lose speed. Our system also al-lows you to save and revisit your questions and answers within the video timeline. This feature serves as a personalized learning companion, en-abling you to reinforce your understanding and quickly access previously sought information dur-ing review. Whether you are a student, teacher, or lifelong learner, our platform provides you the equipment that you need to learn more efficiently and effectively. Our ultimate goal is to democratize access to high-quality educational content, making it more engaging, interactive, and effective for learners of all backgrounds and abilities. By harnessing the power of technology, we're committed to creating a learning experience that's both enjoy-able and impactful – one that helps you achieve your goals and reach your full potential.

4. Scope

The scope of our project is carefully prepared to bring a strong and interactive Q&A system to life, while also acknowledging the limitations and boundaries that will guide our development pro-cess. We'll be focusing on creating a web-based interface that seamlessly integrates a You Tube video player with a Q&A functionality, allowing users to ask questions and receive accurate answers in a timely manner. This will involve develop-ing a range of features, including a user interface that embeds a YouTube video player, timestamp capture, question input mechanism, backend API for Q&A submission, video transcript retrieval, NLP processing module, external search integra-tion, answer generation or extraction, database stor-age, and API for Q&A retrieval. We will also ap-ply a basic user association facility, which will allow users to link their saved Q&A to their pro-file. However, we also know about all the bound-aries and exclusion that will shape the scope of our project. For instance, while we aim to provide low-latency question answering, we can't guaran-tee instantaneous responses for complex queries due to processing time and API dependencies. Our system will mainly depend on textual content such as transcript, and will not focus on understanding or answering questions based on visual elements or complex audio nuances. Additionally, we'll be focusing on supporting question answering in a sin-gle primary language, utilizing existing YouTube transcripts/captions, and won't be developing a browser extension, mobile application, or commu-nity/collaborative features. Furthermore, we can't guarantee 100% factual accuracy or perfect an-swers for all questions, and our error handling and edge case management will be refined over time. Finally, our UI/UX design will prioritize the main functionality, with the design simplicity and ease of use. By acknowledging these limitations and exclusions, we can ensure that our project stays

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0548
e ISSN: 2584-2854
Volume: 03 Issue: 12
December 2025
Page No: 3452 - 3457

focused on delivering a high-quality, interactive Q&A system that meets the needs of our users. interaction, ultimately provides a spontaneous ex-perience for users and brings revolution in the way it interactions with online materials.

Table 1 Software Requirements

Category	Technology / Tool
(Development & Server) Operating System	Windows 10/11, ma-cOS, or Linux (e.g., Ubuntu 20.04+)
Backend Lan-guage & Frame-work	Python (3.8+) with Flask or Django
Frontend Frame-work	JavaScript with React.js
AI / NLP Li-braries	Transformers (Hugging Face), PyTorch / Tensor-Flow
Code Editor / IDE	Visual Studio Code
Database	MongoDB or Post-greSQL
(End-User) Client Software	Modern Web Browser (e.g., Chrome, Firefox, Safari, Edge)
(Deployment) Cloud Platform	Amazon Web Services (AWS), Google Cloud Platform (GCP), or Azure

TABLE 1: The system requires a modern operating system such as Windows, macOS, or Linux to run smoothly. It uses Python (3.8+) with frameworks like Flask or Django for backend development, while the frontend is built using JavaScript and React.js. AI functionality is powered by advanced libraries such as Hugging Face Transformers, Py-Torch, and TensorFlow. Developers primarily use Visual Studio Code as the code editor for build-ing and maintaining the project. For data storage, the system supports popular databases like Mon-goDB or PostgreSQL. Finally, deployment is sup-ported on major cloud platforms including AWS, Google Cloud Platform, or Microsoft Azure, and users can access the system through any modern web browser. The combination of these technologies ensures reliable performance, scalability, and seamless integration across system components.

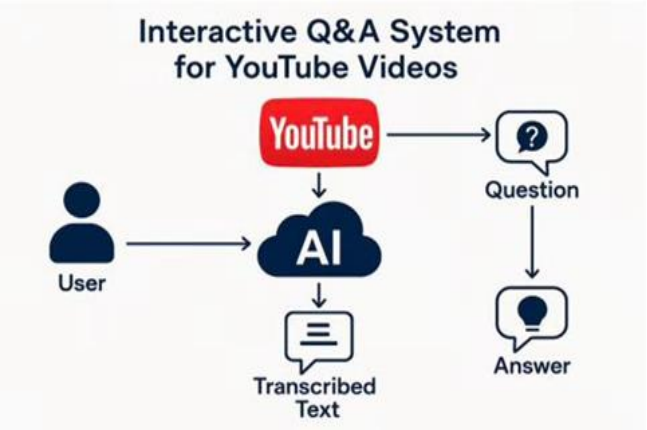

Figure 1 Architecture Design

Figure 1: illustrates the full workflow of the Interactive Q&A System for YouTube videos. The system is designed to handle both real-time and de-layed question answering, allowing users to pause at any moment in the video and ask a question. It then provides intelligent, context-aware responses based on the exact point in the video where the query was raised.

5. Results

The design of the experiments were centered around evaluating the system's core functionalities through controlled testing. The primary objective was to ensure that users can present timestamp-based questions during YouTube video playback, obtaining relevant AI-in-operated answers that is generated using natural language processing (NLP) and video tape, and have these Q&A pairs persis-tently stored and retrieved at the appropriate video timestamps. To comprehensively assess the sys-tem, various types of testing were conducted, in-cluding functional testing of the end-to-end Q&A flow, compatibility testing across different devices and browsers, and performance testing focusing on response time and scalability. The results of the experiments have shown that the system effectively captures the exact video

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0548
e ISSN: 2584-2854
Volume: 03 Issue: 12
December 2025
Page No: 3452 - 3457

moment when a question is asked and generates suitable answers relevant within 2-5 seconds. The Q&A pairs were accurately saved and retrieved during subsequent video viewings. The application performed consistently across multiple platforms including Windows, ma-cOS, Linux, Android, and iOS, and was compatible with major browsers such as Chrome, Firefox, Safari, and Edge. The user interface proved to be responsive and adaptive to various screen sizes, from desktop monitors to mobile devices. Additionally, the system handled errors gracefully by providing clear feedback for issues such as invalid inputs, missing transcripts, or AI delays.

Figure 2 Home Page Server

Figure 2: illustrates the home page of the application, designed as the main interaction point with a clean, responsive layout for enhanced us-ability. It features a welcome message, an intu-itive navigation bar, an interactive QA panel for timestamp-based questions, and a trending ques-tions section to boost engagement. The interface integrates smoothly with the YouTube API and backend engine, ensuring a seamless learning ex-perience across both desktop and mobile devices.

Figure 3 Login Page

Figure 3: illustrates the login interface allows registered users to access their QA history, track answered questions, and engage in community discussions. The authentication process ensures secure access, preventing unauthorized activity responses using AI-generated answers and user-provided insights. Users can upvote/downvote answers, ensuring the most relevant responses appear at the top. Figure 4: illustrates the system presents

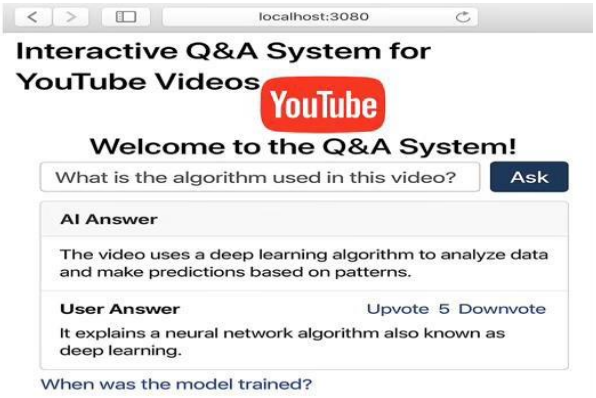

Figure 4 Output Page

Discussion

The experimental results shows that the interac-tive Q&A system successfully transforms passive video watching into an active learning experience. The ability to submit timestamp-based questions and receive AI-generated answers within seconds indicates a significant step toward contextualized, real-time support for learners. This shows the system's potential to bridge the gap between traditional classroom interaction and asynchronous video-based education by simulating an instructor-like presence that responds to user queries as they arise. Moreover, the consistent performance across different platforms and devices reflects the system's accessibility and scalability. This cross-platform reliability ensures that users from different educational and technical backgrounds can engage with the system without facing compatibility issues. The responsive design and intuitive interface further contribute to a seamless user experience, which is crucial for encouraging regular use in both academic and informal learning settings. The result is an effective use of AI and NLP

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0548
e ISSN: 2584-2854
Volume: 03 Issue: 12
December 2025
Page No: 3452 - 3457

in understanding and generating relevant reactions. While the response time of 2–5 seconds is not instantaneous, it is well within acceptable limits for educational applications, and it reflects a bal-ance between computational complexity and user experience. Additionally, the persistent Q&A stor-age and personalized annotation retrieval demon-strated that learners can revisit and reinforce pre-vious queries highlighting the system's value as a revision and reflection tool. The upvote/downvote mechanism also detects an important behaviour in-sight: users attach more when they can contribute to the quality of the material, which adds a collabo-rative and community-operated dimension to learn-ing. Overall, the results indicate that the system addresses key limitations in current video learn-ing platforms—specifically the lack of real-time interaction, contextual question answering, and long-term learning traceability. However, minor limitations such as dependence on available tapes and the absence of visual/audio visual interpreta-tion highlight areas for future growth

Conclusion

In this project, we introduced an interactive Ques-tion & Answering (Q&A) system designed to en-hance the learning experience on online video plat-forms like YouTube. Our system empowers a user, for instance, a student named Alex, to actively engage with educational video content by asking questions at specific timestamps. The system uses Artificial Intelligence (AI) and Natural Language Processing (NLP) to understand these questions and provides relevant answers, which are drawn from either the video or comprehensive external knowledge sources of the video. A major innova-tion is the frequent storage of in Q&A interactions, which creates individual, relevant annotations that can recreate Alex. It turns into an active, participa-tion process to watch passive videos. The system ensures that users can get clarification on time and create a deep understanding without signifi-cantly interrupting their learning flow. We have underlined a structure that displays a practical and effective approach to make video-based learning more dynamic, individual and user-central, which improves understanding and knowledge retention and proves to be more efficient to the users.

Acknowledgements

I am genuinely grateful to all those who have played a part in supporting me throughout the course of this project. To begin with, I would like to extend my sincere thanks to my project guide, Assistant Professor Mahalakshmi B, for her contin-uous support, insightful suggestions, and encour-agement, which have been instrumental in shaping the direction and progress of this work. I am also indebted to the researchers whose work formed the backbone of this study - Pragathi N, Prajwal H, Prekshitha S S, and Sanketh S. Their published research offered essential insights and served as a valuable reference for developing the methodology and structure of this project. My appreciation also goes to AMC Engineer-ing College and the Department of Computer Sci-ence and Engineering for providing the facilities, resources, and an encouraging academic environ-ment necessary to carry out this research success-fully. Finally, I would like to thank my family and friends for their constant motivation, emotional support, and belief in me throughout this journey. This project has been made possible only through the guidance and contributions of all the individuals mentioned above. I sincerely thank each one of you.

References

[1]. M. Waseem, M. U. G. Khan, and S. K. Khur-shid, "LCGD: Enhancing Text-to-Video Gen-eration via Contextual LLM Guidance and U-Net Denoising," IEEE Access, vol. 13, pp. 47068–47086,Mar.2025,doi:10.1109/ACCESS.2025.3550945.

[2]. J. Gonzalez-Dominguez, D. Eustis, I. Lopez-Moreno, A. Senior, F. Beaufays, and P. J. Moreno, "A Real-Time End-to-End Multi-lingual Speech Recognition Architecture," IEEE J. Sel. Topics Signal Process., vol. 9, no. 4, pp. 749–759, Jun. 2015, doi:10.1109/JSTSP.2014.2364559.

[3]. S. Repp, A. Groß, and C. Meinel, "Browsing within Lecture Videos Based on the Chain Index of Speech Transcription," IEEE Trans. Learn. Technol., vol. 1, no. 3,

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0548
e ISSN: 2584-2854
Volume: 03 Issue: 12
December 2025
Page No: 3452 - 3457

pp. 145–156, Jul.–Sep. 2008, doi: 10.1109/TLT.2008.22.

[4]. J. Yang et al., "A Generative Adversar-ial Network-Based Extractive Text Summa-rization Using Transductive and Reinforce-ment Learning," IEEE Access, vol. 13, pp. 65490–65512, Apr. 2025, doi: 10.1109/AC-CESS.2025.3558266.

[5]. S. Debeeprasad Das, P. K. Bala, and S. Das, "Exploiting User-Generated Content in Product Launch Videos to Compute a Launch Score," IEEE Access, vol. 12, pp. 49624–49638, Apr. 2024, doi: 10.1109/AC-CESS.2024.3381541.

[6]. H. Park, Y. Chung, and J.-H. Kim, "Deep Neural Networks-Based Classification Methodologies of Speech, Audio and Music, and its Integration for Audio Metadata Tagging," J. Web Eng., vol. 22, no. 1, pp. 1–26, Apr. 2023, doi: 10.13052/jwe1540-9589.2211.

[7]. H. Zhu et al., "A Cross-Curriculum Video Recommendation Algorithm Based on a Video-Associated Knowledge Map," IEEE Access, vol. 6, pp. 57562–57575, Oct. 2018, doi: 10.1109/ACCESS.2018.2873106.

[8]. F. Shamsi and I. Sindhu, "Condensing Video Content: Deep Learning Advancements and Challenges in Video Summarization Innova-tions," IEEE Access, accepted and in press, 2025, doi: 10.1109/ACCESS.2025.3526068.

[9]. [J. Bhowmik, V. Frings-Hessami, G. C. Oliver, and M. K. Hossain, "Information Access via Voice Commands on YouTube: Empir-ical Evidence on the Consequences for a Marginalised Community in Bangladesh," Information Research, Mar. 2025, doi: 10.47989/ir30iConf46945.

[10]. A. M. Klein, K. Kölln, J. Deutschländer, and M. Rauschenberger, "Design and Evalu-ation of Voice User Interfaces: What Should One Consider?" Lecture Notes in Com-puter Science, HCII 2023, Jul. 2023, doi: 10.1007/978-3-031-35921-7_12.

[11]. Z. Kuang and X. Tie, "A Survey of Mul-timedia Technologies and Robust Algo-rithms," arXiv preprint, arXiv:2103.13477v2 [cs.MM], Mar. 2021.

[12]. A. Zadeh, M. Chen, S. Poria, E. Cambria, and L.-P. Morency, "Tensor Fusion Network for Multimodal Sentiment Analysis," in Proc. Conf. Empirical Methods in Natural Lan-guage Processing (EMNLP), Copenhagen, Denmark, Sep. 2017, pp. 1103–1114.

[13]. S. Chen, T. Yao, and Y.-G. Jiang, "Deep Learning for Video Captioning: A Review," in Proc. 28th Int. Joint Conf. Artif. In-tell. (IJCAI), Macao, China, Aug. 2019, pp. 6283–6290.

[14]. A. Arnab, M. Dehghani, G. Heigold, C. Sun, M. Lucić, and C. Schmid, "ViViT: A Video Vision Transformer," arXiv preprint, arXiv:2103.15691v2 [cs.CV], Nov. 2021.

[15]. M. Abdrakhmanova, A. Kuzdeuov, S. Jarju,Y. Khassanov, M. Lewis, and H. A. Varol, "SpeakingFaces: A Large-Scale Multimodal Dataset of Voice Commands with Visual and Thermal Video Streams," Sensors, vol. 21, no. x, Art. no. xxxx, 2021, doi: 10.3390/s1010000.