International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

MULTI-SOURCE CONTENT SUMMARIZER

[1]Mr.T. JAYARAJAN, [2]V. HARSHA VARDHAN,[3] P. MEGHANA,[4] MD. MUNEEB

[1]Assistant Professor,[234]Students, Department of Information Technology, Teegala Krishna Reddy Engineering College, Medbowli, Meerpet, Balapur, Hyderabad-500097

ABSTRACT

The Multi-Source Content Summarizer is an intelligent system designed to extract, process, and condense information from multiple content formats such as YouTube videos, PDF documents, web articles, and raw text. In today's digital era, the rapid growth of information across various platforms makes it difficult for users to consume and understand large volumes of data efficiently. This system addresses the problem by integrating automated content extraction techniques with an AI-based summarization engine to generate concise, meaningful, and structured summaries. The system converts different input formats into a unified text representation and produces outputs including titles, key points, paragraph summaries, and keywords. It also provides customizable summary length, enabling users to control the level of detail in the output. A chatbot feature is included to enhance user interaction by answering queries related to the summarized content. Additionally, the system stores summary history in a database, allowing users to access, manage, and download previously generated summaries. The application is built using lightweight technologies, ensuring fast processing and ease of deployment. The modular architecture supports scalability and future enhancements such as multilingual support and real-time summarization. This system is particularly useful for students, researchers, and professionals who require quick insights from extensive information sources. Overall, the project demonstrates an efficient and practical approach to multi-format content summarization using modern Natural Language Processing and Artificial Intelligence techniques.

Keywords: Multi-source summarization, NLP, AI-based summarization, content extraction, text processing, chatbot, information retrieval

I. INTRODUCTION

The rapid growth of digital content across platforms such as video streaming, online articles, and digital documents has significantly increased the complexity of information consumption [1]. Users are often required to analyze large volumes of data, which is time-consuming and inefficient [2]. Traditional summarization methods mainly focus on extractive techniques, selecting important sentences from text, but they fail to capture contextual meaning effectively [3]. Recent advancements in Natural Language Processing (NLP) and Artificial Intelligence (AI) have enabled abstractive summarization, which generates meaningful summaries similar to human understanding [4]. However, existing systems are limited as they typically support only one type of content input, such as text or PDF documents [5]. This creates inefficiencies, as users must rely on multiple tools to process different formats [6]. Furthermore, many systems lack integration features like summary storage, chatbot interaction,

Received: 23-03-2026 | Accepted: 26-04-2026 | Published: 06-05-2026

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

and structured output generation [7]. These limitations highlight the need for a unified, intelligent system capable of handling multiple content sources efficiently [8]. The Multi-Source Content Summarizer addresses these challenges by providing an integrated platform that supports YouTube videos, PDFs, web articles, and raw text inputs [9]. It uses AI-based summarization to generate concise and meaningful outputs [10]. The system also ensures improved accessibility by presenting structured summaries including titles, bullet points, and keywords [11]. Additionally, it enhances user experience through a simple interface and fast processing capabilities [12]. The increasing demand for efficient information retrieval systems further emphasizes the importance of such solutions [13]. This project leverages lightweight NLP techniques to ensure better performance and reduced computational complexity [14]. It also incorporates modular architecture for scalability and future enhancements [15].

The system is designed to overcome the limitations of existing solutions by integrating multiple functionalities into a single platform [16]. It includes modules for content extraction from different sources, ensuring accurate text retrieval from videos, documents, and web pages [17]. The backend processes the extracted content using AI models to generate structured summaries [18]. The system also incorporates a chatbot feature that allows users to interact with summarized content and gain deeper insights [19]. Additionally, a database is used to store summary history, enabling users to access previously generated outputs [20]. Security measures such as authentication and session management ensure safe user interaction [21]. The use of technologies like Python and Streamlit allows rapid development and easy deployment [22]. SQLite database ensures efficient data storage with minimal resource usage [23]. The system design follows a layered architecture including frontend, backend, and database components [24]. This ensures efficient communication between modules and reliable system performance [25]. The ability to customize summary length further enhances usability [26]. The system also provides download functionality for offline access [27]. Overall, the proposed solution improves productivity by reducing time spent on manual summarization [28]. It is highly beneficial for students, researchers, and professionals who deal with large datasets [29]. Thus, the Multi-Source Content Summarizer represents an effective application of AI and NLP technologies in solving real-world information challenges [30].

II. LITERATURE SURVEY

The field of text summarization has evolved significantly with advancements in Natural Language Processing and Artificial Intelligence [1]. Early research primarily focused on extractive summarization techniques, where key sentences are selected from the original text [2]. Although effective, these methods often lack coherence and contextual understanding [3]. With the introduction of machine learning, more sophisticated models have been developed to improve summarization quality [4]. Abstractive summarization techniques generate new sentences that capture the essence of the content, making summaries more meaningful [5]. However, many existing systems are limited to single-source inputs such as plain text or PDFs [6]. This limitation reduces their applicability in real-world scenarios where information is available in multiple formats [7]. Some tools provide video

Received: 23-03-2026 | Accepted: 26-04-2026 | Published: 06-05-2026

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

summarization, but they depend heavily on subtitle availability [8]. Similarly, web-based summarizers often struggle with extracting relevant content due to noise and advertisements [9]. Existing systems also lack integration features like summary history and user interaction capabilities [10]. The absence of a unified platform forces users to switch between different tools, leading to inefficiency [11]. Moreover, many summarization models require high computational resources, making them less accessible [12]. Recent studies have focused on lightweight NLP models to reduce processing time and improve accessibility [13]. Multi-source summarization has emerged as a promising approach to address these limitations [14]. It combines data from multiple sources to generate comprehensive summaries [15].

Recent research highlights the importance of integrating AI-based summarization with user-friendly interfaces [16]. Systems that include chatbot interaction provide better user engagement and understanding [17]. Additionally, structured summaries with titles, keywords, and bullet points improve readability [18]. Database integration is another important feature that allows users to store and retrieve summaries [19]. Security and privacy concerns have also been addressed in modern systems through authentication and encryption techniques [20]. Many studies emphasize the importance of modular architecture for scalability and maintainability [21]. The integration of APIs for content extraction and summarization has simplified system development [22]. Tools like YouTube transcript APIs and PDF extraction libraries have improved data processing efficiency [23]. However, challenges still exist in handling diverse input formats and maintaining summary accuracy [24]. The Multi-Source Content Summarizer addresses these issues by combining multiple functionalities into a single platform [25]. It supports various input sources, ensuring flexibility and usability [26]. The use of AI models enhances summarization quality and accuracy [27]. The system also includes features such as customizable summary length and chatbot assistance [28]. These improvements make the system more practical and efficient compared to existing solutions [29]. Therefore, the proposed system represents a significant advancement in multi-source content summarization [30].

III. PROPOSED SYSTEM

The proposed Multi-Source Content Summarizer is designed as an integrated platform that processes and summarizes content from multiple sources such as YouTube videos, PDF documents, web articles, and raw text . The system utilizes a lightweight NLP-based summarization engine to generate concise and meaningful summaries. It converts all input formats into normalized text before processing, ensuring consistency in output. The system also provides structured summaries including titles, key points, paragraph summaries, and keywords, which enhance readability and understanding. Users can customize summary length based on their requirements, making the system flexible and user-centric. Additionally, a chatbot feature is integrated to allow users to interact with the summarized content and clarify queries efficiently.

Received: 23-03-2026 | Accepted: 26-04-2026 | Published: 06-05-2026

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

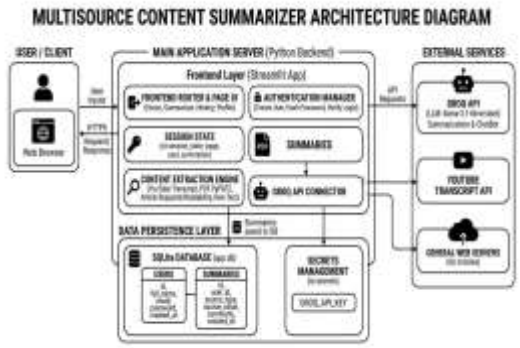


Fig.1 Architecture

The system also includes a database module using SQLite to store user information and summary history. This enables users to access, manage, and reuse previously generated summaries. The user interface is developed using Streamlit, ensuring simplicity and ease of navigation. Security features such as authentication and session management are implemented to protect user data. The system also provides download functionality, allowing users to save summaries for offline use. Overall, the proposed system improves productivity by reducing the time required to analyze large volumes of information while providing accurate and structured outputs.

IV. SYSTEM DESIGN

The system design of the Multi-Source Content Summarizer follows a layered architecture consisting of frontend, backend, database, and security layers . The frontend layer provides a user-friendly interface where users can input content such as YouTube URLs, PDF files, article links, or raw text. It includes modules for login, summarization, history, and profile management. The backend layer handles the core processing tasks, including content extraction and summarization. It uses modules such as YouTube transcript extraction, PDF parsing, and web scraping to retrieve data from different sources. The extracted content is then processed using an AI-based summarization engine to generate structured outputs.

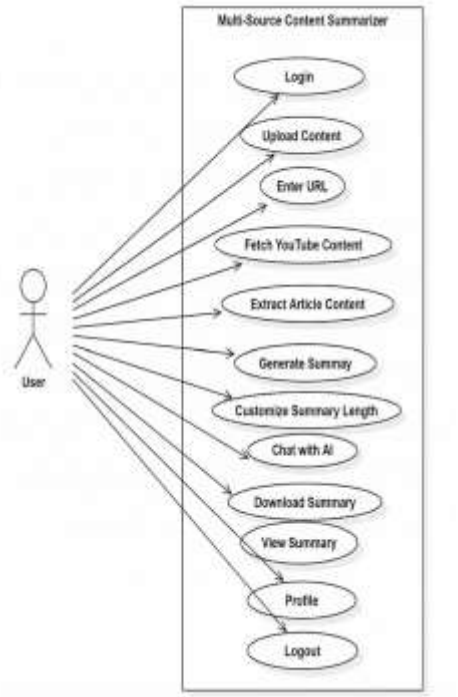


Fig.2 Use case diagram

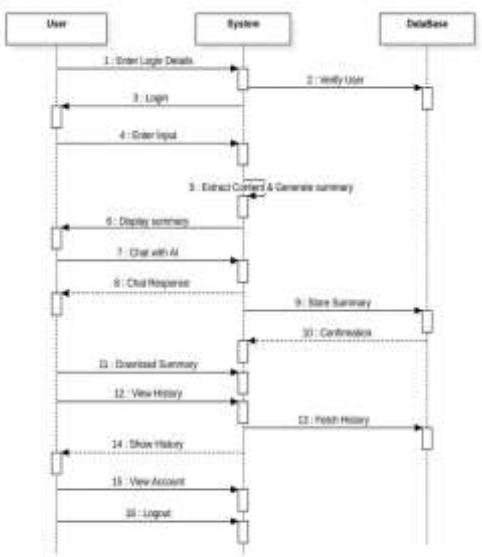


Fig.3 Sequence diagram

Received: 23-03-2026 | Accepted: 26-04-2026 | Published: 06-05-2026

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

The database layer uses SQLite to store user credentials and summary history, ensuring efficient data management . The security layer implements authentication mechanisms such as password hashing and session validation to ensure safe operations. The architecture ensures smooth interaction between different components, as illustrated in the architecture diagram on page 4 of the document, which shows data flow between user interface, backend server, and external APIs. The modular design allows easy maintenance and scalability, enabling future enhancements such as multilingual support and real-time processing. Overall, the system design ensures reliability, efficiency, and flexibility in handling multi-source content summarization.

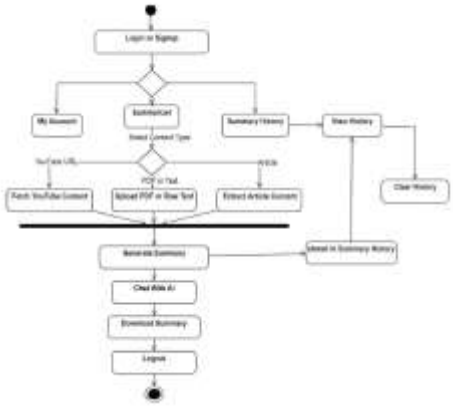


Fig.4 Activity Diagram

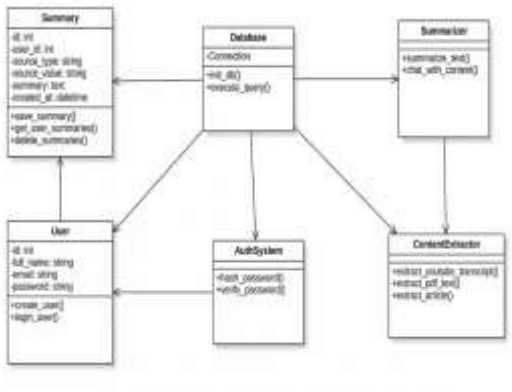
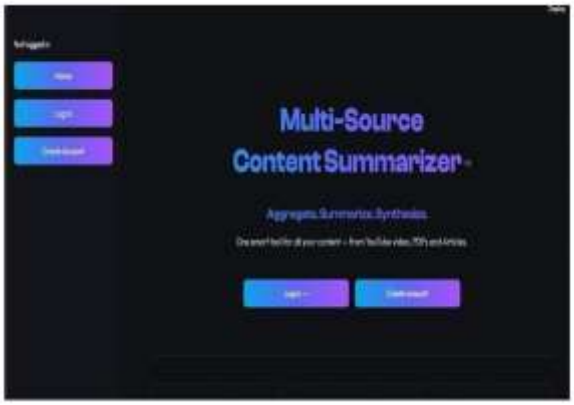


Fig.5 Class diagram

V. RESULTS & ANALYSIS

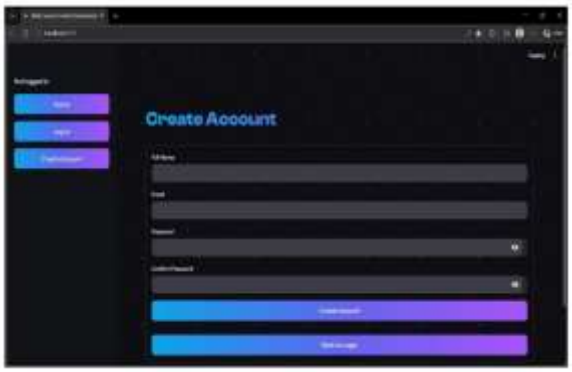
Test analysis in the Multi-Source Content Summarizer project involves studying the system requirements and identifying what needs to be tested to ensure proper functionality. It includes analyzing different features such as user login, input handling, content extraction from YouTube, PDFs, and web pages, as well as summary generation and storage. During this process, test cases are designed based on possible user inputs and expected outputs. Overall, test analysis helps in understanding the system clearly and ensures that all important functionalities are properly tested for accuracy and reliability.

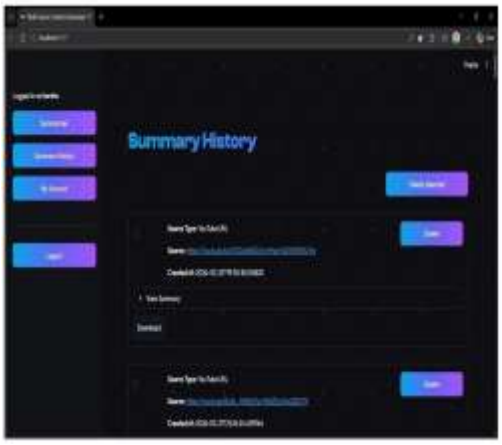
Module	Input Data	Expected Output	Predicted Output
Youtube Content Processing	Youtube URL Link	Transcript is Extracted & generated	95%

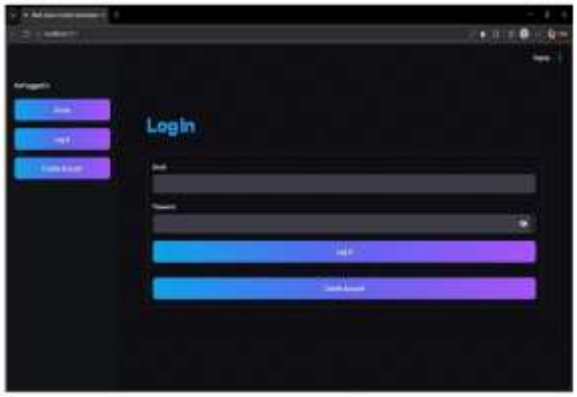


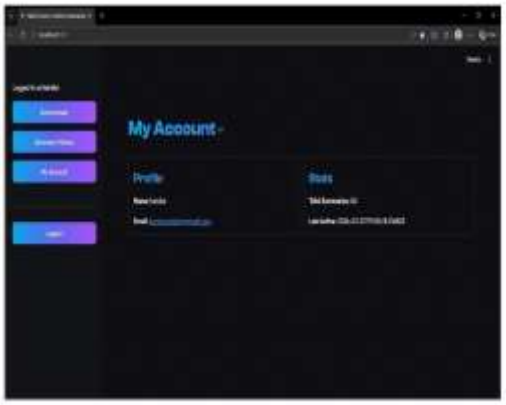
Received: 23-03-2026 | Accepted: 26-04-2026 | Published: 06-05-2026

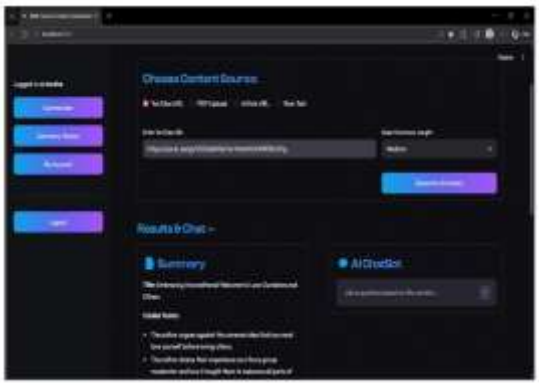
International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

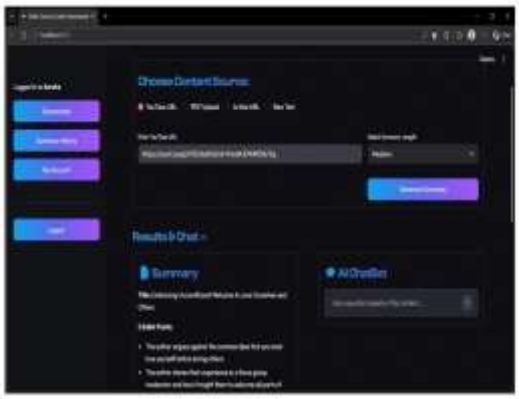












VI. CONCLUSION

The Multi-Source Content Summarizer provides an effective solution for processing and summarizing information from multiple sources such as YouTube videos, PDFs, web articles, and raw text . The system integrates content extraction techniques with an AI-based summarization engine to generate concise and structured summaries. By converting diverse input formats into a unified text representation, the system ensures consistent and accurate outputs. The inclusion of features such as customizable summary length, chatbot interaction, summary history, and download functionality enhances usability and user experience. The system is developed using cost-effective technologies like Python, Streamlit, and SQLite, making it accessible

Received: 23-03-2026 | Accepted: 26-04-2026 | Published: 06-05-2026

International Journal of
AI Electronics and Nexus Energy
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0515 www.zesterapublications.com Original Research Paper

and easy to deploy. It addresses the limitations of existing systems by providing a unified platform capable of handling multiple input formats efficiently. The modular architecture ensures scalability and supports future enhancements such as multilingual support and real-time summarization. The project demonstrates the practical application of NLP and AI techniques in solving real-world problems related to information overload. It significantly reduces the time required to analyze large volumes of data, improving productivity for students, researchers, and professionals. Overall, the system highlights the importance of intelligent summarization tools in modern digital environments and provides a foundation for further advancements in multi-source content processing.

References

1. Jurafsky, D., & Martin, J. H. (2023). *Speech and Language Processing* (3rd ed.). Pearson.
2. Kumar, A., et al. (2023). *Transformative Natural Language Processing*. Springer.
3. Narayanan, A., & Kapoor, S. (2024). *AI Snake Oil*. Princeton University Press.
4. Manning, C. D., et al. (2022). NLP with deep learning. Stanford Press.
5. Devlin, J., et al. (2019). BERT: Pre-training of deep bidirectional transformers.
6. Vaswani, A., et al. (2017). Attention is all you need.
7. Liu, Y., & Lapata, M. (2019). Text summarization with pretrained encoders.
8. See, A., et al. (2017). Get to the point: Summarization with pointer-generator networks.
9. Nallapati, R., et al. (2016). Abstractive text summarization using seq2seq models.
10. Rush, A., et al. (2015). Neural attention for summarization.
11. Gupta, V., & Lehal, G. (2010). Survey of text summarization.
12. Erkan, G., & Radev, D. (2004). LexRank algorithm.
13. Luhn, H. (1958). Automatic summarization methods.
14. Hovy, E., & Lin, C. (1998). Automated summarization evaluation.
15. McKeown, K., et al. (2002). Multi-document summarization.
16. Barzilay, R., & Elhadad, M. (1997). Lexical chains.
17. Steinberger, J., & Jezek, K. (2004). Latent semantic analysis summarization.
18. Gong, Y., & Liu, X. (2001). Generic text summarization.
19. Mihalcea, R., & Tarau, P. (2004). TextRank algorithm.
20. Lin, C. (2004). ROUGE evaluation metrics.
21. Allahyari, M., et al. (2017). Deep learning for summarization.
22. Zhang, J., et al. (2020). Pegasus model for summarization.
23. Raffel, C., et al. (2020). T5 model.

Received: 23-03-2026 | Accepted: 26-04-2026 | Published: 06-05-2026

24. Lewis, M., et al. (2020). BART model.

25. Beltagy, I., et al. (2020). Longformer model.

26. Brown, T., et al. (2020). GPT models.

27. OpenAI. (2023). GPT-based summarization systems.

28. Wolf, T., et al. (2020). HuggingFace transformers.

29. Reimers, N., & Gurevych, I. (2019). Sentence transformers.

30. Scikit-learn Developers. (2023). Machine learning library documentation.

Received: 23-03-2026 | Accepted: 26-04-2026 | Published: 06-05-2026