International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2(1) (2026)

ijerst.editor@gmail.com

editor@ijerst.com

ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

GENAI DRIVEN TEXT SUMMARIZATION FOR NEWS AGGREGATION

[1] Mohan, [2] P Akshay Chary, [3] K Sumanth, [4] K Akhil Kumar, [5] M Harshil
[1]AssistantProfessor, [2345]Students
Department of AIML
Siddhartha Institute of Technology & Sciences, Narapally

dr.m.mohnrao@siddhartha.org.in, 24tq1a66c3@siddhartha.co.in, 24tq1a6668@siddhartha.co.in, 24tq1a66b6@siddhartha.co.in, 25tq5a6609@siddhartha.co.in

Abstract

The GENAI-Driven Text Summarization for News Aggregation project is an advanced application of Generative Artificial Intelligence designed to address the growing challenge of information overload caused by the rapid expansion of digital news media. In today's digital world, users are exposed to massive amounts of news content from multiple sources, making it difficult to identify important information and extract meaningful insights efficiently. This project provides an intelligent solution by using Large Language Models (LLMs) to automatically summarize, organize, and analyze complex news articles in a concise and structured manner.

The system is developed using the Streamlit framework and powered by Llama 3.1 integrated through the Groq API to enable high-speed and low-latency AI processing. Unlike traditional text summarization systems that simply shorten articles, the proposed platform acts as an intelligent news aggregation and analysis dashboard capable of transforming unstructured news data into structured and insightful information. The system automatically extracts important topics, geopolitical entities, sentiment trends, key highlights, and contextual summaries from news content.

The core architecture utilizes a dual-engine approach consisting of an Analysis Engine and an Investigation Engine. The Analysis Engine processes news articles and converts them into structured JSON data while identifying important entities, keywords, sentiment patterns, and topic relationships. The Investigation Engine enables users to perform deeper analysis and detailed exploration of specific news topics through an interactive session-based interface. The application also integrates Plotly visualizations to generate graphical representations such as sentiment analysis charts, trend visualizations, and geographic news distribution maps, providing users with a multidimensional understanding of global events.

I. Introduction

In the modern digital era, the volume of information generated every day has grown exponentially due to the rapid expansion of online news platforms, social media, blogs, and digital publishing systems. Millions of news articles and reports are published continuously across the globe, making it increasingly difficult for users to manually read, analyze, and extract meaningful insights from large amounts of information. This phenomenon, commonly referred to as information overload,

creates significant challenges for researchers, analysts, journalists, organizations, and general users who need to stay informed in real time. The challenge is no longer accessing information, but rather filtering, organizing, summarizing, and understanding it efficiently.

Traditional news aggregation platforms mainly provide collections of headlines, article links, or short snippets, requiring users to spend considerable time reading and comparing multiple sources to understand the full context of events. These systems often lack intelligent analysis capabilities and fail to provide concise, context-aware summaries or meaningful insights from large-scale news data. As digital content continues to grow, there is an increasing need for intelligent systems capable of automatically processing and summarizing complex news narratives in a fast, scalable, and accurate manner.

Recent advancements in Generative Artificial Intelligence (GenAI) and Large Language Models (LLMs) have introduced powerful capabilities for automated text understanding, semantic analysis, and natural language generation. Modern AI models such as Llama 3.1 are capable of understanding the meaning, context, sentiment, and relationships within textual information, enabling the generation of human-like summaries and structured insights. These technologies have transformed the field of Natural Language Processing by enabling AI systems to process unstructured data and convert it into concise, meaningful, and contextually accurate information automatically.

The GENAI-Driven Text Summarization for News Aggregation project is developed to address these challenges by providing an intelligent AI-powered platform for automated news summarization and analysis. The system utilizes Large Language Models to not only shorten articles but also perform semantic synthesis by identifying important topics, extracting key highlights, analyzing emotional sentiment, and mapping geopolitical entities related to news events. This enables users to understand complex global developments quickly and efficiently without reading large volumes of raw text.

II. Literature Survey

The field of news aggregation and text summarization has experienced a major transformation over the years, evolving from traditional Natural Language Processing (NLP) techniques to advanced Generative Artificial Intelligence (GenAI) systems. In the early stages, news summarization mainly relied on Extractive Summarization Methods, where algorithms selected important sentences directly from the original text to create summaries. Techniques such as TF-IDF (Term Frequency–Inverse Document Frequency), TextRank, and graph-based ranking methods were commonly used to identify key sentences from news articles. Although these methods maintained factual consistency by preserving original sentences, they often lacked coherence, contextual understanding, and narrative flow. Extractive approaches also struggled to summarize complex news stories effectively because they could not synthesize information across multiple paragraphs or sources.

A major breakthrough occurred with the introduction of Transformer architectures, proposed by Vaswani et al. in 2017 through the “Attention Is All You Need” model.

Transformer-based models revolutionized Natural Language Processing by introducing self-attention mechanisms capable of understanding long-range contextual relationships within text. This advancement led to the development of Abstractive Summarization Systems, where AI models generate entirely new sentences rather than simply extracting existing ones. Unlike extractive systems, abstractive models understand semantics, context, and sentence relationships, enabling them to generate concise, fluent, and human-like summaries that preserve the overall meaning of the original content.

The emergence of Large Language Models (LLMs) such as GPT, BERT, T5, and the Llama (Large Language Model Meta AI) series significantly advanced the field of automated text summarization and news aggregation. Research studies conducted in 2024 and 2025 demonstrate that modern LLMs can perform not only summarization but also semantic structuring, where unstructured news articles are converted into machine-readable and context-aware formats. These systems can identify important entities, geopolitical information, emotional sentiment, topic relationships, and key events within news reports automatically. This capability improves information organization and supports intelligent news analysis and decision-making processes.

Recent literature also highlights the growing importance of Human-in-the-Loop Intelligence Dashboards in AI-driven information systems. Studies show that users prefer interactive AI systems that allow deeper exploration of summarized information through conversational interfaces rather than receiving static summaries alone. Interactive AI-driven dashboards enable users to ask follow-up questions, investigate specific events, and receive contextual insights dynamically. This approach transforms news consumption from passive reading into active information investigation and analysis.

Another important advancement discussed in current research is the integration of high-speed inference technologies such as Groq's Language Processing Units (LPUs), which significantly reduce latency in real-time AI processing systems. Earlier AI-based summarization platforms often faced performance bottlenecks due to slow inference speeds and high computational requirements. Modern inference engines enable near real-time processing of large volumes of news data, making AI-powered summarization systems more practical and scalable for real-world applications.

The literature also emphasizes the role of modern web application frameworks such as Streamlit in democratizing access to Generative AI technologies. Streamlit enables developers to build interactive AI-powered dashboards and deploy cloud-based applications efficiently without requiring complex frontend development. Combined with data visualization libraries such as Plotly, these frameworks allow AI-generated summaries, sentiment trends, and topic relationships to be presented visually, improving user understanding and interaction.

Despite these advancements, several challenges remain in the field of AI-driven text summarization. Existing systems may sometimes generate inaccurate or hallucinated summaries, fail to preserve factual consistency, or introduce bias during content generation. Managing misinformation, ensuring transparency, maintaining contextual relevance, and handling multilingual news content are ongoing research challenges.

Furthermore, balancing summarization quality with real-time performance and scalability remains a critical area of development.

The GENAI-Driven Text Summarization for News Aggregation project builds upon these advancements by integrating Large Language Models, high-speed inference systems, interactive dashboards, semantic structuring, and dynamic data visualization into a unified intelligent platform. The project aims to reduce information overload by transforming large-scale unstructured news data into concise, structured, and actionable insights for modern digital information consumption.

III. System Analysis

The GENAI-Driven Text Summarization for News Aggregation system is designed to automatically process, summarize, and analyze large volumes of digital news content using Generative Artificial Intelligence technologies. The system focuses on reducing information overload by converting lengthy and complex news articles into concise, structured, and context-aware summaries. It utilizes Large Language Models such as Llama 3.1 along with Natural Language Processing techniques to understand semantic meaning, identify key information, and generate human-like summaries dynamically. The application also performs sentiment analysis, entity extraction, and geopolitical mapping to provide deeper insights into global events. The system is developed using the Streamlit framework to create an interactive dashboard that allows users to explore summarized news content efficiently. An Investigation Engine enables users to perform detailed inquiries and multi-turn conversational analysis on specific news topics. Plotly-based visualizations transform raw textual data into charts, trend graphs, and geographic representations for improved understanding. High-speed inference using the Groq API ensures real-time response generation and low-latency performance. The system improves information accessibility, decision-making efficiency, and user interaction with large-scale digital news content. The modular architecture supports future integration of multilingual summarization, fake news detection, and predictive analytics. Overall, the system provides a scalable and intelligent solution for modern AI-powered news aggregation and summarization.

Existing System

In the existing system, news aggregation platforms mainly provide collections of headlines, article links, or short snippets gathered from multiple news sources. Users are required to manually read and analyze large amounts of information to understand the context and significance of events. Traditional text summarization systems mainly relied on extractive summarization methods such as TF-IDF and TextRank algorithms, which selected important sentences directly from source articles without generating new contextual summaries. These systems often produced fragmented summaries lacking narrative flow and semantic understanding. Existing rule-based summarization systems also struggled to handle complex news relationships, sentiment analysis, and topic interconnections effectively. Many traditional platforms provide static summaries without interactive investigation capabilities or personalized analysis. Existing systems may also suffer from information redundancy, inconsistent summarization quality, and limited contextual interpretation. Earlier AI models faced latency issues and computational limitations that reduced real-time usability for large-

scale news processing. Traditional news aggregation systems also lacked advanced visualization and semantic structuring features.

Disadvantages of Existing System

- Information overload due to excessive news content.
- Dependence on manual reading and analysis.
- Extractive summaries lack contextual understanding.
- Poor narrative flow and coherence in summaries.
- Limited semantic and sentiment analysis capabilities.
- Static summaries without interactive investigation.
- Difficulty handling large-scale real-time news data.
- Limited personalization and topic exploration.
- Lack of advanced data visualization features.
- Latency and performance limitations in older AI systems.

Proposed System

The proposed GENAI-Driven Text Summarization for News Aggregation system is designed to provide intelligent, automated, and context-aware summarization of digital news content using Generative Artificial Intelligence technologies. The system utilizes Large Language Models such as Llama 3.1 to perform abstractive summarization, semantic analysis, sentiment detection, and entity extraction dynamically. Unlike traditional extractive summarization systems, the proposed platform generates concise and human-like summaries that preserve contextual meaning and narrative flow. The Analysis Engine processes news articles into structured JSON data containing summaries, sentiment trends, geopolitical entities, and key highlights. An interactive Investigation Engine allows users to perform deeper analysis and multi-turn conversational inquiries related to specific news topics. The application also integrates Plotly visualizations to represent sentiment analysis, topic distributions, and geographic news trends graphically. Streamlit-based responsive dashboards improve user interaction and accessibility across devices. The system utilizes the Groq API for high-speed inference and low-latency response generation. The modular architecture supports future enhancements such as multilingual summarization, fake news detection, recommendation systems, and predictive analytics. Overall, the proposed system provides a scalable, intelligent, and efficient AI-driven solution for modern news aggregation and information analysis.

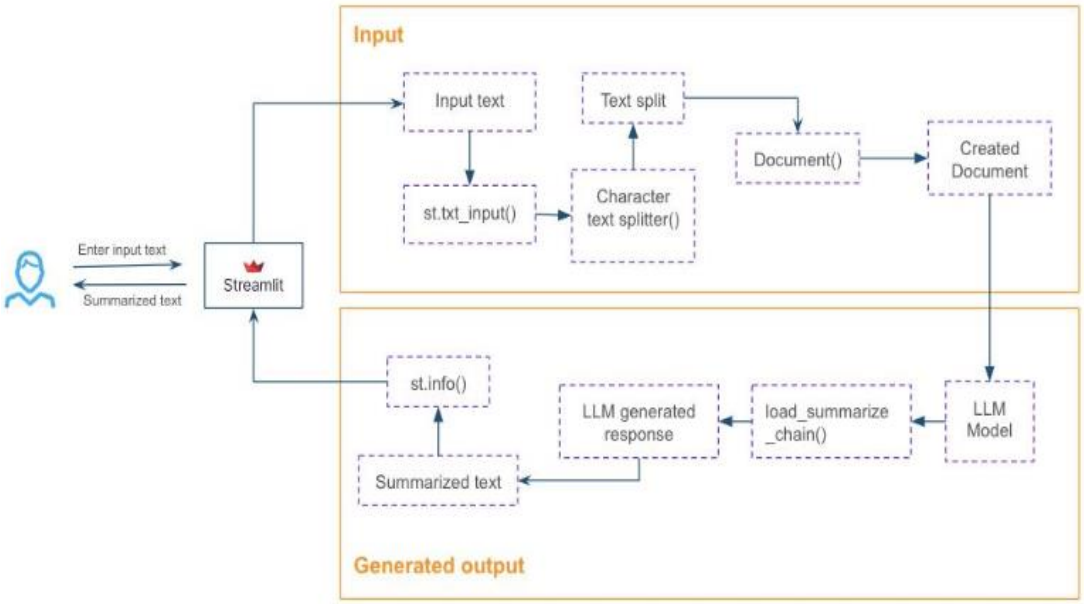
Advantages of Proposed System

- Reduces information overload effectively.
- Generates human-like and context-aware summaries.
- Supports real-time news processing and analysis.
- Provides sentiment analysis and entity recognition.
- Interactive investigation and conversational querying.
- Improved semantic understanding of news content.
- Faster response generation using high-speed inference.
- Enhanced data visualization and trend analysis.
- Scalable for large-scale digital news platforms.
- Supports future AI and multilingual integrations.

IV. Methodology

The development methodology of the GENAI-Driven Text Summarization for News Aggregation system includes data collection, preprocessing, model integration, summarization, analysis, visualization, testing, and deployment phases. Initially, news articles and digital content were collected from various online news sources for processing and analysis. Text preprocessing techniques such as tokenization, cleaning, normalization, and semantic parsing were applied to prepare the data for AI processing. Large Language Models such as Llama 3.1 were integrated to perform abstractive summarization and semantic synthesis of news content. The Analysis Engine extracts important entities, sentiment information, and key highlights while converting unstructured data into structured JSON formats. An Investigation Engine was developed to enable interactive multi-turn querying and deeper contextual exploration of summarized information. Plotly visualization tools were integrated to display sentiment trends, topic relationships, and geographic news analysis through interactive charts and graphs. The Streamlit framework was used to build an interactive and responsive dashboard interface. Testing was conducted to evaluate summary quality, contextual accuracy, response generation speed, and overall system performance. Optimization techniques were applied to reduce latency and improve scalability using the Groq API infrastructure. Finally, the system was deployed as a cloud-based AI-powered news summarization and aggregation platform. The methodology ensures scalability, maintainability, and efficient real-time AI-driven news analysis functionality.

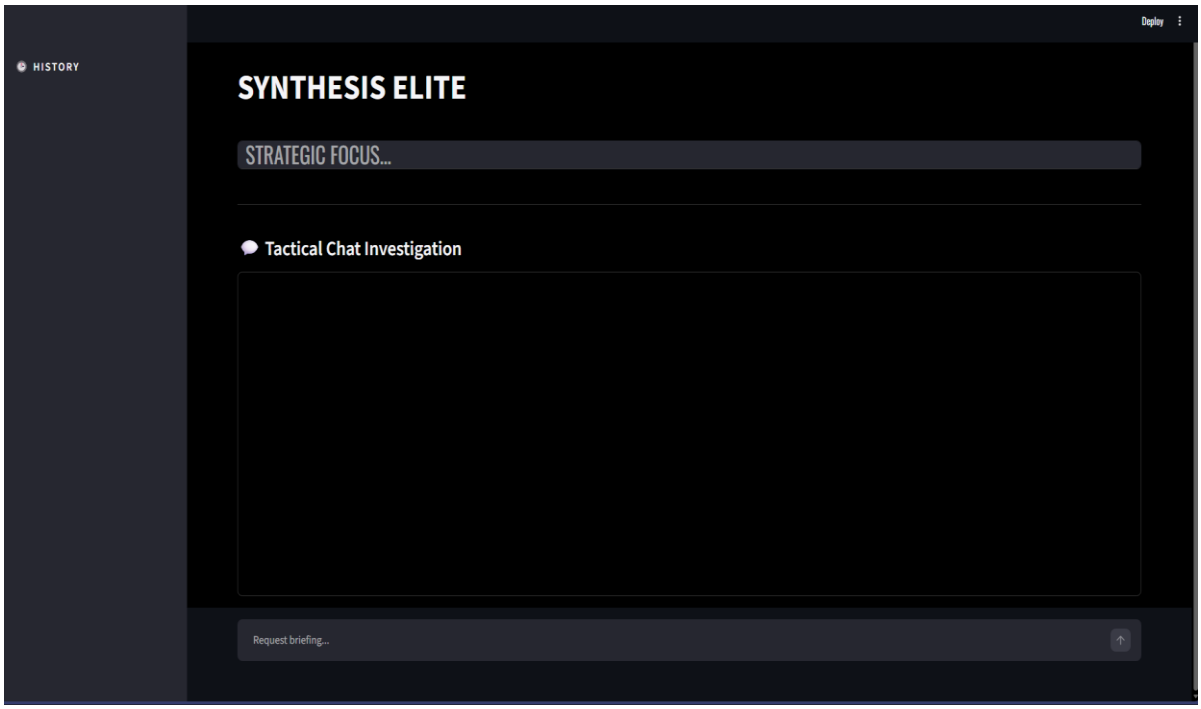
System Architecture



The system architecture of the GENAI-Driven Text Summarization for News Aggregation follows a layered client-server architecture consisting of data ingestion, preprocessing, AI processing, analysis, visualization, backend, and database layers. The data ingestion layer collects news articles and textual information from online news sources and APIs. The preprocessing layer performs text cleaning,

normalization, tokenization, and semantic preparation for efficient AI processing. The AI processing layer integrates Large Language Models such as Llama 3.1 through the Groq API to perform abstractive summarization, sentiment analysis, and entity extraction dynamically. The Analysis Engine converts processed news data into structured JSON formats containing summaries, geopolitical entities, sentiment trends, and key insights. The Investigation Engine supports conversational querying and contextual exploration of news topics. The visualization layer integrates Plotly charts, graphs, and geographic maps to provide multidimensional analysis and interactive user insights. The backend layer manages application logic, request processing, API communication, and session management operations. The database layer stores news articles, summaries, user interaction history, and analytical data securely for future reference and reporting. Security modules ensure safe API handling and protected data processing. The modular architecture also supports future integration of multilingual summarization, recommendation systems, and predictive intelligence modules. Overall, the architecture provides a scalable, intelligent, and efficient framework for AI-powered news aggregation and summarization systems.

V. Result and Output



VI. Conclusion

The GENAI-Driven Text Summarization for News Aggregation project successfully demonstrates the application of Generative Artificial Intelligence in transforming large volumes of unstructured news data into concise, meaningful, and context-aware information. By integrating Large Language Models such as Llama 3.1 with Natural Language Processing techniques, the system provides intelligent abstractive summarization, semantic analysis, sentiment detection, and entity extraction for modern digital news environments.

The project effectively addresses the growing problem of information overload by enabling users to quickly understand important news events without reading lengthy articles manually. Unlike traditional extractive summarization systems, the proposed platform generates fluent and human-like summaries while preserving contextual meaning and narrative coherence. The integration of sentiment analysis, geopolitical mapping, and structured JSON outputs further improves analytical depth and user understanding.

The implementation of interactive dashboards using Streamlit and Plotly enhances user experience by providing visual insights such as trend analysis, sentiment visualization, and geographic event mapping. The Investigation Engine allows users to perform deeper contextual analysis through interactive and conversational exploration of news topics, transforming passive news consumption into active information investigation.

The use of high-speed inference technologies through the Groq API significantly improves system performance and enables real-time summarization and analysis of large-scale news data. The scalable and modular architecture also supports future enhancements such as multilingual summarization, fake news detection, recommendation systems, predictive analytics, and advanced intelligence dashboards.

References

[1] Kumar, R. D., Prudhviraj, G., Vijay, K., Kumar, P. S., & Plugmann, P. (2024). Exploring COVID-19 through intensive investigation with supervised machine learning algorithm. In Handbook of Artificial Intelligence and Wearables (pp. 145-158). CRC Press.

[2] Swathi, B., Vijay, K., Sushanth Babu, M., & Dinesh Kumar, R. (2024, November). Machine Learning Techniques in Cloud Based Intrusion Detection. In The International Conference on Artificial Intelligence and Smart Environment (pp. 557-564). Cham: Springer Nature Switzerland.

[3] Sv satyakrishna, shirisha rangu ,bhargavi nalacheruve.(2024) Prospective investigation on colorectal cancer with SMOTE on machine learning Algorithm

[4] Dr.G.Vishnu Murthy, BhargaviNalacheruve 1Professor, Department of computer Science & engineering, Anurag University, TS, India. 2Student, Department of computer Science & engineering, Anurag University, TS, India.

[5] V. N. S. Manaswini, K. K, C. Nigam, S. S. Ali, R. Niranjana, and Suman, “Real-Time Object Detection in Drone Surveillance Using YOLOv5,” in Proc. 2025 3rd Int. Conf. IoT, Communication and Automation Technology (ICICAT), Gorakhpur, India, 2025, pp. 1–6, doi: 10.1109/ICICAT68430.2025.11414670.

[6] B. Soundarya, V. N. S. Manaswini, M. Ayyakrishnan, R. D. Kumar, “Contextual Analysis of Big Data Analytics in Intelligent Transportation Frameworks,” in Intersection of Artificial Intelligence, Data Science, and Cutting-Edge Technologies: From Concepts to Applications in Smart Environment, Lecture Notes in Networks and Systems, vol. 1353, Cham: Springer, 2025, doi: 10.1007/978-3-031-88304-0_79.

[7] R. D. Kumar, V. N. S.Manaswini, “Applications of blockchain in smart cities: detecting fake documents from land records using blockchain technology,” in Blockchain for Smart Cities, Elsevier, 2021, pp. 105–117, doi: 10.1016/B978-0-12-824446-3.00017-X.

[8] Tejavath Veeramma, Badarla Anil, Guguloth Ravinder, "An advanced movie recommender using collaborative filtering and sentiment analysis," International Research Journal of Modernization in Engineering Technology and Science, vol. 7, no. 7, July 2025, doi: 10.56726/IRJMETS81618.

[9] Ravi Kumar Banoth, Ramana Murthy B V, "Automatic crop recommendation system using LightGBM and decision tree machine learning models," Journal of Machine and Computing, vol. 5, no. 1, pp. 343, Jan. 2025, doi: 10.53759/7669/jmc202505026.

[10] Ravi Kumar Banoth, Dr. B.V. Ramana Murthy, "Smart agriculture through IoT and machine learning for analyzing carbon footprints," in Proc. Int. Conf. Computer Science and Communication Engineering (ICCSCE), Apr. 2025.

[11] Ravi Kumar Banoth, B. V. Ramana Murthy, "Soil image classification using transfer learning approach: MobileNetV2 with CNN," SN Computer Science, vol. 5, art. no. 199, 2024, doi: 10.1007/s42979-023-02500-x.