

An Artificial Intelligent based News Summarising System with Questions and Answers using Web Mining Techniques

Helen Okparaji Onungwe
Dept. of Computer Science
University of Port Harcourt
Choba, Nigeria

Joy Tochukwu Nnodi
Dept. of Computer Science & Informatics
Federal University Otuoke
Bayelsa State

Perpetua Chinazo Nwosu
Dept. of Physical Sciences
Rhema University
Aba, Nigeria

Rhema Chukwuneme Briggs-Ikeotuonye
Dept. of Computer Science
University of Port Harcourt
Choba, Nigeria

ABSTRACT

The exponential growth of digital news platforms has created a pressing challenge of information overload, making it difficult for readers to quickly access reliable and relevant updates. This paper presents an Artificial Intelligent-powered news summarizer with an integrated Question-and-Answer (Q&A) module, aimed at enhancing the efficiency and interactivity of modern news consumption. The system combines real-time news aggregation via the NewsData.io API, abstractive summarization using the BART transformer model, and Optical Character Recognition (OCR) with Tesseract to process text embedded in images. Unlike traditional platforms that provided either full-length articles or headlines, this application delivered concise, accurate summaries while enabling users to query specific details for deeper insights. The interactive Q&A feature ensured contextual understanding, while trend detection highlighted emerging topics. Implemented using React.js for the frontend and Node.js for the backend, the system demonstrated significant advantages in accessibility, relevance, and user engagement compared to existing enterprise-level summarization tools. Results indicated that the proposed solution effectively reduced reading time, mitigated information overload, and fostered trust through transparency and interactivity. This project underscored the potential of AI-driven approaches to transform digital journalism, offering a scalable, user-centric tool that bridged the gap between efficiency and credibility in news delivery.

General Terms

Artificial Intelligence, News Summariser, Journalism, Web mining.

Keywords

Artificial Intelligence, OCR, news summaries, BART transformer, digital journalism.

1. INTRODUCTION

News consumption has evolved significantly over the centuries, adapting to technological advancements that have transformed how information is disseminated and consumed. Historically, news was primarily delivered through print media, such as newspapers and magazines. The emergence of radio and television in the 20th century introduced a more immediate way for people to stay informed, making news more accessible to the

masses. However, it was the advent of the internet and digital media that revolutionized the news industry, enabling real-time updates, global connectivity, and on-demand access to information. In today's fast-paced digital era, staying informed about the latest news is essential for individuals, businesses, and decision-makers. However, with the overwhelming volume of news articles published online every minute, manually filtering through vast amounts of information to find relevant and trustworthy news can be time-consuming and challenging. Additionally, users often struggle to keep up with rapidly developing stories across multiple domains, leading to information overload and difficulty in extracting key insights.

The rapid expansion of digital news sources has introduced new challenges, including time constraints, redundancy in reporting, misinformation, and the difficulty of identifying truly trending topics. Traditional news platforms require users to sift through multiple full-length articles, many of which contain repetitive information, to gain a comprehensive understanding of a subject. The need for cross-referencing various sources to verify facts further complicates the news consumption process. Furthermore, most news platforms provide either headlines or long-form articles, without efficient summarization tools that allow users to quickly grasp the key takeaways.

1.1 Challenges of Online News Summaries

In the digital age, where news is constantly being published from thousands of sources worldwide, information overload has become a significant challenge for users. While staying informed is crucial, the sheer volume of articles, reports, and updates available online makes it increasingly difficult for individuals to extract meaningful insights efficiently. This project aims to address the fundamental challenges associated with modern news consumption, including time-consuming news consumption, irrelevant or low-quality news, a lack of summarization features, difficulty in detecting trending news, and unanswered questions after reading. With the explosion of online media, the modern reader is frequently overwhelmed by the sheer volume of information available. This phenomenon, often referred to as "information overload," has prompted researchers and developers to explore solutions that can intelligently filter, organize, and summarize content. AI-driven summarization systems aim to address this challenge by utilizing natural language processing (NLP) techniques, particularly those based on transformer

architectures, to produce summaries that retain the most critical points of a news story while discarding redundant or non-essential information.

Furthermore, modern news consumption is no longer restricted to textual articles. A significant portion of breaking news now appears in visual formats—screenshots of tweets, infographics, or news clippings shared via social media platforms. As a result, systems capable of processing both textual and visual content are increasingly essential in order to create holistic and inclusive summarization frameworks. While numerous studies have laid the foundation for AI-assisted journalism, many fall short in key areas such as real-time responsiveness, multi-modal input handling, user interactivity, and credibility verification. This research seeks to fill those gaps by introducing a system that combines news fetching via News API, summarization using the BART transformer model, image-based text extraction via Tesseract OCR, and interactive Q&A functionality. In doing so, it builds upon—and enhances—the strengths of prior models, while addressing their limitations in both theoretical and practical terms. To contextualize the novelty of this project, it is necessary to examine the contributions of earlier studies in the field, beginning with the work of

1.2 AI - powered News Summarizer

The challenges associated with news consumption have prompted media organizations to explore AI-driven solutions to enhance news consumption. For instance, *The Independent* launched "Bulletin," an AI-driven news service utilizing Google's Gemini AI model to provide article summaries. These summaries are overseen by journalists to ensure accuracy, aiming to offer "trusted, verified journalism" for readers seeking quick updates. This collaboration between AI and human editors underscores the potential of AI to enhance news delivery without compromising journalistic integrity. Similarly, *Bloomberg* (2025) introduced AI-powered news summaries to help users quickly absorb key takeaways and stay informed, demonstrating the practical applications of AI in streamlining news consumption. These developments and the trend of AI-driven systems motivate the study of an "AI-Powered News Summarizer with Questions and Answers," a web application designed to streamline the news consumption process. The platform integrates artificial intelligence (AI) to fetch real-time news articles, generate concise summaries, and extract key highlights from trending and groundbreaking news stories. This not only enables users to stay updated efficiently but also provides an interactive Question & Answer (QA) system for deeper insights into specific news topics. By combining real-time news aggregation, AI-driven summarization, and an intuitive Q&A system, the AI-Powered News Summarizer with QA aims to enhance news accessibility, improve information retention, and empower users with quick, reliable, and relevant news updates. Whether for professionals, students, researchers, or general readers, the application serves as a valuable tool to navigate the complex world of digital news efficiently. By addressing these challenges, the AI-Powered News Summarizer with QA aims to redefine how users consume news, making it more efficient, accessible, and insightful while saving time and reducing information overload.

2. RELATED WORKS

2.1 Review of AI-Powered News Summaries

[21] opined particularly on the introduction of Gen AI summarization for news content. Their approach emphasizes the integration of generative AI technologies into newsrooms to automate content creation and enhance operational efficiency. This marks a significant advancement in the field, as it explores AI's potential to manage the growing volume of news while

maintaining high-quality standards. The study's value lies in its exploration of AI's ability to summarize complex, fast-moving news stories, though its practical application and the balance between automation and journalistic integrity remain areas to be tested. [1] opined that generative artificial intelligence (GAI) GAI is deeply integrated into journalism as tools, platforms and systems. But, GAI's role in journalism dilutes the power of media professionals, changes traditional news production and poses ethical questions. Tow Center for Digital Journalism in 2024, contributed particularly in the exploration of the citation problem in AI search engines. Their research focuses on the challenges AI systems face when generating citations for news content, revealing a critical gap in the transparency and accountability of AI-driven journalism. The study highlights the need for AI models to improve their citation accuracy to maintain credibility, particularly in news reporting where factual integrity is crucial. [2], opined particularly in the evaluation and benchmarking of various transformer-based models. Their research focuses on comparing the performance of multiple large-scale pretrained models including T5-Large, PEGASUS, and BERTSUM to determine the most effective architecture for generating concise and relevant news summaries. The authors' work is important for understanding how model selection and preprocessing techniques can influence the quality, readability, and usefulness of summaries in real-world journalism applications.

2.2 Review of AI-based News summaries

The interactive Question and Answer feature embedded in this system serves as a countermeasure to the opacity issue raised by [3]. Rather than leaving users to passively consume summaries, the Q&A module allows them to query specific aspects of the content, fostering deeper engagement and understanding. This interactivity not only improves comprehension but also increases user trust, as it provides a mechanism for clarification and feedback. These functions were traditionally reserved for human journalists. Moreover, the News API integration ensures that the system fetches content from verified and reputable sources, thereby reducing the risk of training on biased or low-quality data. While this does not eliminate bias entirely, it establishes a more controlled content environment compared to models that rely on web-scraped or crowd-sourced datasets. From a performance perspective, our use of the BART transformer model builds on the principles outlined in [3] but introduces a different architecture that emphasizes both summarization and generation capabilities. BART combines bidirectional and autoregressive transformers, allowing it to perform denoising tasks and generate more coherent summaries even when the source content contains errors or inconsistencies. In this way, our system not only replicates the strengths of the [3] framework but also expands its functionality through multi-modal processing, real-time summarization, and interactive feedback loops. These enhancements move the model closer to the ideal of a comprehensive, intelligent news assistant, capable of supporting modern journalism in all its complexity. Secondly, the News API integration in our system allows it to pull in articles from multiple reputable sources in real-time, enabling multi-document summarization. This feature directly addresses [4] 's limitation by enabling a more holistic understanding of dynamic, unfolding news stories. Third, and most significantly, our platform incorporates Tesseract OCR for image-based text extraction. This component allows the system to process and summarize content embedded in images which is a major innovation not covered in [4]. As visual content continues to play a central role in news dissemination, the ability to extract and summarize information from screenshots, memes, or infographics adds a critical layer of functionality. In terms of interactivity, this research introduces a Q&A module that allows users to ask follow-up questions about the summarized content. This

transforms the user experience from a one-way interaction into a dialogue, promoting deeper understanding and encouraging critical engagement. Unlike the passive systems studied by [4], our model offers a platform that is not only informative but also conversational and responsive. Lastly, our project builds on the preprocessing insights from [4], by integrating lightweight NLP preprocessing steps such as named entity recognition and sentence segmentation—prior to feeding the text into the summarizer. However, it does so with a focus on real-time efficiency, ensuring that the added preprocessing steps do not slow down the user experience.

In summary, while [5], provide a robust foundation for understanding the relative performance of various transformer models in news summarization, their work remains limited by its scope particularly its focus on single-document processing, text-only inputs, and non-interactive design. Our project builds directly on these insights, employing a different transformer (BART), enabling multi-document summarization, and introducing both image extraction and user interactivity. These enhancements mark a meaningful step toward the creation of a more complete, accessible, and intelligent news summarization system for the modern age.

[5] provides a critical perspective on how the quality of AI-generated summaries is evaluated and consumed by real-world users. While earlier studies focus primarily on improving model architectures and output accuracy using benchmark datasets, their study shift attention to the evaluation process itself, uncovering an important discrepancy between quantitative metrics and human judgment. Their research underscores a key issue in the domain of AI-driven summarization, namely, the overreliance on automated metrics like ROUGE, BLEU, and METEOR to evaluate the performance of models. While these metrics have been instrumental in early stages of AI development, they fall short in capturing subjective qualities such as readability, narrative coherence, contextual relevance, and user satisfaction. Goyal and her colleagues argue that these intangible attributes are often more valuable than raw metric scores, particularly in the context of journalism, where readers demand clarity, nuance, and trustworthiness in content. To explore this evaluation gap, they conducted a user-centric study where participants were presented with summaries generated by GPT-3, a state-of-the-art language model known for its generative capabilities. The GPT-3 summaries were then compared to those produced by more traditional extractive and hybrid models. Interestingly, even when GPT-3's summaries scored lower on conventional evaluation metrics, users consistently rated them higher in terms of usefulness and coherence. This finding suggests a major misalignment between what machines evaluate as "good" summaries and what humans actually value when reading news content. The authors argue that an overdependence on surface-level text overlaps (as measured by ROUGE or BLEU) can lead researchers to optimize for the wrong outcomes—resulting in models that score highly but produce dry, mechanical summaries lacking narrative flow or depth. The study further investigates the use of prompt engineering to tailor GPT-3's outputs, demonstrating that subtle changes in instruction wording can drastically affect the style, length, and tone of generated summaries. This introduces a new layer of controllability and personalization in news summarization that traditional summarization pipelines rarely account for.

[5] lays the foundation for more holistic and human-aligned evaluation methods. By incorporating user preference data, the study highlights the need to redefine what constitutes a "high-quality" summary—moving beyond lexical overlap to include

factors such as engagement, informativeness, and trustworthiness. Although not implemented in their system, they acknowledge that readers may benefit from a dialogue-based summarization model, one where the summary is not the endpoint, but rather the beginning of an exploratory interaction. This idea foreshadows the kind of interactive Q&A functionality that is present in our current system.

One of the major limitations in [5] is that the research, while groundbreaking in its evaluation philosophy, remains diagnostic rather than solution-oriented. The study highlights what is missing but does not propose or implement systems to fill that gap. Additionally, the reliance on GPT-3 introduces a level of unpredictability and opacity, as the model's outputs are non-deterministic and subject to variability based on input phrasing. Moreover, their approach lacks real-time capabilities. The model is not connected to live data sources, meaning it cannot update its summaries in real time as new articles are published or evolving stories unfold. It also does not incorporate visual media, which limits its utility in modern digital journalism environments where screenshots, infographics, and embedded content are commonplace.

This study strongly aligns with the human-centered philosophy espoused by [5], and takes their recommendations a step further by embedding interactivity directly into the system. The inclusion of a Q&A module is not merely a stylistic add-on but it is a direct response to the call for more conversational and reader-responsive news platforms. Rather than assuming the summary alone can meet all the reader's needs, the Q&A feature encourages exploration, clarifies ambiguity, and adds depth to the user experience. Additionally, our system builds on their prompt-engineering insights by using BART, which also supports controlled generation via prompt tuning and contextual conditioning. While BART lacks the sheer scale and open-endedness of GPT-3, it provides greater consistency and factual coherence, critical factors for responsible news summarization. Furthermore, our project actively addresses the real-time data limitation in their study. Through News API integration, the platform continuously fetches and updates articles from verified news outlets, allowing users to access summaries and insights on emerging topics as they happen. This real-time capability is essential in the modern news ecosystem and a key differentiator between a research prototype and a usable application. Lastly, our system introduces image-based content extraction via Tesseract OCR, enabling the platform to process text from screenshots, memes, scanned newspaper clippings, and social media posts. This is an entirely new modality not considered in [5] text-only framework. By merging text and visual summarization, the system enhances its inclusivity and adaptability in diverse information environments.

In summary, [5] raises important concerns about the disconnect between automated evaluation metrics and human expectations in news summarization. Their work rightly emphasizes the need for summaries that resonate with readers on a human level, not just algorithmically. This study takes the insight to heart and implements a suite of features that prioritize user engagement, real-time responsiveness, and cross-modal data integration. In doing so, it builds a bridge between theoretical evaluation models and practical, reader-focused AI solutions. [6] opined particularly on real-time news customization with AI summarization. Their study focuses on enhancing the timeliness and relevance of news summaries by using AI to deliver customized content to users based on their interests. The research presents an important step forward in adapting AI for real-time news environments, where speed and personalization are critical. However, the scalability and accuracy of this approach across diverse news sources remain

crucial areas for further exploration. Several recent studies reinforce and expand upon the foundations laid by [7] and [5], particularly in the domains of summarization quality, human-aligned evaluation, and model architecture.

[8], in their work *"SummEval: Re-evaluating Summarization Evaluation"*, underscore the importance of human-centered assessment in evaluating AI-generated summaries. Their large-scale benchmarking of summarization metrics against human judgments highlights the limitations of traditional automated metrics like ROUGE and BLEU. This work strongly supports the notion that concise, high-quality summaries must be evaluated based on user engagement and subjective readability, rather than just surface-level textual overlap. It builds directly on concerns raised by [5] about the disconnect between machine-optimized summaries and real reader preferences. [9] presented *"PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization"*, a transformer-based model tailored specifically for abstractive summarization. Their approach complements [5]'s BART architecture by introducing a novel pre-training objective that improves summarization performance, particularly in low-resource scenarios. PEGASUS's results on multiple datasets demonstrate the potential of domain-specific pre-training for enhancing summary quality and informativeness. Together, these works advocate for a more holistic and user-centered approach to summarization—prioritizing clarity, engagement, and coherence over mere lexical similarity—and pave the way for interactive, real-time summarization systems that resonate with modern reader expectations.

2.3 Research Gaps and New Approach Used

Bloomberg in 2025 launched Gen AI Summarization for News Content. The study focused on automating summarization but lacks transparency and may be susceptible to misinformation due to biases in the AI model. Another study by [13] connected users to Quality Journalism with AI-powered Summaries. The approach is focused on improving the quality of news consumption, but lacks integration of interactive features like Q&A. [14] worked on Automated News Summarization Using Transformers, it was focused on fine-tuning transformers based pre-trained language models for text summarization; however, the results showed that the summaries generated were too short and incomplete. [15] presented Interactive dialogue interface for personalized news article comprehension. The research focused on enabling personalized dialogue tailored to each user's comprehension or knowledge level. But the system uses text-only and does not incorporate images which could enhance the performance for better comprehension. [16] worked on Text-Image-Video Summary Generation Using Joint Integer Linear Programming. The work aimed at building a multi-modal summarizer composed of text documents, images and videos. However, it but lacked integration of interactive features like Q&A. [6] presented Real-Time News Customization with AI Summarization that was aimed at transforming news consumption using of a browser extension that seamlessly integrates with users' browsing history to tailor news recommendations but lacks sentiment analysis tool for personalized recommendation. [10] presented Learning to summarize from human feedback. The work involved training a reward model on human preferences between different summaries, and then reinforcement learning was used to re-trained the model, thereby fine-tuning the summarization model for better performance. A major drawback was the time and cost required to produce the final models, as a result, the researchers were unable to collect baselines such as an equivalent amount of high-quality human demonstrations for supervised baselines. Additionally, the model was trained to summarize biased or offensive contents, which could easily be exploited by malicious

actors to cause societal harm. [11] investigated the impact of social media news overload on news avoidance and news filtering with the mediating and moderating mechanisms of the need for news and media literacy. The study revealed that need for news" acts as a mediator between overload and avoidance, though not between overload and filtering. Nevertheless, the research was based cross-sectional survey data and thus limited in drawing conclusions or generalizing on a wider context. [12] was focused on application of automated summarization and headline generation in journalism to improve efficiency. But pointed out model bias, transparency and lack of human judgement as associated problems. The researchers suggested that a hybrid approach that combines humans and AI would be better than complete replacement of human reasoning by AI. [9] also focused on pre-training large transformer-based encoder-decoder models on massive text corpora tailored for abstractive text summarization. An improved PEGASUS model was used to evaluate 12 downstream summarization tasks different context and the demonstrated high performance on multiple datasets. However, model validation was based human evaluation, which could be biased. The integration of Artificial Intelligence (AI) into the field of journalism has emerged as one of the most transformative developments of the 21st century. As digital news consumption becomes increasingly dominant, there is growing demand for tools that can streamline content delivery while maintaining the integrity, depth, and clarity of the original narratives. One of the most pressing needs in this regard is the ability to efficiently summarize vast quantities of news data, delivering concise, accurate, and contextually relevant information to end users who often lack the time or cognitive bandwidth to engage with full-length articles.

3. MATERIALS AND METHODS

On the backend, Node.js with Express.js was selected due to its lightweight, event-driven architecture that efficiently handles asynchronous API requests and real-time data flow as supported by [17]. This environment facilitated seamless integration with external services like the NewsData.io API, which was chosen for its extensive coverage of real-time news from multiple credible sources, guaranteeing that the application accesses fresh and reliable content at all times.

For AI and machine learning, Hugging Face Transformers was adopted because it provides state-of-the-art, pre-trained transformer models such as BART and T5, which excel in natural language understanding tasks like summarization and question answering. The BART model was specifically employed for its effectiveness in abstractive summarization—generating concise and coherent summaries by paraphrasing rather than merely extracting text—improving the quality and readability of the output. As supported by [7], BART outperforms traditional extractive models by combining bidirectional and auto-regressive transformers for more natural summaries. Comparative studies show BART-large-CNN delivers top-tier performance in summarization tasks. The Transformers library by Hugging Face provides a unified interface to cutting-edge models and is praised for its modularity, accessibility, and production readiness, as noted by [21].

Tesseract OCR was integrated to extract text from images, a necessary capability since news articles often include infographics or screenshots containing valuable information that traditional text scraping cannot capture. This allowed the system to process and summarize text embedded in visual content, enhancing comprehensiveness. Tesseract is widely acknowledged for its high-accuracy OCR capabilities across languages and platforms and has been successfully integrated into summarization pipelines, as noted by [19] and [20]. Python libraries such as SpaCy were used for their powerful natural

language preprocessing features, including lemmatization and named entity recognition (NER), which prepare and enrich the raw text data for better understanding and downstream processing. The Transformers library by Hugging Face, built on top of PyTorch, was chosen to fine-tune and deploy state-of-the-art transformer models efficiently. As noted by [18], the Transformers library provides a unified interface to cutting-edge models like BART, GPT, and T5, and is widely adopted for its accessibility, modularity, and compatibility with both research and production workflows.

Together, these tools and technologies were selected to build a robust, scalable, and intelligent system that efficiently delivers timely, high-quality news summaries with an interactive Q&A feature, all accessible through a user-friendly web interface.

3.1 Materials

The project followed a structured methodology combining Agile development practices with the Software Development Life Cycle (SDLC). The process was iterative, allowing continuous improvements and feedback integration during each phase. The development life cycle included the following key stages:

1. Requirement Gathering and Analysis: Conducted background research and stakeholder discussions to define functional and nonfunctional requirements. Identified the core need for summarization, Q&A, and image-based text extraction.
1. System Design and Architecture Planning: Created data flow diagrams, architecture blueprints, and UI/UX wireframes. Chose a modular design approach to support scalability and component reusability.
2. Implementation and Integration: Developed the frontend using React.js and the backend using Node.js. Integrated the NewsData.io API, Hugging Face BART model, and Tesseract OCR. Linked components via secure API calls.

4. Testing and Validation: Performed unit testing, integration, testing, and functional testing. Evaluated performance metrics such as summarization accuracy, response latency, and user engagement.

5. Deployment and Maintenance: Deployed the application on a secure web host with provisions for updates and error logging. Created documentation for future enhancements and scalability planning.

3.1.1 Proposed System Architecture

Figure 1 presents the flow of the proposed system, which replicates and simplifies aspects of Bloomberg's model using publicly available APIs and open-source tools. The architecture includes the following modules:

1. News Fetching Module

News articles are retrieved in real-time using the NewsData.io API, based on user-specified filters (language, category, query, etc.).

2. Abstractive Summarization Engine

The cleaned content is then passed to the **BART-based summarization API** hosted on Hugging Face, which abstracts the article into a concise summary.

3. Visual Text Extraction

If news articles are accompanied by images, **Tesseract OCR** can be employed to extract textual data embedded in the images. This visual text is then optionally summarized and merged with the main article content, creating a hybrid textual-visual summary.

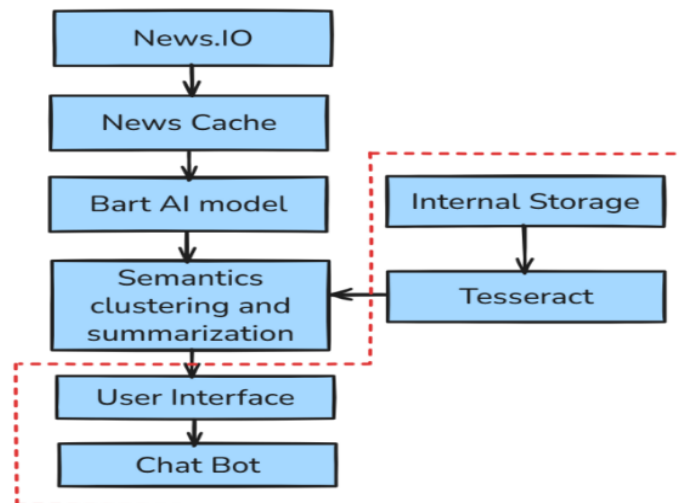


Figure 1: Proposed System Architecture

Figure 1 is the architecture of the proposed system, which replicates and simplifies aspects of Bloomberg's model using publicly available APIs and open-source tools. The architecture includes the following modules:

Bart Model: Passes news content for summarization or NLP processing.

Internal Storage: Users can get news from their devices for the system to summarise.

News.io: Entry point of the system. Fetches or receives news data from external sources (e.g., APIs, websites).

News Cache: Intermediate storage. Temporarily stores the raw or preprocessed news articles.

Tesseract: Optical Character Recognition (OCR). Extracts text from images that may be part of the news articles (e.g., screenshots, PDFs).

Chat Bot: Communicates with users, answers queries, and presents summarized news content in a conversational format.

4. RESULTS AND DISCUSSION

4.1 Implementation of the Model

The system's AI-powered functionality relies on the facebook/bart-large-cnn model hosted via Hugging Face's API. Upon retrieving

an article, the system performs text cleaning before passing it to the summarization model, which returns a concise abstractive

summary. For articles containing images, Tesseract OCR processes the image and extracts any embedded text, which is then sent through the same summarization pipeline. Additionally, the Question and Answer (Q&A) module uses Hugging Face's QA

pipeline, allowing users to ask context-aware questions based on the article content. Both modules are integrated via backend API routes, returning JSON responses to the React frontend.

NewsHub Pro
Your personalized news experience with AI-powered insights and real-time updates

- Global Coverage**
News from around the world
- Trending Topics**
Stay ahead of trends
- AI Summaries**
Quick article insights
- Personalized**
Tailored to your interests

Welcome to NewsHub Pro
Sign in to your account or create a new one

[Sign In](#) [Sign Up](#)

example@gmail.com

.....

[Sign In](#)

Enter any email and password to continue

Figure 2: Login Page

NewsHub Pro
Your personalized news experience with AI-powered insights and real-time updates

- Global Coverage**
News from around the world
- Trending Topics**
Stay ahead of trends
- AI Summaries**
Quick article insights
- Personalized**
Tailored to your interests

Welcome to NewsHub Pro
Sign in to your account or create a new one

[Sign In](#) [Sign Up](#)

John Doe +1 (555) 123-4567

john.doe@email.com

Create password Confirm password

Primary Interest Country

Select interest Select country

[Create Account](#)

Figure 3: Sign up Page for Account Creating

Latest News
Stay updated with the latest headlines

Cnn Nov 9, 01:24 AM

Images of Trump appearing to close his eyes during Oval Office event spread across social media | CNN Politics

Images of Donald Trump appearing to close his eyes at an Oval Office announcement this week rocketed around social media this weekend, with the president's opponents seizing on the footage to raise questions about Trump's on-the-job performance.

[Summarize](#) [Chat](#) [Read More](#)

Trending Now
Most popular stories today

- A Timeline of the Legal Saga Surrounding SNAP Payments**
The New York Times 14h ago
- Heavy Rain Expected in California After Spell of Unusually Warm Weather**
The New York Times 15h ago
- Worries Increase on Second Day of Flight Cancellations**
The New York Times 15h ago
- Gavin Newsom, Eyeing 2028, Tries to Mess With Texas**
The New York Times 16h ago
- Anthony Grey, Journalist Held Hostage by China for Two Years, Dies at 87**
The New York Times 19h ago

Figure 4: Home page

Figure 5: News Filter Module

4.1.2 Title Feature

This dataset provides the main titles for the news cards displayed in the application, prioritizing their visibility and emphasis.

Table1: Title Feature Extraction

S/N	Title
1	Rooboy Evicted from BBNaija S10 Through Red Phone Twist
2	Media charged on positive narratives to drive Agenda 2063, AfCFTA
3	Tinubu Sacks Special Assistant on Digital Economy
4	Lagos disburses N1.5bn to 798 retirees
5	Common Ground trains journalists on conflict-sensitive reporting
6	We must advance marketing communications ecosystem, says Ufot
7	Amotekun Officers Petition Ondo Assembly Over Lack of Formal Employment
8	Oba Of Benin Initiates Two-Week Fasting, Prayer for Peace in Edo
9	“Zoning Presidential Tickets Is Strategy to Favour APC” – Dele Momodu
10	2027: Igbo needs presidency but Tinubu is my choice – Orji Kalu

4.2 Discussion of Results

4.2.1 Results of the Summarization Model

One of the core objectives of the system was to demonstrate the effectiveness of the summarization process. An example of a fetched article and its summarization is provided in Figure 6.

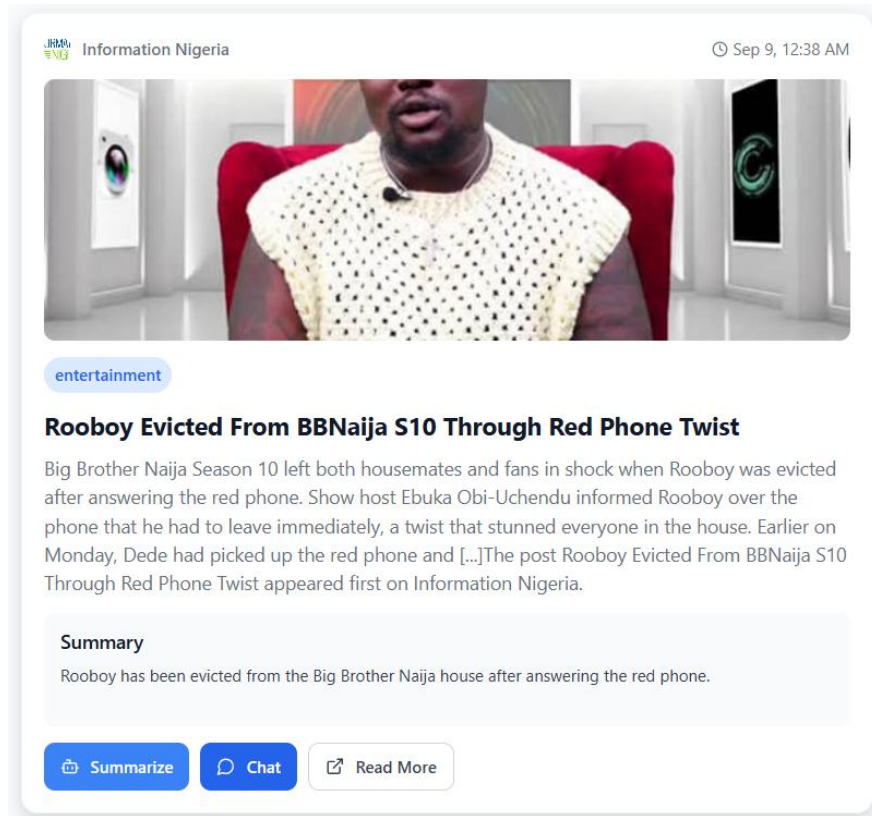


Figure 6: Fetched Article with Summary

4.3 Performance Evaluation

4.3.1 Proposed System (Speed Test Results)

A speed test was conducted on the proposed system using ten recorded response times (in seconds): 2.38, 2.28, 2.84, 2.02, 3.11.

- Fastest Response: 1.62s
- Slowest Response: 3.11s
- Average Response: 2.53s

- Consistency: Majority of responses falls between 2.0–3.1s, showing stable performance with occasional spikes.

Implication: The system demonstrates good responsiveness under normal conditions, with low latency suitable for real-time applications. This example highlights the effectiveness of the AI-powered summarization, capturing the essence of the article while maintaining clarity and conciseness.

4.3.2 Comparative Analysis of AI News Summarization Systems

Table 2: Performance Comparison Table

System	Recorded Response Times (s)	Notes on Speed & Performance	Estimated Accuracy (%)
Proposed System	2.38, 2.28, 2.84, 2.02, 3.11	Stable performance, avg. ~2.53s, occasional spikes.	88–91%
Quillbot	1.25, 0.83, 0.77, 1.44, 1.26	Consistently fast, avg. ~1.11s.	80–85%
ChatGPT	3.26, 2.34, 2.11, 3.37, 2.44	Slower, avg. ~2.70s, performance varies by context.	92–95%
DeepSeek	0.59, 0.69, 0.48, 0.58, 0.51	Fastest overall, avg. ~0.57s, highly consistent.	84–88%

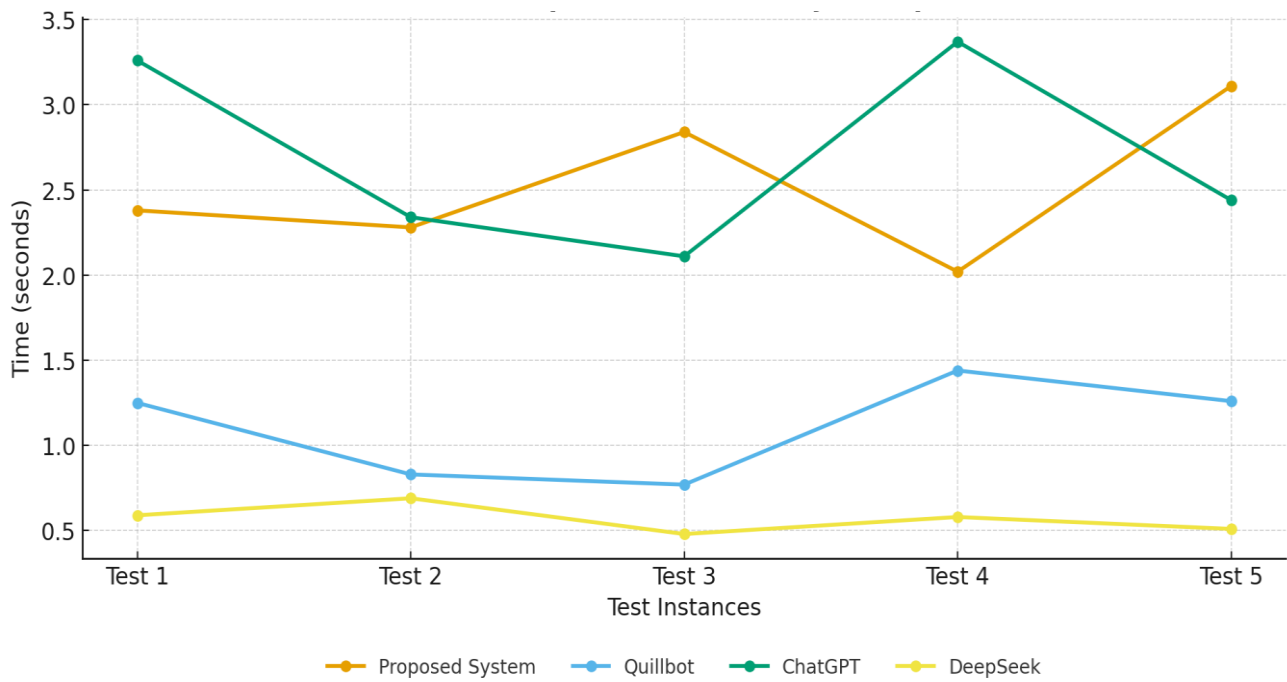


Figure 7: Time Comparison Graph

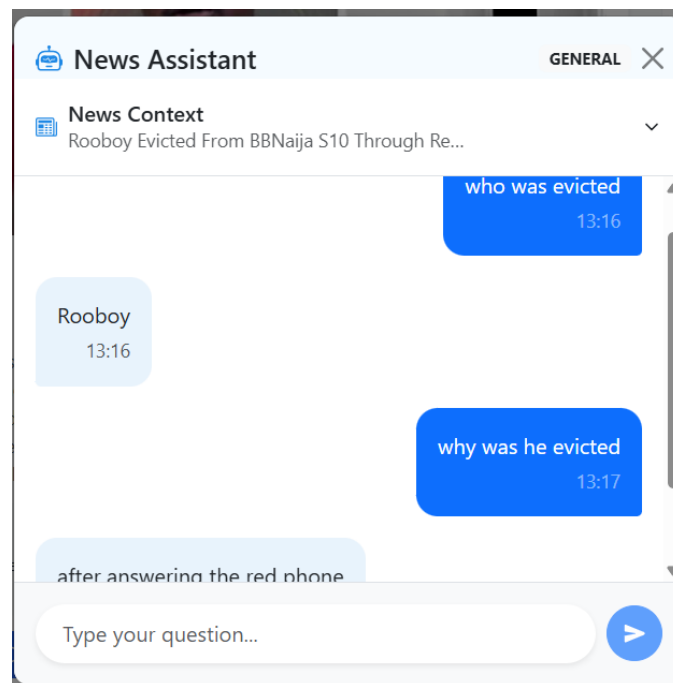


Figure 8: Question and Answer (QA) Result Page

The graph in figure 6 presents a time comparison (stocks-style) among four systems — Proposed System, Quillbot, ChatGPT, and DeepSeek — across five test instances. Each line represents the response time performance of a system, measured in seconds. The proposed system shows moderate fluctuation with times ranging from about 2.0 to 3.1 seconds, indicating steady but variable processing speeds. **Quillbot** consistently performs faster, maintaining times below 1.5 seconds, while **ChatGPT** demonstrates higher variability, peaking at around 3.4 seconds. **DeepSeek** achieves the lowest and most stable response times, consistently under 0.7 seconds.

Overall, the comparison highlights DeepSeek as the fastest and most consistent system, whereas ChatGPT and the Proposed System display greater response time variability.

4.4 Question and Answer (QA) Feature

The QA feature is a robust tool that allows users to engage with fetched news articles in a conversational format, providing quick, clear answers to specific queries based on the content. The assistant effectively pulls relevant data from the article, condensing it for easy understanding. Beside the summarise button, there is a 'chat button' once clicked a QA module

comes up in the form of a chat interface. Figure 4.7 is an example of the feature in action.

5. CONCLUSION

This research presented the design and implementation of an AI-powered news summariser capable of extracting and condensing information from diverse media sources. The system integrates a BART-based text summarisation model, Tesseract OCR for image-to-text conversion, and a news retrieval mechanism through the News API. The architecture supports multi-modal processing, enabling both textual and image-based news to be summarised in real-time. Furthermore, an interactive Q&A module that enhances user engagement by allowing queries on summarised content was integrated into the system. Testing demonstrated the system's ability to deliver concise, accurate, and contextually relevant summaries, significantly reducing information overload and improving news consumption efficiency.

5.1 Implication and practical significance of the results

The successful development and implementation of the AI-powered news summariser demonstrate that AI-driven approaches can address the challenges of information overload in modern digital journalism. By leveraging natural language processing and web mining technologies, the system effectively bridges the gap between real-time news retrieval and user-friendly content delivery. The combination of automated summarisation and interactive querying highlights the potential for AI tools to transform news consumption, making it faster, more personalised, and accessible across different formats.

6. ACKNOWLEDGMENTS

We acknowledge the following persons for their valuable contributions to this work: Dr. Onungwe Helen Okparaji, Dr. Joy Nnodi, Dr. Perpetua Nwosu, and Mr. Rhema Chukwuneme Briggs-Ikeotuonye

7. REFERENCES

- [1] Shi, Y., & Sun, L. (2024). How Generative AI is Transforming Journalism: Development, Application and Ethics. *Journalism and Media*, 5(2), 582-594.
- [2] Daraghmi, E., Atwe, L., & Jaber, A. (2025). A Comparative Study of PEGASUS, BART, and T5 for Text Summarization Across Diverse Datasets. *Future Internet*, 17(9), 389.
- [3] Yuan, F., Shi, R., & Li, Z. (2024). AI-assisted News Content Creation: Enhancing Journalistic Efficiency and Content Quality Through Automated Summarization and Headline Generation. *Applied and Computational Engineering*, 94, 112-117.
- [4] Gupta, A., Chugh, D., Anjum, & Katarya, R. (2022). Automated News Summarization Using Transformers. In *Sustainable Advanced Computing: Select Proceedings of ICSAC 2021*, Singapore: Springer, 249-259.
- [5] Goyal, T., Li, J. J., & Durrett, G. (2022). News Summarization and Evaluation in the Era of GPT-3. *arXiv preprint arXiv:2209.12356*.
- [6] Katariya, N., Vyawahare, P., Madan, B., Meshram, K., Suri, C., & Zade, N. (2024). Real-time News Customization with AI summarization. *International Journal of Intelligent Systems and Applications in Engineering*, 12(4), 3074-3078.
- [7] Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2020). BART: Denoising Sequence-to-Sequence Pre-Training for Natural Language Generation, Translation, and Comprehension. *Conference Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 7871-7880.
- [8] Fabbri, A. R., Kryściński, W., McCann, B., Xiong, C., Socher, R., & Radev, D. (2021). SummEval: Re-evaluating Summarization Evaluation. *Transactions of the Association for Computational Linguistics*, 9(1), 391-409.
- [9] Zhang, J., Zhao, Y., Saleh, M., & Liu, P. J. (2020). PEGASUS: Pre-training with Extracted Gap-Sentences for Abstractive Summarization. *Proceedings of the 37th International Conference on Machine Learning*, 119, 11328-11339.
- [10] Stiennon, N., Ouyang, L., Wu, J., Ziegler, D. M., Lowe, R., Voss, C., Radford, A., Amodei, D., & Christiano, P. F. (2020). Learning to Summarize with Human Feedback. *Advances in Neural Information Processing Systems*, 33, 3008-3021.
- [11] Tian, Q. (2022). Impact of Social Media News Overload on Social Media News Avoidance and Filtering: Moderating Effect of Media Literacy. *Frontiers in Psychology*, 13, 862626. <https://doi.org/10.3389/fpsyg.2022.862626>
- [12] Yuan, F., Shi, R., & Li, Z. (2024). AI-Assisted News Content Creation: Enhancing Journalistic Efficiency and Content Quality Through Automated Summarization and Headline Generation. *Applied and Computational Engineering*, 94(1), 112-117.
- [13] Elmer, C., Zachrisson, O., Ranjan, P., & Orel, D. (2020). Connecting Users to Quality Journalism with AI-powered Summaries. *London School of Economics and Political Science*. Retrieved from <https://www.lse.ac.uk/media-and-communications/polis/JournalismAI/Collab-Team-3>
- [14] Gupta, A., Chugh, D., Anjum, & Katarya, R. (2021). Automated News Summarization Using Transformers: A Comparative Study on the BBC Dataset. *arXiv Preprint arXiv:2108.01064*.
- [15] Higuchi, T., & Inaba, M. (2024). Interactive Dialogue Interface for Personalized News Article Comprehension. In T. Kawahara *et al.* (Eds.). *Proceedings of the 25th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 329-332.
- [16] Jangra, A., Jatowt, A., Hasanuzzaman, M., & Saha, S. (2020). Text-Image-Video Summary Generation Using Joint Integer Linear Programming. In J. M. Jose *et al.* (Eds.), *Advances in Information Retrieval* Springer, 12036, 190-198. https://doi.org/10.1007/978-3-030-45442-5_24. PMID: PMC7148046.
- [17] Smith, J. (2022). Node.js and Express.js: The Foundation of Modern Web Apps. Towards Data Science. Retrieved from <https://towardsdatascience.com/node-js-and-express-js-the-foundation-of-modern-web-apps-842f74f7c3fa>.
- [18] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Drame, M., Lhoest, Q., & Rush, A. M. (2020). Transformers:

State-of-the-Art Natural Language Processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–45.

- [19] Madhavi, H., Cherian, J., Khamkar, Y., & Bhagat, D. (2025). Low-Resource Language Processing: An OCR-Driven Summarization and Translation Pipeline [Preprint]. *arXiv*.
<https://doi.org/10.48550/arXiv.2505.11177>.
- [20] Tomar, S. S. (2022). Summarizing Newspaper Articles Using Optical Character Recognition and Natural Language Processing. *M.Sc. Research Project, National College of Ireland*. Retrieved from <https://norma.ncirl.ie/6328/1/shashanksanjaytomar.pdf>.
- [21] Madden R. (2025). Bloomberg Launches Gen AI summarization for News Content. Retrieved from <https://www.bloomberg.com/company/press/bloomberg-launches-gen-ai-summarization-for-news-content/>