

---

# "SentiNEWS: Bilingual Summarization and Sentiment Analysis with SerpAPI for Efficient and Accessible News Consumption"

---

**Prolay Shankar Mazumder**  
prolay23065@iiitd.ac.in  
IIIT Delhi

**Shaksham Singhal**  
shaksham23087@iiitd.ac.in  
IIIT Delhi

**Shivani Sharma**  
shivani23142@iiitd.ac.in  
IIIT Delhi

**Argharupa Adhikary**  
argharupa23020@iiitd.ac.in  
IIIT Delhi

**Arunoday Ghorai**  
arunoday23023@iiitd.ac.in  
IIIT Delhi

**Sandip Pal**  
sandip23080@iiitd.ac.in  
IIIT Delhi

## ABSTRACT

In an era inundated with information, efficient and insightful news consumption is paramount. SentiNEWS emerges as a pioneering solution, redefining the landscape of news consumption through its innovative approach to bilingual summarization and sentiment analysis. This platform integrates advanced text processing techniques, including NLTK and Newspaper Library Integration, to sift through vast amounts of web content, utilizing Google's own information retrieval engine via SerpAPI to curate the most relevant articles. The application employs a sophisticated summarization technique, coupled with TF-IDF analysis, to distill articles into concise summaries while preserving essential context. Additionally, SentiNEWS dynamically generates relevant images using Hugging Face's Standard Diffusion Model, enhancing user engagement and comprehension by visually representing key article themes. Furthermore, SentiNEWS optimizes performance through GPU acceleration and a pre-trained image generation model, ensuring swift and visually captivating results. Moreover, SentiNEWS goes beyond traditional news consumption methods by incorporating text-to-speech conversion functionality. This feature enables users, including those with visual impairments, to access news content effortlessly through audio summaries of articles. By seamlessly integrating sentiment analysis, keyword-based image generation,

text to speech conversion and refined summarization processes, SentiNEWS offers users a comprehensive and user-friendly tool for navigating today's dynamic information landscape.

## KEYWORDS

SentiNEWS, Sentiment Analysis, Bilingual, SerpAPI, NLTK, TF-IDF, Hugging Face, CUDA

## 1 PROBLEM STATEMENT

In today's world, finding news that's right for you is hard because there's so much of it. Existing platforms often lack efficient summarization in more than one language, sentiment analysis, and visually engaging representations. Thus, there is a need for a comprehensive application which can integrate advanced technologies to streamline news consumption, ensuring users access timely, summarized, and sentiment-analyzed news with visually captivating representations.

## 2 INTRODUCTION

In today's fast-paced digital era, the abundance of news sources combined with time constraints poses significant challenges to comprehensive news consumption. Moreover, accessibility issues persist, especially for individuals with disabilities, finding it difficult in quickly analyzing

sentiment amidst the sheer volume of content. In response to these challenges, the development of SentiNEWS emerges as a crucial solution. By leveraging state-of-the-art techniques in text processing, summarization, sentiment analysis, text to speech conversion and dynamic image generation, SentiNEWS aims to transform the landscape of news consumption, providing users with efficient, accessible, and insightful means of staying informed. Navigating the complexities of information overload, sentiment analysis accuracy, and accessibility, SentiNEWS addresses fundamental challenges inherent in modern news consumption. Algorithmic biases, scalability concerns, and the complexity of multimodal integration are among the key challenges the platform tackles. By effectively addressing these challenges, SentiNEWS stands poised to revolutionize news consumption methods, benefiting a diverse array of users, including those seeking efficient news consumption methods, individuals with disabilities requiring accessible news, and individuals lacking time for extensive reading yet still desiring to remain informed. SentiNEWS not only addresses pressing challenges but also introduces novel approaches to news consumption. Furthermore, SentiNEWS integrates sentiment analysis seamlessly, offering deeper insights into emotional tone and subjective opinions, thus enhancing the overall user experience. By synthesizing cutting-edge technologies and innovative approaches, SentiNEWS represents a significant advancement in news consumption methods. Through its comprehensive suite of features, including text processing and summarization, sentiment analysis, dynamic image generation, and crucial keyword extraction, SentiNEWS empowers users to navigate the dynamic information landscape effectively. As such, the development of SentiNEWS holds immense promise in revolutionizing the way individuals consume news, offering a user-friendly and impactful solution for staying informed in today's fast-paced world.

### 3 MOTIVATION

We are driven to create SentiNEWS because we see how tough it can be for people to keep up with the news nowadays. There's just so much information out there, and it's hard to find the time to read it all. Also, not everyone can access news easily, especially those with disabilities. We believe there is a better way to do things. By using smart technology like natural language processing and sentiment analysis, we can make news easier to understand and more accessible to everyone. Our goal with SentiNEWS is to help people stay informed and engaged with the world around them, no matter how busy life gets.

## 4 LITERATURE REVIEW

In 2017, Tarun B. Mirani and Sreela Sasi from Gannon University, in their paper: Two-level Text Summarization from Online News Sources with Sentiment Analysis [4], employed a two-level text summarization approach with sentiment analysis to extract important sentences from online news articles covering various topics such as politics, sports, health, science, and movie reviews. The study utilized an extractive summarization method to generate summaries from two to three news articles for each topic. The extraction-based approach involved identifying important sentences and arranging them in order of their importance. Sentiment analysis was then performed to understand variations in news articles from different sources, determining positive, negative, or neutral opinions. Additionally, they used the ROUGE metric to evaluate the performance of the summarization. The study concluded that the positive polarity in the text processing indicated a public inclination towards government policy. The research utilized Python language and various packages for tasks such as fetching URLs, pre-processing news articles, and evaluating the performance of the summarization using the ROUGE metric. Overall, they presented a novel approach to text summarization and sentiment analysis, contributing to the advancement of automatic text summarization and sentiment analysis from online news sources.

In 2013, Alexandra Balahur, Ralf Steinberger, Mijail Kabadjov, Vanni Zavarella, Erik van der Goot, Matina Halkia, and Bruno Pouliquen, in their paper: Sentiment Analysis in the News, focused on sentiment analysis in news articles [2], particularly in English language news. They identified the differences between sentiment analysis in highly subjective text types like product reviews and news articles, emphasizing the need to clearly define the target of any sentiment expressed and to restrict the analysis to the immediate context of the target. The authors conducted experiments to test the relative suitability of various sentiment dictionaries and attempted to separate positive or negative opinion from good or bad news. They also discussed the impact of category-defining word lists on sentiment analysis and the challenge of interpreting sentiment that is not expressed lexically in news texts. The authors outlined their experiments and evaluations, including the impact of using different lexicons and word windows, as well as the errors encountered in sentiment classification, such as misclassifying sentences as neutral due to lack of explicit sentiment words, misclassifying sentiment due to irony, and problems related to co-reference. The paper concluded by outlining future work,

including evaluating the impact of using negation and valence shifters, extending the lexica for additional languages, and assessing methods to compare opinion trends across sources and time.

In 2009, Mijail Kabadjov, Josef Steinberger, Bruno Pouliquen, Ralf Steinberger, and Massimo Poesio presented a paper on "Multilingual Statistical News Summarisation: Preliminary Experiments with English." [1]. The paper aimed to develop a generic approach for summarizing multilingual news clusters, particularly those produced by the Europe Media Monitor (EMM) system. The authors highlighted the challenges of developing a multilingual news summarizer within the EMM system and emphasized the need for multilingual capabilities, which have been less addressed in the text summarization literature. The paper introduced the use of Latent Semantic Analysis (LSA) and multilingual entity disambiguation to enhance the summarization of news clusters, aiming to provide succinct and comprehensive summaries for decision-makers within the European Union who utilize the EMM system. The authors discussed various approaches to text summarization, including shallow linguistic analysis, machine learning, and more sophisticated approaches, emphasizing the multi-faceted nature of text summarization research. They also presented preliminary experiments with the TAC 2008 data, demonstrating promising improvements over a summarization system ranked in the top 20 percent at the TAC 2008 competition. The paper concluded by outlining future steps, including incorporating intra-document co-reference resolution and expanding the range of normalized entity information to enhance the summarization process.

In 2015, Ubale Swati, Chilekar Pranali, and Sonkamble Pragati conducted a study on sentiment analysis of news articles using a machine learning approach [3]. The study aimed to determine the overall sentiment expressed in news articles. The authors highlighted the importance of sentiment analysis in tracking a company's behavior over time and its application in social media monitoring to gauge consumer sentiment towards a brand. The study also delved into the challenges of fine-grained sentiment analysis, particularly in the context of news articles, and the techniques and algorithms used to improve accuracy. The authors discussed the framework and techniques employed for sentiment analysis, including crawling and extraction, data preprocessing, feature extraction, sentiment identification, scoring, and classifier training. They also referenced relevant works and algorithms, positioning sentiment analysis as a valuable tool for extracting insights from news articles and social media

to aid decision-making processes for businesses and organizations. They also referenced works that addressed semantic parsing and fine-grained sentiment analysis, highlighting the challenges posed by the variety of ways in which opinions can be expressed in news articles. Overall, the study provided a comprehensive overview of sentiment analysis of news articles using a machine learning approach, emphasizing its significance in various domains and the reliability and efficiency of the proposed methodology.

## 5 METHODOLOGY

### 5.1 SerpApi

SerpAPI plays a pivotal role in sourcing the best news pages available on the internet based on user queries. Leveraging the power of the Google News search engine, SerpAPI swiftly generates relevant and high-quality news articles based on the user's input. When user inputs a query, through SerpApi, it generates best news related to user's input query.

We imported the Serpapi module, which is a Python client for the SerpApi service. By utilizing SerpAPI, SentiNEWS enhances the efficiency and accuracy of news retrieval, empowering users to access timely information on a diverse range of topics.

### 5.2 Keywords extraction

The keyword extraction process, facilitated by the `extract_keywords` function, analyzes a given text to identify the top 5 keywords using TF-IDF (Term Frequency-Inverse Document Frequency) vectorization. Initially, the function computes the TF-IDF matrix for the input text, identifies feature names (words), computes TF-IDF scores, sorts the words by their scores, and finally selects the top 5 keywords with the highest scores. The `extract_keywords(text)` method efficiently extracts crucial keywords from summarized news articles, offering valuable insights into the primary topics and themes covered.

### 5.3 Sentiment Analysis

The sentiment analysis process, executed by the `analyze_sentiment` function, evaluates the sentiment of a given text using the VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment analyzer. The function calculates sentiment scores for the input text and classifies it as "Positive", "Negative", or "Neutral" based on the compound sentiment score. The methodology incorporates NLTK's `SentimentIntensityAnalyzer` class from the sentiment module to analyze pre-processed text data.

#### 5.4 Text Processing and Summarization

The text processing and summarization methodology encompasses several key steps. At its core, the `search_and_summarize_news()` function drives the main functionality by prompting users to input news queries and subsequently searching for relevant articles using the SerpApi client. Summarization involves extracting titles, summaries, top keywords, and sentiment analysis from each top link, aided by functions like `remove_stopwords`, `extract_keywords`, and `analyze_sentiment`. Further refinement involves removing stopwords and alphanumeric characters to enhance coherence and readability, while keyword extraction ensures focused summaries. Leveraging NLTK for preprocessing tasks and integrating the newspaper library streamline text extraction and summarization, while TF-IDF analysis improves summary quality by extracting significant terms. Together, these processes provide users with concise, informative summaries and facilitate efficient news consumption.

#### 5.5 Text to Speech Conversion

The text-to-speech (TTS) conversion process begins by generating audio using the gTTS (Google Text-to-Speech) library. The text to be converted and the desired language are passed as parameters to the gTTS function, which generates the audio file. This audio file is then stored in memory using a BytesIO object to facilitate playback. Subsequently, the audio is played using the pygame library, which initializes the audio mixer and loads the audio file for playback. While the audio is playing, the program continuously checks if the audio playback is complete, ensuring a smooth and uninterrupted listening experience. Additionally, a `play_button` function is defined to create interactive buttons for playing and stopping the audio, enabling user control over the audio playback.

#### 5.6 Dynamic Image Generation

Dynamic image generation enhances the visual representation and engagement of news content through the utilization of Hugging Face's Standard Diffusion Model. The `generate_image(prompt)` function drives this process by taking a text prompt as input and generating an image based on it using the `StableDiffusionPipeline` class from the `diffusers` module. This function initializes a `StableDiffusionPipeline` object, sends the prompt to the pipeline, and generates an image based on the prompt. Leveraging CUDA for GPU acceleration and a pre-trained model ensures optimized performance

and visually captivating results. The integration of the Diffusers library from HuggingFace facilitates the utilization of state-of-the-art diffusion models for image generation, while the Diffusion Pipeline simplifies the complexity of the diffusion system into an easy-to-use API.

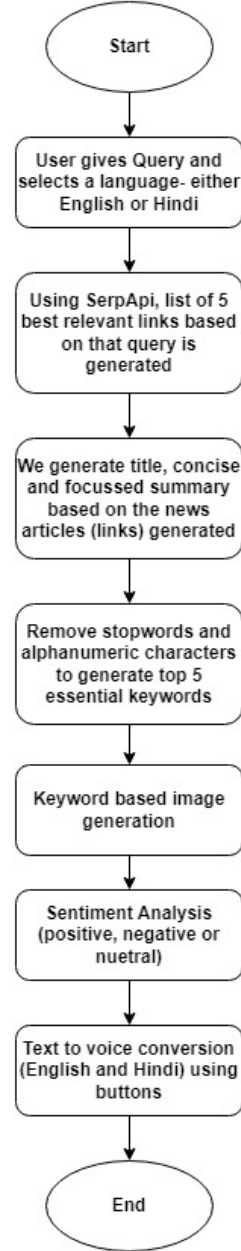


Figure 1: **Workflow**

#### 5.7 User Interface

In UI part, upon entering a query, users are prompted to select their preferred language - English or Hindi. After selecting a language, the system retrieves the five most relevant news links based on the query, accompanied by concise and

focused summaries, top five keywords, and sentiment analysis. Additionally, an image is dynamically generated based on the extracted keywords. The interface features three buttons: one for initiating text-to-speech conversion, another for stopping audio playback, and a third for displaying ROUGE scores. Users are also presented with the option to initiate a news search or exit the interface. This intuitive and interactive design ensures efficient news consumption and enhances user engagement.

Figure 2: Home page with User query

Figure 3: Select Language option

Figure 4: Search Again option

Print ROUGE Scores		
Rouge-1	Rouge-2	Rouge-l
{r: 0.17258883248730963, p: 1.0, f: 0.29437229186184666}	{r: 0.13533834586466165, p: 0.972972972972973, f: 0.2376237602322131}	{r: 0.17258883248730963, p: 1.0, f: 0.29437229186184666}

Figure 5: Rogue Score calculation

Rouge-1	Rouge-2	Rouge-l
{r: 0.014139590854392299, p: 1.0, f: 0.027884900347999387}	{r: 0.006769462203836028, p: 0.9642857142857143, f: 0.013444541128819595}	{r: 0.014139590854392299, p: 1.0, f: 0.027884900347999387}

Figure 6: Results for the input query in English

Rouge-1	Rouge-2	Rouge-l
{r: 0.42152466367713004, p: 0.8894736842105263, f: 0.5911949643635538}	{r: 0.3206997084548105, p: 0.9565217391304348, f: 0.4803493412172728}	{r: 0.42152466367713004, p: 0.8894736842105263, f: 0.5911949643635538}

Figure 7: Results for the input query in Hindi

## 6 EVALUATION

### 6.1 Performance Metrics

Performance metric that we used is the ROUGE score, which measures the similarity between the system-generated summary and a reference summary. For our evaluation, we focus on the example query "farmer protest" chosen for its topical relevance and potential diversity of results.

Using the Serp Api, we retrieve five pertinent links corresponding to the "farmer protest" query. These links serve as our primary data source for evaluating the summarization system. Before calculating ROUGE scores, we preprocess the retrieved articles to ensure uniformity and consistency in the evaluation process. ROUGE scores are calculated for each link, encompassing ROUGE-1, ROUGE-2, and ROUGE-L metrics.

Below table includes columns for link URL, ROUGE-1, ROUGE-2, and ROUGE-L scores (rounded to 2 decimal points), facilitating easy comparison and analysis of the summarization performance across different links

Website Link	ROUGE Scores		
	ROUGE-1	ROUGE-2	ROUGE-L
Link 1	Recall: 0.23 Precision: 0.93 F1 Score: 0.38	Recall: 0.06 Precision: 0.27 F1 Score: 0.10	Recall: 0.23 Precision: 0.93 F1 Score: 0.38
Link 2	Recall: 0.33 Precision: 0.84 F1 Score: 0.47	Recall: 0.12 Precision: 0.36 F1 Score: 0.18	Recall: 0.33 Precision: 0.84 F1 Score: 0.47
Link 3	Recall: 0.5 Precision: 0.77 F1 Score: 0.60	Recall: 0.13 Precision: 0.26 F1 Score: 0.17	Recall: 0.49 Precision: 0.76 F1 Score: 0.60
Link 4	Recall: 0.23 Precision: 0.88 F1 Score: 0.37	Recall: 0.05 Precision: 0.21 F1 Score: 0.08	Recall: 0.23 Precision: 0.88 F1 Score: 0.37
Link 5	Recall: 0.64 Precision: 0.80 F1 Score: 0.71	Recall: 0.30 Precision: 0.41 F1 Score: 0.34	Recall: 0.64 Precision: 0.80 F1 Score: 0.71

Figure 8: Table containing ROUGE scores for the retrieved links related to the "farmer protest" query

Upon analyzing the obtained ROUGE scores, we discern trends and patterns that shed light on the system's efficacy. Variations in summarization quality among different links offer valuable insights into the system's strengths and weaknesses, informing potential areas for improvement.

## 6.2 Sentiment Analysis

Based on the sentiment analysis results for the query on the "farmer protest", there were 2 positive sentiment, 3 negative sentiment and no neutral sentiment. It can be inferred that there is a predominance of negative sentiment regarding this topic. The presence of 3 negative sentiments suggests that there are dissenting or unfavorable opinions expressed. On the other hand, the presence of 2 positive sentiments indicates that there is significant support or favorable opinions expressed in the articles related to the farmer protest. The absence of any neutral sentiments suggests that the articles tend to express clear opinions rather than remaining neutral on the issue. Overall, this analysis indicates a mixed sentiment landscape surrounding the farmer protest topic, with a slightly higher prevalence of negative sentiment.

## 7 NOVELTY

Our work is different from others in the sense that:

- **Text-to-Speech Summaries:** We have introduced a feature that converts news summaries into speech format that caters to visually challenged individuals, ensuring accessibility to news content. Currently there is option for 2 languages-English and Hindi. This new feature helps more people access information by giving them a different way to understand it, not just through reading.

- **Keyword-Based Image Crafting:** By generating images based on essential keywords extracted from news articles, this feature enhances engagement and comprehension for all users. Visual representations of key concepts or events offer an additional layer of understanding and appeal to diverse learning preferences.

- **Bi-lingual News Generation:** By offering news content in both Hindi and English broadens the platform's accessibility and acceptability, catering to a wider audience base. This bilingual approach accommodates users more comfortable with either language, thereby expanding the platform's reach and impact.

- **Seamless Integration of Advanced Models and APIs:** The seamless fusion of advanced models and external APIs enhances the platform's capabilities and performance. By leveraging state-of-the-art technologies, users benefit from accurate summarization, sentiment analysis, and content retrieval, providing a comprehensive and efficient news consumption experience.

- **Refined Keyword Extraction for Text Summarization:** Elevating text summarization through refined keyword extraction techniques ensures the generation of concise and relevant summaries. By focusing on extracting the most essential keywords, the platform delivers high-quality summaries that capture the core essence of news articles, facilitating efficient information consumption.

- **Tone Analysis of Articles:** Incorporating tone analysis enables users to gain deeper insights into the emotional context and subjective opinions expressed in news articles.

## 8 FUTURE SCOPE

- Using RAG LLM model with or in place of SerpAPI for more accurate/relevant Link generation.

- Language support will be extended beyond English and Hindi to accommodate a diverse user base, ensuring accessibility and inclusivity.

- The user interaction and experience will be improved through a streamlined and intuitive interface, enhancing usability and engagement.

- Collaborate with news organizations to ensure the freshness and relevance of content

## CONCLUSION

In conclusion, SentiNEWS represents a significant leap forward in the realm of news consumption, offering a comprehensive and user-friendly solution for navigating the complexities of today's information landscape. By integrating advanced technologies such as natural language processing, sentiment analysis, and dynamic image generation, SentiNEWS not only streamlines the process of accessing news but also enhances its accessibility and engagement. The inclusion of features like text-to-speech for visually impaired users, bilingual support, and keyword-based image crafting demonstrates a commitment to inclusivity and innovation. Moving forward, SentiNEWS holds immense promise in revolutionizing how individuals stay informed, ensuring that everyone, regardless of their preferences or abilities, can access timely and relevant news content efficiently.

## REFERENCES

- [1] M. Kabadjov, J. Steinberger, B. Pouliquen, R. Steinberger, and M. Poesio, "Multilingual Statistical News Summarisation: Preliminary Experiments with English," Jan. 2009, pp. 519–522
- [2] A. Balahur, R. Steinberger, M. Kabadjov, and V. Zavarella, "Sentiment Analysis in the News".
- [3] U. Swati, C. Pranali, and S. Pragati, "SENTIMENT ANALYSIS OF NEWS ARTICLES USING MACHINE LEARNING APPROACH," vol. 2, no. 4.
- [4] T. B. Mirani and S. Sasi, "Two-level text summarization from online news sources with sentiment analysis," in 2017 International Conference on Networks and Advances in Computational Technologies (NetACT), Thiruvanthapuram, India: IEEE, Jul. 2017, pp. 19–24.
- [5] Facebook, "BART Large CNN Model," Available: <https://huggingface.co/facebook/bart-large-cnn>
- [6] Hugging Face, "Diffusers Documentation - Basic Training," Available: [https://huggingface.co/docs/diffusers/en/tutorials/basic\\_training](https://huggingface.co/docs/diffusers/en/tutorials/basic_training)