Ethical AI-Powered News Summarization with Core Insights

Mrs. J. A Lavanya[1] M. Tech (PhD), Assistant Professor, ANITS
B. Revanth[*2], M. A. Surya Prakash[3], M. Vamsi Sai Krishna[4], M. Vinay Kumar[5]
[1-5] Computer Science Engineering with Data Science, Anil Neerukonda Institute of Technology and Sciences, Visakhapatnam

---

Abstract - The rise of digital news makes staying informed challenging. This paper presents an AI-powered news summarization system using NLP and machine learning to extract, analyze, and generate concise, factual summaries. It integrates keyword extraction, sentiment analysis, and ethical evaluation, ensuring unbiased, multi-perspective insights. Leveraging GPT-based models and multi-stage attention mechanisms, the system enhances responsible journalism and information transparency.

***Key Words*:** GPT-3, NLP, Multi-Stage Attention Mechanisms, OLLAMA, LLAMA 3.2, Gemini API

1. INTRODUCTION

Digital news platforms, social media, and online journalism have transformed how people engage with news. Readers frequently find it challenging to process vast amounts of information. The rise of clickbait headlines, misinformation, and biased reporting has raised concerns about the credibility and ethical standards of news sources. In an age where false information spreads faster than facts, developing automated solutions is crucial to help users efficiently access accurate, relevant, and unbiased news summaries.

AI-driven news summarization offers a potential solution to this challenge. Most existing models prioritize content reduction but overlook ethical integrity and relevance.

The abstract summary is for generating coherent summaries. There is ethical bias detection that filters misinformation and biased narratives. To highlight the most critical insights.

1.1 The Necessity for AI-Powered News

Individuals must read lengthy articles, compare various sources, and manually filter out irrelevant or misleading content in traditional news consumption. Artificial intelligence offers a solution to these challenges.

Concise summaries of lengthy articles are generated, incorporating ethical analysis to eliminate bias or misinformation. Key insights are emphasized while ensuring factual accuracy. Existing text summarization methods fall into two categories: extractive summarization, which selects key sentences, and abstractive summarization, which generates new text based on the original content's meaning.

Maintaining factual accuracy, reducing bias, and ensuring ethical integrity in news summaries are major challenges in modern NLP-based summarization models. This paper proposes an advanced AI-driven news summarization system that combines NLP techniques with ethical bias detection and key insight extraction to deliver accurate, unbiased, and socially responsible news summaries.

1.2 Challenges in Existing AI- Powered News Summarization

Despite advancements in text summarization, challenges remain. Biases in training data can lead to skewed summaries, while a lack of context awareness may cause misleading interpretations. Large language models can also introduce factual inaccuracies, affecting reliability.

Most systems fail to assess misinformation, propaganda, or sensationalized content, making it difficult to ensure ethical reporting. This can contribute to the spread of misleading narratives.

To solve this, our model integrates bias detection, fact verification, and insight extraction, ensuring summaries are accurate, fair, and reliable for informed decision-making.

1.3 Research Objectives

This research aims to develop an advanced news summarization system. A reliable model is essential to produce clear, concise, and well-structured summaries. By integrating an ethical analysis framework, the system can detect bias, misinformation, and political influence, ensuring more accurate and trustworthy news delivery.

Users can get actionable and structured information by getting key insights from articles. ROUGE scores, factual consistency tests, LLAMA 3.2 summaries and user

feedback are used to evaluate the model.

1.4 Contributions of This Paper

The work's main contributions are as follows:

- Developing an advanced AI-based news summarization model.
- Identifying and mitigating bias and misinformation in news content.
- Extracting and presenting key insights from articles.
- Comparing the system's performance with existing summarization techniques.

2. RELATED WORK

Due to the explosion of online content, the field of automatic text summarization has gained significant attention. Various approaches have been proposed to tackle this problem.

2.1 Extractive vs. Abstractive Summarization

The traditional methods for news summarization were based on selecting key sentences from the text. Graph-based ranking methods can be used to determine the importance of a sentence. Many of the methods produce disorganized summaries.

In contrast, abstractive summarization uses deep learning models to generate summaries that are similar to human summaries. Transformer-based architectures have greatly improved the summarization quality. New benchmarks have been set by learning from large-scale data.

2.2 AI-Based Ethical News Summarization

Bias detection and ethical analysis remain underdeveloped despite the advancement of summarization. Research was done by Baly et al. Political bias in news articles was explored in 2020.

Deep learning was used for misinformation detection. There are few works that integrate fact verification into summarization models. The Fact-Checking Transformer and Truth-aware Summarization tried to address this issue.

Our work combines abstractive summarization with ethical analysis to ensure fairness, factual accuracy, and reliability in news content.

3. SYSTEM ARCHITECTURE

Our proposed AI-powered news summarization system consists of multiple components designed to extract, analyze, and refine news content efficiently. 3.1 Architecture Overview The system comprises five major modules:

3.1.1. News Data Collection: Scrapes news articles from multiple sources (RSS feeds, APIs, web crawling).

3.1.2. Text Preprocessing: Cleans the extracted text (removes HTML tags, stop words, duplicates).

3.1.3. Summarization Engine: Uses LLM-based models for abstractive summarization.

3.1.4. Ethical Analysis Module:

- Bias detection (identifies political, social, and sensational biases).
- Fact verification (cross-checks statements with verified sources).

3.1.5. Key Insight Extraction: Highlights actionable insights, key statistics, and major events from articles.

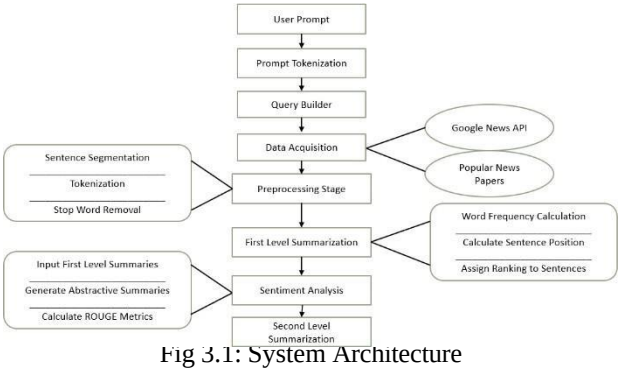

Fig 3.1: System Architecture

3.2 Workflow

- User Input: The user provides a query (e.g., "Recent AI advancements").
- Article Retrieval: The system fetches relevant articles based on the query.
- Summarization: The model generates a concise summary of each article.
- Ethical Review: The system detects biases, misinformation, and missing context.
- Final Output: The user receives a summary, ethical rating, and key insights in an easy-to-read format.

4. IMPLEMENTATION DETAILS

4.1 Technology Stack

Our system is built using:

- Programming Language: Python
- Web Framework:
  - TypeScript
  - ReactJS
- LLM Framework: Ollama (serving Llama 3.2 for text summarization)
- AI & LLM Integration:
  - langchain
  - langchain_ollama
  - google.generativeai
- Web Requests & APIs:
  - requests
  - httpx
  - aiohttp
- Web Scraping:
  - Beautiful Soap
- Environment & Configuration Management:
  - python-dotenv
  - PyYAML
- Parsing & Data Handling:
  - Beautiful Soap
  - LLAMA 3.2
- Additional Tools & Utilities:
  - Gemini API
  - Google Custom Search Engine
  - Google Search API

4.2 System Setup and Requirements

The Ollama should be downloaded locally or kept in docker to run the whole system that is designed. The system is designed to run on a server-based architecture, where Ollama serves as the engine for handling text summarization. The Llama model is used to create summaries. It's possible to deploy the application on a cloud-based infrastructure or a local server.

4.3 News Summarization with Llama 3.2

News summaries with LLAMA the summarization process follows a structured path. The articles are retrieved from trusted sources which are mentioned in our google custom search engine. Text Preprocessing cleans the news content, removing unnecessary elements like advertisements and stop words.

The pre-processed text is passed to the Llama model which creates on point summaries from each source while preserving the key points that are mentioned in the query as per user requirements . The summary has been refined to make it easier to read which is handled by LLAMA.

4.4 Ethical Analysis and Bias Detection

The system uses ethical analysis to ensure accurate summaries. This is achieved through:

- Emotional bias can be identified by grouping the news as neutral, positive or negative.
- Analysing political leaning and media bias by detecting language patterns that indicate subjectivity.

Key claims are cross-referenced with reliable fact-checking database to detect misinformation.

5. IMPLEMENTATION DETAILS

5.1. News Articles Fetching (Data Collection)

The provided user query is first broken down and transformed into google search terms. The most relevant articles are retrieved from the most relevant keywords. The URLs are log for validation purposes.

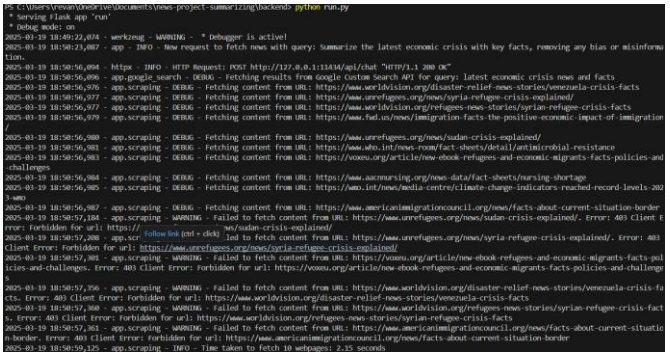

Fig 5.1: Fetching Articles Using Google Search Engine

5.2. Preprocessing and Data Structuring

Rephrase Data is retrieved from articles. The system takes the text from the articles. The data contains complete paragraphs from each source.

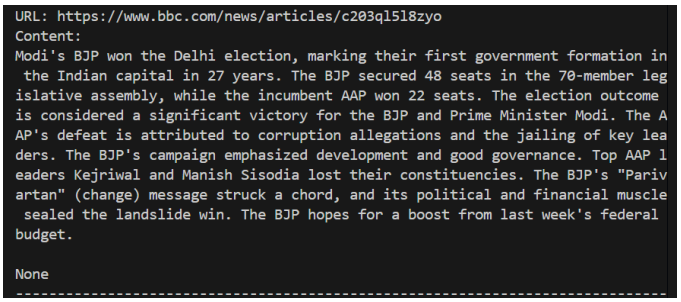

Fig 5.2: Data Extraction from Articles

5.3. Metadata Structuring for Classification

Metadata structuring is used for classification. The data is used to categorize and filter the articles. This includes:

- Taxonomy of the article is a related topic.
- The latest news flag tells you if the article reports breaking news.
- "Politics" and "India" are broad classifications. The original user input is kept for reference.

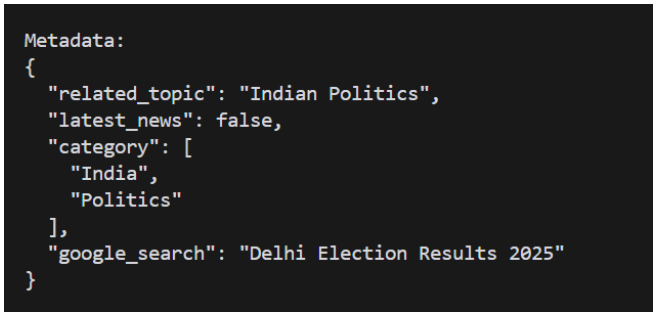

Fig 5.3: Extracted articles are classified into categories like "Politics" or "India."

5.4. Queueing Data for Summarization

Data can be queued for analysis. The summarization process involves storing the content in a queue. The queue structure ensures efficient processing.

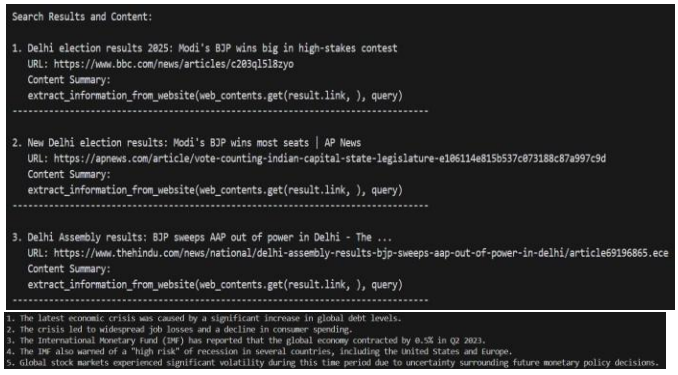

Fig 5.4: Extracted data is queued for efficient processing

5.5. Final Summarization Output

A summary is generated from multiple sources. It gives key insights without being redundant.

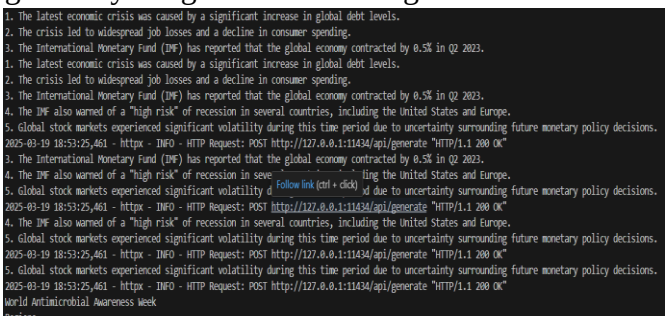

Fig 5.5: System converts raw extracted text into meaningful insights

5. EVALUATION METRICS

We evaluate summarization quality, ethical analysis accuracy, and insight relevance.

5.1 Summarization Evaluation

ROUGE Score (Recall-Oriented Understudy for Gisting Evaluation) is a measure of textual overlap with human-written summaries. The Bilingual Evaluation Understudy has a BLEU score.

5.2 Ethical Analysis Evaluation

The bias detection accuracy was evaluated using a labelled dataset. Fact Verification F1 Score is a measure of accuracy and recall of false claims detection.

5.3 Insight Extraction Evaluation

Insight extract evaluation Experts rate the insights on relevance. Measure how well the system extracts information. The proposed system is superior to baseline models by maintaining higher factual accuracy and ethical fairness.

6. FUTURE SCOPE

Our system has some areas for improvement.

- The system focuses on English-language news. It is possible to enhance accessibility by extending support to multilingual news sources.

- Real-time fact-checking can be used with external databases to improve misinformation detection.
- Users wonder how artificial intelligence makes summaries. Future versions should have explainable techniques.
- A more personalized experience can be offered by giving users the ability to adjust summary length, bias sensitivity and ethical parameters.
- There is integration with news platforms. The system can be used as a mobile app or a browser extension.

REFERENCES

1, Liu, Y., Zheng, Y., Wan, X., & Sun, X. (2022). "News Summarization and Its Ethical Challenges". arXiv preprint arXiv:2209.12356. Retrieved from https://arxiv.org/pdf/2209.12356
2, Narayan, S., Cohen, S. B., & Lapata, M. (2018). "Don't Give Me the Details, Just the Summary! Topic-Aware Convolutional Neural Networks for Extreme Summarization". Proceedings of EMNLP 2018.
3, Zhang, J., Zhao, Y., Saleh, M., & Liu, P. J. (2020). "PEGASUS: Pre-training with Extracted Gap-Sentences for Abstractive Summarization". Proceedings of ICML 2020.
4, Bae, S., & Lee, Y. (2021). "Fact-Check or Fake? Analyzing the Credibility of News Content Using NLP Models". Journal of Computational Journalism, 9(1), 25-40.
5, Brown, T., Mann, B., Ryder, N., et al. (2020). "Language Models Are Few-Shot Learners". NeurIPS 2020.
6, Wang, L., Yang, Y., & Su, Z. (2022). "Bias Mitigation in News Summarization: Challenges and Opportunities". Computational Journalism Conference 2022.
7, Raffel, C., Shazeer, N., Roberts, A., et al. (2020). "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer". Journal of Machine Learning Research.
8, C. Guan, A. Chin, and P. Vahabi, "Enhancing News Summarization with ELearnFit through Efficient In-Context Learning and Efficient Fine-Tuning," 2024.
9, T. Goyal, J. J. Li, and G. Durrett, "News Summarization and Evaluation in the Era of GPT-3," 2023.