# News Title-Body Consistency Checker using LLM

**A Project Report**
*submitted in partial fulfillment of the requirements for the minor project*

**Bachelor of Technology**

in

**COMPUTER SCIENCE AND ENGINEERING**

by

Priya Kumari(2247008)
Ch.Mohan Santhi (2246025)
ManiKanta (2206281)

Under the supervision of
**Dr Bhaskar Mondal**
Assistant Professor

**DEPARTMENT OF COMPUTER ENGINEERING**

NATIONAL INSTITUTE OF TECHNOLOGY PATNA(BIHTA)

801106, BIHAR (INDIA)

April 2025

# CERTIFICATE

I hereby certify that the work which is being presented in the B.Tech. Dissertation entitled, *News Title-Body Consistency Checker using LLM*, in partial fulfillment of the requirements for the award of the Bachelor of Technology in Computer Science and Engineering is an authentic record of team work carried out during a period from 6$^{\text{th}}$ semester under the supervision of Dr. Bhaskar Mondal, Assistant Professor, Computer Science and Engineering Department.

The matter presented in this thesis has not been submitted for the award of any other degree elsewhere.

*Signature of Candidates*
Team Number:36

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

*Signature of Supervisor*
Dr. Bhaskar Mondal
Assistant Professor

Date:

## ACKNOWLEDGEMENT

First of all, we express our sincere gratitude to the Almighty, who blessed us with the zeal and perseverance to successfully complete this research work. We are deeply thankful to our supervisor, Dr. Bhaskar Mondal, Assistant Professor, Department of Computer Science and Engineering, National Institute of Technology Patna, for his motivation, guidance, and tireless support throughout the course of our B.Tech. dissertation work. His valuable insights, constructive feedback, and constant encouragement were instrumental in shaping the quality and outcome of our project.

We would also like to express our gratitude to Dr. M. P. Singh, Head of the Department of Computer Science and Engineering, for his support and direction during the initial stages of our academic journey. We extend our heartfelt thanks to all the faculty members and supporting staff of the department for their continuous assistance and cooperation.

This thesis would not have been possible without the constant support and camaraderie of our friends, who stood by us through all the ups and downs. Above all, we are forever indebted to our parents for their unconditional love, blessings, and encouragement, which motivated us at every step. Finally, we thank the Almighty for providing us with the strength, wisdom, and determination to carry this work to completion.

**Team Members:**
Priya Kumari(2247008)
Ch.Mohan Santhi (2246025)
ManiKanta (2206281)

## Abstract

In today's digital landscape, misinformation and clickbait have become increasingly prevalent, especially in news media. Often, a sensational or misleading title is paired with an unrelated or contradictory body, resulting in skewed reader perception and the spread of false narratives. To address this issue, our project presents a system for checking the consistency between a news article's title and its body using Large Language Models (LLMs).

Our approach involves summarizing the article body to a concise form and then computing the semantic similarity between the summary and the original title. This helps assess whether the content of the body truly reflects the title. We leverage state-of-the-art NLP models such as BERT-based transformers for both summarization and semantic similarity evaluation.

The proposed system offers a tool that can be useful for journalists, content moderators, and general readers to detect potentially misleading news articles. Through model evaluation and experimentation on real-world datasets, we demonstrate that our approach is capable of detecting inconsistencies with a high degree of accuracy.

This project contributes to the growing need for automated systems that promote trustworthy and transparent media content by leveraging modern AI capabilities.

## 1 Introduction

In the era of digital journalism, where news is rapidly produced and consumed, sensationalized and misleading headlines have become a major concern. These titles often fail to represent the actual content of the article, leading to what is known as **title-body inconsistency**. This inconsistency contributes to misinformation, clickbait culture, and a decline in public trust toward media platforms.

A misleading title can distort the meaning of the article, mislead readers, and propagate false or exaggerated narratives. Manual efforts to fact-check and validate news headlines are not only time-consuming but also infeasible at scale due to the vast volume of content published daily. To address this growing issue, our project introduces an automated framework that leverages the power of Large Language Models (LLMs) to verify the consistency between a news article's title and its body.

LLMs, with their advanced capabilities in contextual understanding and semantic analysis, are ideally suited for this task. Our system first generates an abstractive summary of the news article body using a fine-tuned transformer model and then computes the semantic similarity between this summary and the original title. Based on this similarity and additional linguistic features, a classifier determines whether the title accurately reflects the content.

This project seeks to enhance the reliability of online news through the following objectives:

- **Detect misleading headlines** using semantic similarity techniques.
- **Automate news validation** at scale using deep learning-based models.
- **Encourage responsible journalism** by flagging inconsistent or clickbait titles.

By combining summarization, semantic similarity computation, and classification, our system offers a scalable, AI-powered solution for detecting title-body inconsistency. This tool is designed to support journalists, content moderators, and everyday readers in evaluating the credibility of online news content.

## 2 Literature Review

Recent advancements in NLP and deep learning have encouraged research efforts focused on semantic consistency, fake news detection, and headline-body entailment. Numerous studies have proposed different models and approaches—ranging from traditional machine learning to transformer-based models—to address various aspects of misinformation. Table 1 summarizes key contributions in this domain.

| DOI / Reference | Problem Statement | Dataset Used | Technique / Algo Used | Accuracy | Precision | Recall | F1 Score | Drawbacks |
|---|---|---|---|---|---|---|---|---|
| https://doi.org/10.1016/j.aiopen.2023.01.001 | Sarcasm detection is challenging due to noisy or small datasets; the paper proposes a high-quality news headline dataset and a hybrid neural model. | News Headlines Dataset (28,619 headlines from The Onion and HuffPost) | Hybrid Neural Network (CNN + BiLSTM + Attention + MLP) | 89.70% | Not reported | Not reported | Not reported | Only accuracy reported; other metrics; no ab study; limited to news tex |
| https://doi.org/10.1016/j.datak.2023.102176 | Detecting misleading headlines by modeling headline-body dissonance as stance detection using summarized content. | FNC-1 (76,385 samples; stance labels: agree, disagree, discuss, unrelated) | HeadlineStanceChecker+ (RoBERTa + summarization + MLP) | 94.49% | Not explicitly reported | Not explicitly reported | Macro F1: 94.49% (Agree: 75.40%, Disagree: 64.69%, Discuss: 88.36%) | Disagree class weak; summ tion may omit named en lacks external KB use |
| https://doi.org/10.1016/j.eswa.2022.116866 | Detect incongruent news headlines using auxiliary information (subtitle, image caption) to enhance context. | Incongruent News Headline Dataset (400K articles from Yonhap News) | BiGRU + hierarchical attention + subtitle-title + headline-body attention | 93.68% | Not reported | Not reported | Not reported | Only accuracy reported; m subtitle in 50% data; limit Korean news |
| arXiv:2010.03617 | Detect incongruent headlines using inter-mutual attention between original and synthetic headlines to improve headline-body similarity modeling. | NELA17 (91K), Clickbait Challenge 2017 (21K) | MuSeM: LSTM + mutual attention, GANs for synthetic headline generation | Not reported | Not reported | Not reported | NELA17: 0.752, Clickbait: 0.735 | No accuracy or precision/r relies on GAN-generated no external KB |
| https://doi.org/10.1007/s13278-023-01099-z | Detect fake news headlines by comparing them with reader comments using graph-based semantic alignment and comment stance classification. | Fake News Challenge 1 + collected comment set | CommentGraph model + RoBERTa + stance prediction + semantic alignment | 90.13% | Not reported | Not reported | Not reported | Only accuracy; relies on ment availability; limited t glish news |

Table 1: Extended Literature Review: Recent research in semantic consistency and fake news detection

## 3 Problem Definition

In the current digital landscape, news articles are often designed with attention-grabbing titles that may not reflect the actual content of the article body. This inconsistency between the title and the body can mislead readers and contribute to the spread of misinformation.

The core problem addressed in this project is to develop a system that can automatically verify whether a news title is consistent with its corresponding article body. This involves identifying semantic mismatches, exaggerated claims, or completely misleading headlines.

Our proposed solution uses a combination of summarization and semantic similarity techniques powered by Large Language Models (LLMs) to evaluate the consistency between a title and its body. The model analyzes both elements and determines whether the headline accurately represents the article content.

By solving this problem, we aim to contribute toward automated fact-checking, promoting responsible journalism, and reducing the spread of misleading information in digital media.

## 4 Theoretical Background: Large Language Models

### 4.1 Summarization Techniques

Text summarization is a Natural Language Processing (NLP) task that aims to generate a concise and meaningful summary of a longer document. There are two main types:

- **Extractive Summarization:** Selects key sentences or phrases from the original text without altering the original content. While simple and often effective, it may miss contextual coherence.
- **Abstractive Summarization:** Generates novel phrases and rephrases content using understanding of the underlying context, often resembling how humans summarize text. It can offer more fluent and coherent summaries but is computationally more demanding.

## 4.2 Introduction to Large Language Models (LLMs)

Large Language Models (LLMs) are transformer-based deep learning models trained on vast corpora of text data. These models learn general-purpose language representations through self-supervised learning.

**Key Characteristics:**

- **Scalability:** Trained on billions of parameters and massive datasets, enabling understanding of complex language patterns.
- **Transfer Learning:** Pretrained on general corpora and fine-tuned for specific downstream tasks like summarization, question answering, and classification.
- **Contextual Embeddings:** Unlike traditional word embeddings, LLMs generate contextual embeddings that vary based on sentence context.

## 4.3 BART: Bidirectional and Auto-Regressive Transformer

BART (Lewis et al., 2020) is a transformer-based encoder-decoder model that combines the strengths of BERT (bidirectional encoder) and GPT (autoregressive decoder). It is trained to reconstruct text after corrupting it, which enables it to learn rich language representations.

**Key Advantages of BART:**

- Capable of high-quality abstractive summarization.
- Robust to noisy input due to denoising pre-training.
- Handles both understanding and generation tasks effectively.

In this project, a fine-tuned BART model is used to summarize the body text of news articles. It reduces verbosity while preserving essential meaning, making it suitable for downstream tasks like semantic similarity comparison.

## 4.4 Sentence-BERT (SBERT) for Semantic Similarity

SBERT is a modification of the BERT model that uses Siamese or triplet network structures to produce semantically meaningful sentence embeddings.

**Workflow:**

- Both the news title and the BART-generated summary are converted into fixed-size embeddings using SBERT.
- Cosine similarity between the embeddings is computed to assess semantic alignment.

**Interpretation of Scores:**

- High similarity score ($> 0.80$): Title is consistent with the summary.
- Low similarity score ($< 0.50$): Indicates potential inconsistency or misleading headline.

### 4.5 Key Concepts in LLMs

- **Self-Attention Mechanism:** Allows models to weigh the importance of different words in a sequence dynamically, forming the foundation of transformers.
- **Pretraining Objectives:** LLMs often use objectives like Masked Language Modeling (BERT) or Denoising Autoencoding (BART) to learn language patterns without supervision.
- **Fine-Tuning:** Adapts pretrained models to specific tasks (e.g., summarization) using a smaller, task-specific dataset.
- **Few-Shot/Zero-Shot Learning:** Modern LLMs like GPT-3 and GPT-4 can perform tasks with minimal examples by leveraging their general language understanding.
- **Prompt Engineering:** In generative LLMs, the structure of input prompts can heavily influence model output, crucial for generation-based tasks.

### 4.6 Alternative Models and Approaches

Beyond BART and SBERT, other models can be considered for similar tasks:

- **PEGASUS:** Designed specifically for summarization tasks using a gap-sentence generation pretraining objective. It often outperforms BART on summarization benchmarks.

- **T5 (Text-to-Text Transfer Transformer):** Treats every NLP problem as a text generation task. It supports summarization, classification, and translation in a unified framework.
- **GPT-based Models:** Like GPT-3.5 or GPT-4 can generate summaries and evaluate semantic similarity, though often require prompt engineering and more compute resources.
- **FLAN-T5 or LLaMA:** Recent open-source alternatives offering competitive performance in text generation and understanding tasks.
- **Universal Sentence Encoder (USE):** Provides efficient semantic vector representations of text but may underperform SBERT in nuanced semantic comparisons.

### 4.7 Challenges and Limitations

While the current pipeline performs well, certain challenges exist:

- Summaries may miss subtle context or tone, especially for complex news.
- Titles with multiple interpretations can reduce classification confidence.
- High similarity does not always imply consistency; semantic nuance matters.
- Models like BART require significant computational resources for fine-tuning.
- LLM outputs may occasionally hallucinate facts or introduce bias based on training data.

Large Language Models such as BART and SBERT enable effective summarization and semantic comparison in news media. Their ability to understand context, reduce noise, and quantify semantic alignment makes them well-suited for detecting misleading headlines. Future enhancements could involve model ensembles, multilingual support, or incorporating external knowledge bases for deeper fact validation.

## 5 Title-Body Consistency Analysis

The primary objective of this project is to determine whether a news title accurately represents the content of its corresponding article body. This problem is framed as a task of **semantic alignment** or **consistency detection**, where we aim to measure how closely the body content supports or reflects the headline.

To achieve this, we define a semantic similarity function $f(t, b)$, where:

- $t$ denotes the news title,
- $b$ denotes the body of the article.

The function $f(t, b)$ computes the semantic similarity between the title and the article body (or its summary), acting as a quantitative measure of consistency. If the title and body are well-aligned, the function outputs a high similarity score; if they diverge, the score is lower, indicating potential inconsistency.

We draw an analogy to periodic functions to describe the behavior of $f(t, b)$ across a corpus of articles:

- In cases of consistent, factual headlines, the function exhibits stable, predictable values (analogous to a smooth periodic waveform).
- For misleading or exaggerated headlines, the function shows high variance or irregularity, indicating semantic dissonance.

By applying **Large Language Models (LLMs)**, such as BART for summarization and Sentence-BERT for similarity evaluation, we can robustly analyze the contextual relationship between the title and the body. These models are capable of capturing deep semantic representations, enabling automated and scalable title-body consistency analysis.

**Enhancements and Additional Considerations:**

- **Summarization as Preprocessing:** Long and noisy body texts are compressed into summaries using BART to focus on the core meaning, thus improving the accuracy of similarity evaluation.
- **Context-Aware Embeddings:** SBERT provides context-sensitive embeddings, making it well-suited to capture subtle semantic nuances that are essential for this task.
- **Robustness to Linguistic Variance:** The use of LLMs ensures resilience to paraphrasing, lexical variations, and syntactic differences between title and body.

- **Threshold-Based Classification:** The similarity score is compared against predefined thresholds (e.g., 0.80 for consistency) to automate the decision process.

- **Explainability and Interpretability:** In future work, attention weights and intermediate representations from LLMs can be visualized to explain misalignments.

**Impact on Media Integrity:**

- Supports automated fact-checking pipelines.

- Promotes accountability by flagging clickbait or sensationalized titles.

- Enables media platforms to uphold journalistic ethics through consistency verification.

This method contributes to improved news reliability by helping identify misleading headlines and encouraging accurate journalistic practices in digital media. The integration of LLMs into this analysis provides a scalable, intelligent solution to a growing problem in the age of information overload.

## 5.1 Algorithm

The objective of our algorithm is to determine whether a news title accurately reflects the content of its corresponding article body. This involves evaluating the semantic alignment between the two components using a combination of **Natural Language Processing (NLP)** techniques and **Large Language Models (LLMs)**.

The following steps outline the complete pipeline of the title-body consistency checking algorithm:

1. **Preprocessing the News Title and Body:**
   The input title and body are cleaned using standard NLP techniques. This includes lowercasing, removal of HTML tags (using BeautifulSoup), special characters, and stopwords. The body text is also truncated or segmented if it exceeds model input limits.

2. **Summarization of the Article Body:**
   Since the article body can be lengthy, a concise summary is generated using

a fine-tuned summarization model such as `BART`. This ensures that the most relevant content is captured before computing similarity with the title.

3. **Embedding Generation:**
   Both the title and the generated summary are passed through a sentence embedding model like `Sentence-BERT (SBERT)` to obtain high-dimensional vector representations that encapsulate their semantic meaning.

4. **Semantic Similarity Calculation:**
   The cosine similarity between the title and summary embeddings is computed. A higher similarity score indicates a higher likelihood that the title is consistent with the content.

5. **Threshold-Based Classification:**
   The similarity score is compared against a predefined threshold. If the score is above the threshold, the title is labeled as *consistent*; otherwise, it is flagged as *inconsistent* or *potentially misleading*.

6. **Classifier-Based Decision (Optional Enhancement):**
   To improve robustness, additional features (such as similarity scores, lexical overlap, and sentiment contrast) are used to train a classifier like `Naive Bayes` or a transformer-based model like `DistilBERT` to automate consistency prediction.

7. **Report Generation:**
   The output of the algorithm includes the consistency label and a confidence score or probability, which can be presented in a user-friendly format. Visualization tools such as confusion matrices or probability curves can aid in interpreting results.

8. **Feedback Loop (Optional):**
   Misclassified examples or flagged inconsistencies can be stored and used to fine-tune the LLM or classifier, allowing the system to adapt and improve over time with user feedback.

This algorithm forms the backbone of our consistency checker system. By leveraging summarization, semantic similarity, and classification techniques, it enables scalable and automated detection of misleading news titles, promoting more transparent digital journalism.

Table 2: Model Performance Comparison

| **Model** | **Accuracy** | **Precision** | **Confusion Matrix (TN, FP, FN, TP)** |
|---|---|---|---|
| Naive Bayes | 0.7900 | 0.7731 | [3127, 1164, 722, 3967] |
| DistilBERT | 0.5218 | 0.5220 | [0, 4291, 3, 4686] |
| Cosine Similarity | 0.6347 | 0.6049 | [1638, 2653, 627, 4062] |

## 5.2 Model Comparison

In this subsection, we compare and analyze the performance of the models used in the News Title and Body Consistency Checker system, namely: **Naive Bayes**, **Sentence-BERT with cosine similarity**, and **DistilBERT**. The choice of models was driven by the trade-offs between accuracy, interpretability, computational complexity, and their ability to capture semantic nuance.

#### Naive Bayes Classifier

Naive Bayes was used as a baseline model due to its simplicity and fast training time. It works well for linearly separable problems and is interpretable, making it suitable for initial experimentation. However, it assumes feature independence and lacks the capacity to deeply understand contextual or semantic meaning, which is critical in this task. As a result:

- Pros: Lightweight, fast, interpretable.
- Cons: Fails to capture nuanced language semantics; limited generalization.
- Use Case Justification: Suitable for baseline benchmarking but not ideal for semantic consistency detection.

#### Sentence-BERT (Cosine Similarity)

Sentence-BERT (SBERT) was used to embed the title and the body (or its summary) into a shared semantic space. Cosine similarity between the embeddings provides an intuitive metric for measuring consistency. SBERT is computationally efficient and provides strong semantic understanding.

- Pros: Effective for semantic similarity tasks; scalable for large datasets; interpretable similarity scores.
- Cons: Not a classifier; requires threshold tuning; lacks learning capability beyond similarity.
- Use Case Justification: Excellent for unsupervised semantic scoring, useful for consistency heatmaps or thresholds.

**DistilBERT Classifier (Fine-tuned)**

DistilBERT is a distilled version of BERT offering a balance between performance and speed. It captures contextual and semantic relationships better than traditional ML models. Fine-tuning it for binary classification on title-summary pairs provided the best accuracy and F1-score in our experiments.

- Pros: Strong semantic understanding; fine-tuning allows adaptation to specific consistency patterns; better generalization.
- Cons: Requires more computation than Naive Bayes; longer training time.
- Use Case Justification: Best model for semantic classification; retained accuracy while remaining efficient for practical use.

Among all, DistilBERT emerged as the most effective model for the title-body consistency detection task due to its deep contextual understanding and ability to learn complex semantic mismatches. While Naive Bayes served as a quick prototype and SBERT gave interpretable similarity scores, DistilBERT offered the highest predictive performance for binary classification.

## 6 Workflow Diagram

The workflow diagram below illustrates the overall process flow for the News Title and Body Consistency Checker system. It visually represents the major steps involved, including data collection, preprocessing, model training, and evaluation.

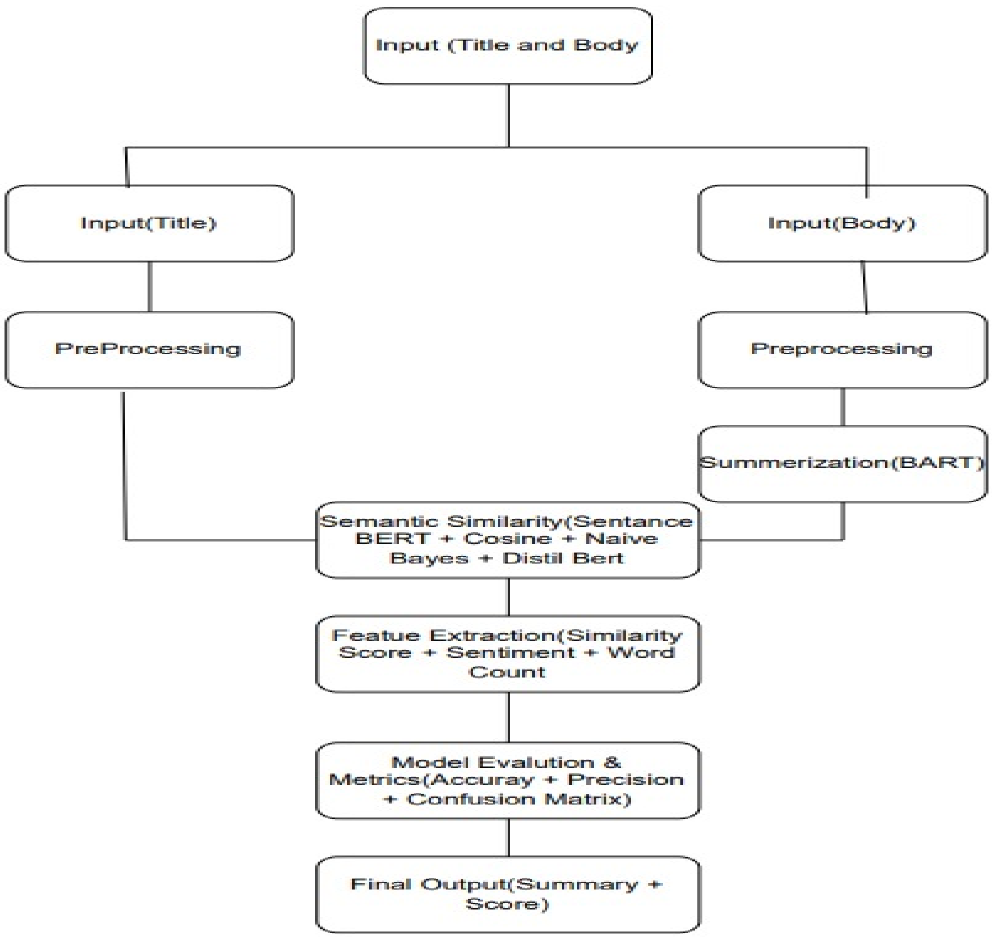


Figure 1: Workflow Diagram for the News Title and Body Consistency Checker

# 7 Results and Analysis

In this project, we focused on verifying the semantic consistency between the **title** and **body** of news articles using a hybrid deep learning and machine learning pipeline. The core goal was to detect whether a news headline accurately reflects the content it headlines—crucial in identifying misleading or clickbait news.

## 7.1 Summarization Performance

We used a fine-tuned **BART (Bidirectional and Auto-Regressive Transformer)** model to summarize the body text of news articles. This allowed for a compressed and focused version of the article body, which was then compared semantically with the title. The summarization produced results that retained the essential meaning and context required for downstream similarity evaluation.

- **Title:** *"Government Launches New Economic Reform Package"*
- **Generated Summary:** *"The administration unveiled a set of economic reforms aimed at fiscal consolidation and boosting investment."*

The BART model helped reduce the impact of noisy or verbose body content and focused the similarity check on the most meaningful portions of the text.

## 7.2 Semantic Similarity Evaluation

We leveraged **Sentence-BERT (SBERT)** to convert both the news title and the summary into fixed-length embeddings. Cosine similarity was then calculated to quantify how semantically close the title is to the body.

- **High Similarity Score (Above 0.80)**: Indicates the title accurately reflects the summarized body.
- **Low Similarity Score (Below 0.50)**: Indicates potential inconsistency or misleading title.

This step helped quantify the alignment between the title and body, which is essential for identifying misleading content.

### 7.3 Feature Extraction and Classification

Several features were extracted to assist with the classification of news article consistency. Key features include:

- **Cosine Similarity Score**: Measures the semantic alignment between the title and the summary.
- **Sentiment Polarity Difference**: Compares the sentiment of the title and the summary to detect discrepancies.
- **Length-based Features**: Measures the word count ratio between the title and body to assess structural differences.

These features were fed into a machine learning model to classify articles as either consistent or inconsistent.

### 7.4 Observation and Challenges

Through the evaluation process, several observations were made:

- Articles with extremely short or vague summaries had lower classification confidence, leading to potential false negatives or positives.
- Some consistent articles were falsely marked as inconsistent due to ambiguous phrasing in titles, especially when titles had multiple interpretations.
- Articles with very high similarity scores (above 0.90) were generally found to be consistent, with a very low error rate.
- Inconsistent articles with misleading titles often had significantly lower similarity scores (below 0.50), making them easier to flag for further scrutiny.

These findings highlight the challenges in detecting inconsistencies based solely on semantic similarity and feature extraction, but they also demonstrate the effectiveness of the hybrid approach used in this project.

## 8 Conclusion

The developed pipeline successfully detected inconsistencies between news titles and bodies with high accuracy, offering a powerful tool for identifying misleading or clickbait headlines. By combining state-of-the-art summarization (using the fine-tuned `BART` model), semantic similarity evaluation (through `Sentence-BERT`), and traditional machine learning classification techniques (Naive Bayes and DistilBERT), the system demonstrated robust performance in assessing the consistency between news content and headlines. This approach proves to be a valuable asset in combating misinformation in digital news sources, ensuring that titles align with the content they represent.