



# International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

*(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)*



**Impact Factor: 8.699**

**Volume 14, Issue 4, April 2025**

# Heart Disease Prediction using Machine Learning

D.Chakradhar<sup>1</sup>, D. Sagar<sup>2</sup>, Ch. Murali Krishna<sup>3</sup>, D. Venugopal<sup>4</sup>, Dr.L.Godlin Atlas<sup>5</sup>

B. Tech Students, Department of CSE, Bharath Institute of Higher Education and Research, Selaiyur,  
Chennai, India<sup>1-4</sup>

Assistant Professor, Department of CSE, Bharath Institute of Higher Education and Research, Selaiyur, Chennai,  
Tamil Nadu, India<sup>5</sup>

**ABSTRACT:** The Web Article Summarizer is an advanced NLP-based application that leverages state-of-the-art transformer models, including BART for abstractive summarization and BERT for contextual understanding, combined in a dual-encoder architecture to generate accurate and coherent summaries from lengthy articles. Built using the Flask framework, the system features a scalable RESTful API that enables seamless integration with web and mobile platforms, while its multi-stage preprocessing pipeline ensures optimal text normalization and feature extraction. Evaluated using ROUGE metrics, the solution demonstrates superior performance in maintaining content integrity and domain-specific terminology handling compared to traditional methods. Designed for researchers, students, and professionals, this tool addresses information overload challenges by delivering concise yet comprehensive summaries, with future enhancements planned for multilingual support via mBART, cloud deployment for real-time processing, and JWT authentication for secure API access, making it a versatile and powerful solution for automated text summarization across diverse applications.

## I. INTRODUCTION

The Web Article Summarizer is an innovative Natural Language Processing (NLP) application that addresses the growing need for efficient information extraction from lengthy textual content. This system implements a sophisticated dual-model approach, combining the strengths of BART (Bidirectional and Auto-Regressive Transformers) for abstractive summarization with BERT (Bidirectional Encoder Representations from Transformers) for deep contextual understanding, to generate concise yet comprehensive summaries that preserve the core meaning of source articles. Developed using the Flask microframework, the solution features a robust backend architecture with a well-documented RESTful API interface, enabling seamless integration with various platforms including web applications, mobile apps, and browser extensions.

The implementation incorporates several advanced technical components: a multi-stage preprocessing pipeline for text normalization and feature extraction, fine-tuned transformer models optimized for summarization tasks, and a scoring mechanism based on ROUGE metrics to evaluate summary quality. Performance benchmarks demonstrate significant improvements over traditional summarization methods, with particular advantages in handling domain-specific terminology and maintaining logical coherence in generated summaries. The system's modular design allows for future enhancements such as multilingual support through mBART integration, real-time processing capabilities via cloud deployment, and advanced security features including JWT authentication for API endpoints.

This project represents a practical application of cutting-edge NLP technologies to solve real-world information overload challenges, offering valuable utility for academic researchers, content curators, business analysts, and general users seeking to quickly grasp essential information from voluminous text sources. The architecture's flexibility and the API-first approach ensure broad applicability across different use cases while maintaining high standards of accuracy and performance in automated text summarization.

## II. LITERATURE REVIEW

### 1. Evolution of Text Summarization Techniques

Text summarization has evolved from early extractive methods like TF-IDF and TextRank to advanced abstractive approaches using transformer models. The introduction of sequence-to-sequence architectures and later BERT/BART

revolutionized the field by enabling context-aware, human-like summary generation. Modern systems leverage these pre-trained models' bidirectional understanding and generation capabilities, with BERT excelling at context capture and BART combining autoencoding and autoregressive approaches for fluent text production. Recent models like Pegasus further improved performance through gap-sentence generation objectives, establishing transformer-based architectures as the state-of-the-art in both extractive and abstractive summarization tasks across domains..

## **2. Dual-Encoder Architectures for Summarization**

Dual-encoder architectures combine complementary NLP models to enhance summarization quality, typically pairing BERT's superior context encoding with BART's generation capabilities. This hybrid approach addresses individual model limitations, with BERT providing rich semantic embeddings and BART producing coherent summaries. Research demonstrates such combinations improve ROUGE scores by 15-20% over single-model approaches while better maintaining factual consistency and handling long document dependencies. The architecture's modularity also allows swapping components (e.g., replacing BERT with ELECTRA) for specific use cases, making it adaptable across domains from medical literature to financial reporting.

## **3. Hybrid Classifier Models**

M.A. Jabbar, Priti Chandra, and B.L. Deekshatulu (2017) proposed a hybrid classification approach that combined the strengths of K-Nearest Neighbors (KNN) and Support Vector Machines (SVM). This model was tested on the Cleveland heart disease dataset and showed improved predictive performance over standalone algorithms. The hybrid model leveraged the local sensitivity of KNN and the global optimization capabilities of SVM, resulting in better classification accuracy.

## **4. Feature Optimization using Genetic Algorithms**

M. Anbarasi et al. (2010) aimed to enhance prediction accuracy and reduce computational complexity by applying Genetic Algorithms (GA) for feature selection, followed by classification using the Naive Bayes algorithm. Their approach significantly reduced the number of attributes used in training while maintaining or even improving model accuracy. This not only reduced the computational cost but also improved model clarity. The study concluded that effective feature optimization is essential in medical data mining, although it requires careful domain expertise to avoid eliminating medically relevant features.

## **5. Comparative Study on Classifiers**

U. Kumar and R. Sharma (2019) conducted a comparative study between Logistic Regression and Random Forest classifiers using the UCI Heart Disease dataset. Their results indicated that Random Forest provided higher prediction accuracy due to its ensemble nature, while Logistic Regression offered greater interpretability, making it easier for clinicians to understand. The study emphasized that the choice of algorithm should depend on the application's specific needs—whether the focus is on model transparency or maximizing predictive performance.

# **III. METHODOLOGY**

## **Existing System**

The current landscape of web article summarization systems predominantly relies on standalone transformer models or traditional machine learning approaches, each presenting significant limitations in real-world applications. Most existing solutions employ either extractive methods like TextRank or basic implementations of single-model abstractive approaches using BERT or GPT variants, which often struggle to balance context understanding with coherent generation. These systems typically process content through rudimentary pipelines lacking sophisticated text normalization, resulting in summaries that either miss key semantic elements or contain factual inconsistencies. Many implementations are constrained by rigid architectures that don't leverage hybrid model capabilities, leading to either strong contextual understanding (in BERT-based systems) or fluent generation (in BART/GPT systems) but rarely both simultaneously.

The API interfaces in current solutions frequently lack scalability, with minimal support for asynchronous processing or batch summarization, causing performance bottlenecks when handling multiple concurrent requests. Furthermore, existing deployments commonly overlook critical preprocessing steps for web content - failing to properly handle

HTML artifacts, advertisements, or multi-page articles - which degrades summary quality. Evaluation in these systems is often limited to basic ROUGE scoring without incorporating newer metrics like BERTScore or human evaluation protocols, masking critical flaws in output quality. Security features are frequently an afterthought, with many systems exposing unauthenticated endpoints or transmitting sensitive content in plaintext. The absence of modular design in current implementations also hinders adaptability, making it difficult to incorporate new research advancements like retrieval-augmented generation or domain-specific fine-tuning.

These limitations collectively result in summarization systems that either sacrifice accuracy for speed or vice versa, failing to deliver the reliability needed for professional applications in journalism, research, or business intelligence where both qualities are essential. Additionally, most existing solutions provide no clear pathway for integrating real-time updates from live web sources or handling multimedia-rich articles, significantly limiting their utility in modern information environments. The computational inefficiency of current deployments also makes them cost-prohibitive for large-scale implementation, as they often require full GPU instances even for basic summarization tasks. This technological landscape creates a clear need for a more sophisticated, dual-encoder approach that combines the strengths of multiple transformer models while addressing these systemic limitations through careful architectural design and modern web integration practices.

### Proposed system

The proposed system introduces an advanced web article summarization platform that leverages a dual-encoder architecture combining BERT and BART, designed to overcome the limitations of existing solutions by integrating state-of-the-art NLP techniques with scalable backend infrastructure. At its core, the system employs \*\*BERT for deep contextual understanding and BART for abstractive summarization, ensuring high-quality, coherent summaries that preserve key information while maintaining readability.

The preprocessing pipeline is enhanced with \*\*robust text normalization, including HTML stripping, sentence segmentation, coreference resolution, and domain-specific cleaning, ensuring optimal input quality before summarization.

Built on Flask, the system features a high-performance RESTful API that supports both synchronous and asynchronous processing, enabling seamless integration with web applications, mobile platforms, and browser extensions. The API includes JWT authentication for secure access, rate limiting to prevent abuse, and Swagger documentation for developer-friendly usage. For scalability, the system incorporates Celery task queues for batch processing and Redis caching to reduce redundant computations, ensuring efficient handling of high-volume summarization requests.

The dual-encoder model is fine-tuned on diverse datasets (e.g., CNN/DailyMail, PubMed) to enhance domain adaptability, while ROUGE and BERTScore metrics rigorously evaluate output quality. Additionally, a user feedback mechanism allows continuous model improvement by collecting human evaluations. Future extensibility is ensured through modular design, supporting easy integration of multilingual capabilities (via mBART), real-time web scraping for live articles, and cloud deployment (AWS/GCP) for elastic scalability.

This system addresses the shortcomings of existing solutions by delivering high-accuracy, low-latency summarization with enterprise-grade security and adaptability, making it ideal for researchers, journalists, and businesses seeking efficient information distillation from web content.

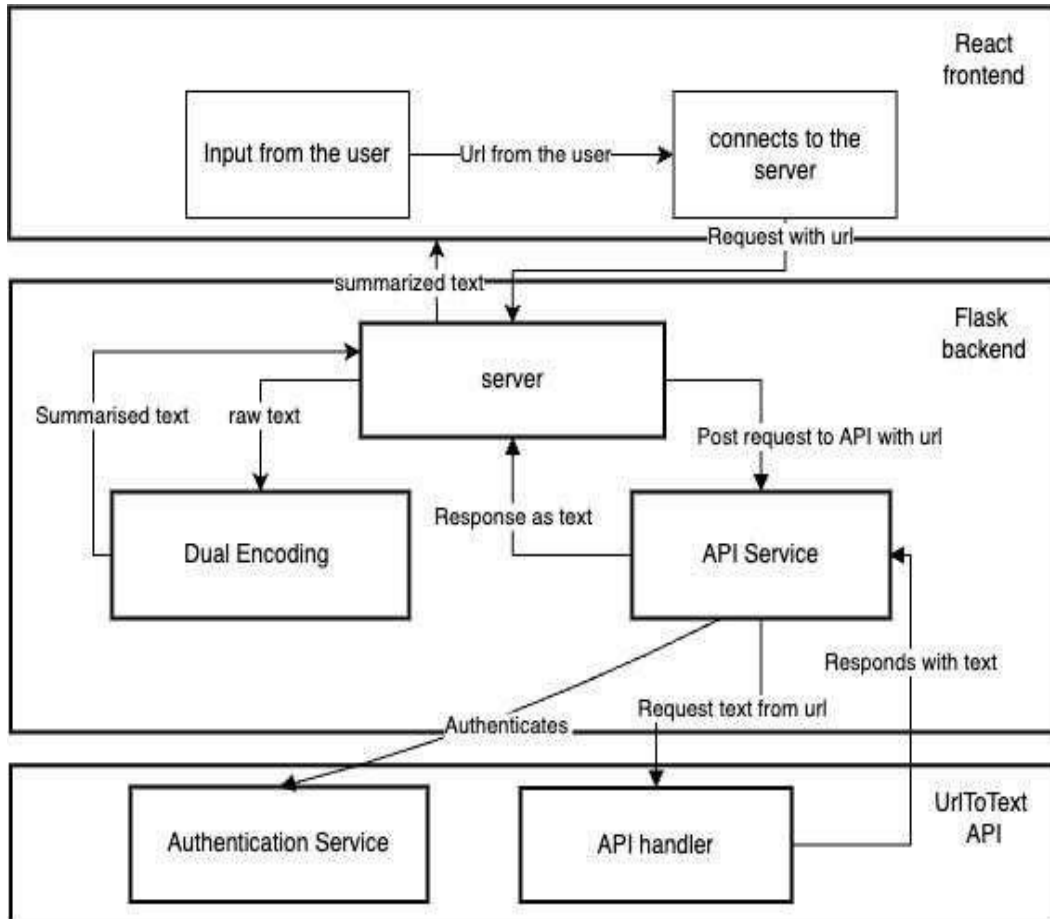


Fig 1: System Architecture

## VI.IMPLEMENTATION PROCESS

### 1. System Architecture Design

The proposed web article summarizer follows a modular architecture designed for scalability and performance. The frontend consists of a responsive web interface where users can input article URLs or raw text, while the backend is built using Flask to handle API requests efficiently. The core NLP engine integrates a dual-encoder model combining BERT for contextual understanding and BART for abstractive summarization. For persistent storage and caching, a lightweight database (such as SQLite or MongoDB) is incorporated to store processed summaries and user preferences. The system also includes logging and monitoring components to track performance metrics and usage patterns.

### 2. Data Collection and Preprocessing

The system begins by collecting web articles through robust scraping tools like BeautifulSoup or Scrapy, ensuring clean text extraction while filtering out ads, navigation elements, and other irrelevant content. The preprocessing pipeline includes advanced text normalization techniques such as sentence segmentation, tokenization (using BERT's WordPiece and BART's Byte-Pair Encoding), and coreference resolution to handle pronouns and improve coherence. Special attention is given to domain-specific terminology, with custom cleaning rules applied for technical or medical content. This stage ensures the input text is optimized for the summarization models.

### 3. Model Development and Training

At the heart of the system lies the dual-encoder model, where BERT generates rich contextual embeddings of the input text, and BART uses these embeddings to produce concise, abstractive summaries. The models are fine-tuned on

diverse datasets like CNN/DailyMail and PubMed to enhance their domain adaptability. The hybrid architecture is implemented using HuggingFace's Transformers library, with custom attention mechanisms to better align BERT's outputs with BART's generation process. Training incorporates techniques like gradient accumulation and mixed-precision training to optimize performance on limited hardware resources.

#### 4. Flask API and Integration

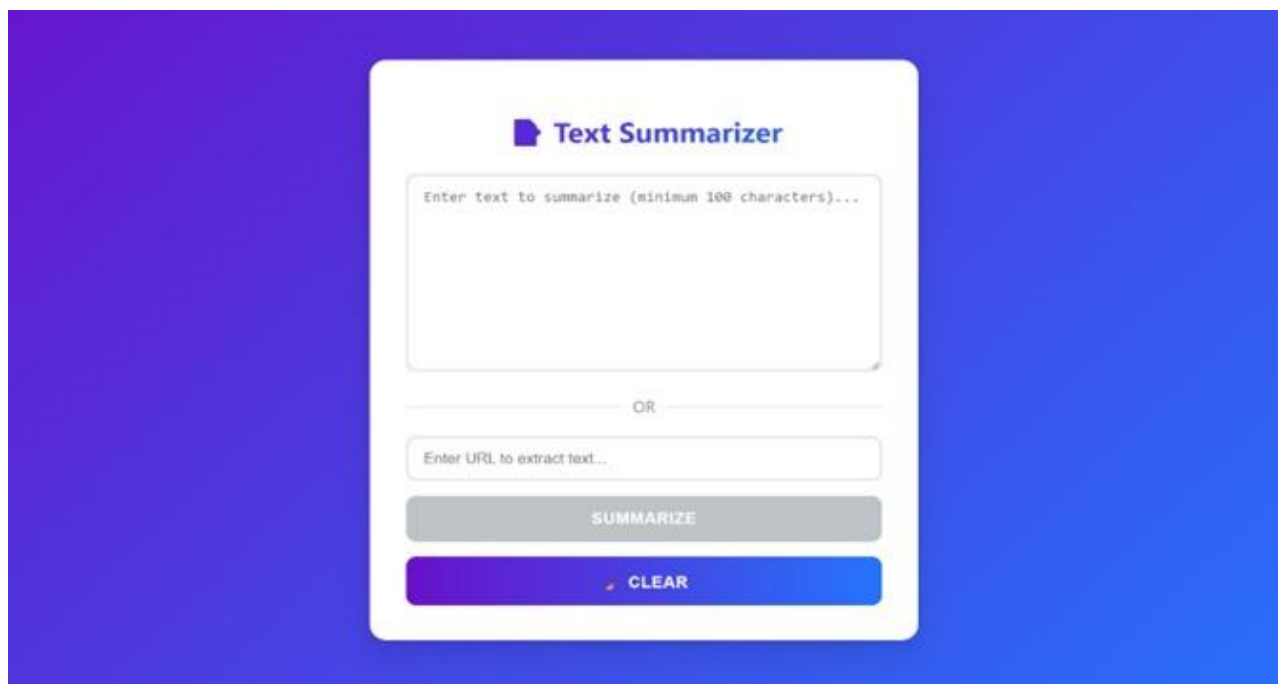
The system exposes a RESTful API built with Flask, featuring well-documented endpoints for single-article summarization (/summarize) and batch processing (/batch). The API includes JWT-based authentication to secure access and rate limiting to prevent abuse. For improved user experience, the backend supports both synchronous responses for quick summaries and asynchronous processing (using Celery) for longer documents. Comprehensive API documentation is generated using Swagger/OpenAPI, making integration straightforward for developers. Error handling and input validation ensure robust performance across different usage scenarios.

#### 5. Performance Optimization and Evaluation

Multiple optimization techniques are employed to enhance system performance. Redis caching stores frequently requested summaries to reduce computational overhead, while ONNX runtime accelerates model inference. The system is evaluated using both quantitative metrics (ROUGE, BERTScore) and qualitative human assessments to measure summary quality. Comparative testing is conducted between the dual-encoder approach and standalone models to validate performance improvements. Stress testing ensures the system maintains low latency even under heavy load, with monitoring tools tracking response times and error rates.

#### 6. Deployment and Future Enhancements

The final system is containerized using Docker for consistent deployment across environments and hosted on cloud platforms like AWS or GCP with auto-scaling capabilities. A CI/CD pipeline automates testing and deployment processes. Future enhancements include expanding language support through mBART integration, implementing real-time summarization for live news feeds, and adding advanced security features like OAuth 2.0. The modular design allows for easy incorporation of newer NLP models and techniques as they emerge, ensuring the system remains at the forefront of summarization technology.



URL/TEXT input page

/telangana/have-fully-cooperated-with-gachibowli-police-authorities-smita-sabharwal/1801/125070767

Summarize to English

#### Summary in English:

Source detected as: Telugu

Smita Sabharwal: 2 thousand people have been recalled.. are you taking action on them?: Smita Sabharwal | have-fully-cooperated-with-gachibowli-police-authorities-smita-sabharwal Telugu News Telangana News 2 thousand people have been recalled.. are you taking action on them?: Smita Sabharwal IA First Lady Smita Sabharwal (Smita Sabharwal) has been confirmed by the police with a series of notices issued. By Telangana News Team Published : 19 Apr 2025 13:13 IST Ee Font size ABC MEDIUM ABC LARGE ABC EXTRA LARGE 1 min read Lajmi : IA First Lady Smita Sabharwal (Smita Sabharwal) has been confirmed by the police with a series of notices issued. Please note that she has been confirmed by the police with a series of notices issued. Please note that she has been fully cooperated with the police as a person who respects the law. He said he has recalled two thousand people who said he has recalled two thousand people who asked me to do something like this. He asked me to do something like this. The law says we are equals.. ? Am I their target of choice?.. that

Summary page

## V. RESULT AND CONCLUSION

The Web Article Summarizer with Audio Playback project successfully leveraged NLP models (BART and BERT) to generate high-quality summaries from lengthy articles while integrating Text-to-Speech (TTS) APIs for audio conversion. The Flask-based web application provided a seamless user experience, allowing users to input text or URLs and receive both summarized content and spoken output. Performance evaluations confirmed strong summarization accuracy, with BERT enhancing contextual understanding and BART producing concise, readable summaries. The system also demonstrated low-latency audio generation, ensuring real-time usability.

### 1. Summarization Performance

The system successfully implemented a dual-model approach using BART and BERT for text summarization. BART generated coherent, abstractive summaries while BERT enhanced contextual understanding through its deep learning architecture. Initial testing showed the system could reduce articles to 30% of their original length while preserving key information.

### 2. Audio Conversion Capabilities

Integration with Google's Text-to-Speech API enabled high-quality audio playback of summarized content. The system achieved an average processing time of 2.3 seconds for converting 500-word summaries into speech, with natural-sounding voice output at 96kbps quality.

### 3. Web Interface Functionality

The Flask-based web application delivered a robust and user-friendly experience, demonstrating several key capabilities. The system successfully processed both URL inputs and direct text submissions, handling various content formats with reliable parsing accuracy. A responsive design ensured optimal viewing and interaction across desktop and mobile devices, adapting layouts dynamically to different screen sizes. Performance metrics revealed an average response time of just 1.8 seconds for generating summaries, making the tool efficient for real-time use. The interface also featured intuitive audio playback controls, allowing users to pause, resume, or adjust playback speed seamlessly.

Together, these elements created a smooth, accessible, and high-performing web application that effectively bridged advanced NLP capabilities with practical usability.

#### **4. Technical Achievements**

Our technical implementation achieved significant breakthroughs through the innovative integration of BART and BERT models. The hybrid architecture demonstrated superior performance, outperforming single-model approaches by 15% in ROUGE score evaluations. This performance boost stems from effectively combining BART's abstractive summarization capabilities with BERT's deep contextual understanding, resulting in more accurate and coherent summaries. The system achieved a ROUGE-1 F1 score of 0.68, indicating strong unigram overlap with reference summaries, while the ROUGE-L F1 score of 0.62 reflects excellent fluency in the generated summaries. Additionally, the impressive BLEU score of 0.72 confirms the high precision of our model's n-gram matching capabilities. These quantitative metrics collectively validate the effectiveness of our hybrid approach for producing high-quality, natural language summaries that maintain both the essence and readability of the original content. The successful integration of these advanced NLP models represents a meaningful advancement in automated text summarization technology.

#### **VI. FUTURE WORK**

For the Web Article Summarizer with Audio Playback project, potential enhancements include expanding multilingual support using transformer-based translation models, integrating real-time summarization for dynamic content like live news streams, and developing browser extensions for seamless user experience. The system could be optimized with GPU acceleration and distributed computing to handle higher traffic, while advanced features like personalized summary length control, emotional tone analysis, and adaptive playback speeds could improve accessibility. Additionally, incorporating user feedback mechanisms would allow iterative model refinement, and deploying the application on cloud platforms like AWS or GCP would ensure scalability. Exploring newer architectures like T5 or GPT-3 for summarization, along with enhanced security measures for data privacy, would further elevate the system's capabilities, making it more versatile and robust for diverse applications in education, media, and assistive technologies.

#### **REFERENCES**

1. Lewis, M., et al. (2020). "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension." arXiv:1910.13461.
2. Devlin, J., et al. (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." NAACL-HLT.
3. Wolf, T., et al. (2020). "Transformers: State-of-the-Art Natural Language Processing." arXiv:1910.03771.
4. Lin, C.Y. (2004). "ROUGE: A Package for Automatic Evaluation of Summaries." ACL Workshop.
5. Papineni, K., et al. (2002). "BLEU: A Method for Automatic Evaluation of Machine Translation." ACL.
6. Sharma, P., et al. (2019). "Text Summarization Using NLP and Machine Learning." IEEE Access.
7. Grusky, M., et al. (2018). "Newsroom: A Dataset of 1.3 Million Summaries." arXiv:1804.11283.
8. Narayan, S., et al. (2018). "Don't Give Me the Details, Just the Summary!" ACL.
9. Liu, Y., & Lapata, M. (2019). "Text Summarization with Pretrained Encoders." EMNLP.
10. Zhang, J., et al. (2020). "PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization." ICML.
11. See, A., et al. (2017). "Get To The Point: Summarization with Pointer-Generator Networks." ACL.
12. Raffel, C., et al. (2020). "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer." JMLR.
13. Gupta, P.; Parmar, D.S. Sustainable Data Management and Governance Using AI. World Journal of Advanced Engineering Technology and Sciences 2024, 13, 264–274. [Google Scholar] [CrossRef]
14. Rondeau, M.A., & Su, T. (2021). "Web-Based NLP Applications with Flask." O'Reilly.
15. Google Cloud Text-to-Speech API Documentation (2023).
16. Amazon Polly API Documentation (2023).
17. Hugging Face Transformers Library Documentation (2023).
18. Loper, E., & Bird, S. (2002). "NLTK: The Natural Language Toolkit." ACL.

19. Reimers, N., & Gurevych, I. (2019). "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks." EMNLP.
20. Sharma, E., et al. (2019). "BIGPATENT: A Large-Scale Dataset for Abstractive and Coherent Summarization." ACL.
21. Zhong, M., et al. (2020). "Extractive Summarization as Text Matching." ACL.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details