AI BASED NEWS SUMMARIZATION BOT

Nityasree G C[1], Shakthisree R C[2], Nisha Devi K[3]

[1] *Student , Information Technology, Bannari Amman Institute of Technology, TamilNadu, India*
[2] *Student,Artificial Intelligence and Data Science, Bannari Amman Institute of Technology, TamilNadu, India*
[3] *Asst Professor , Artificial Intelligence and Data Science, Bannari Amman Institute of Technology, TamilNadu, India*

ABSTRACT

The amount of news available online is overwhelming. In a single day, there are millions of news articles published all over the world. It can be difficult to keep up with the latest news, and even more difficult to find the time to read through long articles. This project proposes a new browser extension that uses generative AI to summarize news articles. The extension would allow users to quickly and easily get the gist of a news article without having to read the whole thing.The extension would use a custom generative AI model that has been trained on a large corpus of news articles. The model would be able to identify the most important information in an article and generate a concise summary. The summary would be presented to the user in a clear and easy-to-read format.The extension would be available for Chrome and Firefox. It would be free to use, and it would not require any personal information from users.The extension has the potential to make it easier for people to stay informed about the news. It could also help people to save time and to focus on the most important information.
In addition to the features mentioned in the draft abstract, the extension would also include the following features - The ability to customize the length of the summary,The ability to save summaries for later reading, The ability to share summaries with others.The extension would be developed using open-source software. This would allow other developers to contribute to the project and to make improvements to the extension.The extension would be evaluated using a user study. The user study would measure the effectiveness of the extension in helping users to stay informed about the news.The extension has the potential to make a significant impact on the way people consume news. It could make it easier for people to stay informed about the news, and it could help people to save time and to focus on the most important information.

Keyword : *news summarizing, generative AI, browser extension, open-source software, user study*

1. INTRODUCTION

In today's fast-paced world, staying informed about current events and news developments can be challenging amidst the vast amount of information available. As the digital age continues to evolve, the need for efficient methods of digesting and comprehending news content becomes increasingly apparent. To address this demand, we introduce a News Summarization Bot powered by state-of-the-art natural language processing (NLP) techniques and the Hugging Face Transformers library.The News Summarization Bot leverages advanced machine learning models trained on extensive datasets to condense lengthy news articles and reports into concise summaries, providing users with a quick overview of the most important information. Utilizing the Hugging Face Transformers API, our bot employs cutting-edge NLP algorithms to analyze and extract key insights from diverse news sources across various topics and domains.

1.1 Significance of the Study

With the exponential growth of online news platforms and social media channels, the volume of news articles generated each day has reached unprecedented levels. As a result, individuals often struggle to keep up with the influx of information, leading to information overload and cognitive fatigue. The News Summarization Bot aims to

alleviate this burden by delivering succinct and informative summaries tailored to the user's preferences and interests. Built on the Python Flask framework, our web service provides a seamless and intuitive interface for users to interact with the News Summarization Bot. Through a user-friendly web application, individuals can input URLs or text snippets of news articles they wish to summarize. The bot then processes the input using advanced NLP models, generating concise summaries that capture the essence of the original content while filtering out extraneous details.

The underlying technology behind our News Summarization Bot relies on state-of-the-art transformer models, such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer), developed by the Hugging Face community. These transformer architectures have revolutionized the field of NLP, enabling machines to understand and generate human-like text with unprecedented accuracy and fluency.

1.2 Contextualizing the Topic

In addition to summarizing news articles, our bot can also perform sentiment analysis to assess the overall tone and sentiment expressed in the text. By analyzing the sentiment of news articles, users gain valuable insights into public opinion and sentiment trends surrounding various topics and events. This feature enhances the user experience by providing deeper insights beyond the basic content summary.

In an era characterized by an overwhelming influx of information from various sources, the traditional approach to news consumption has become increasingly impractical. With the exponential growth of digital media and the internet, users are inundated with vast amounts of text-based content on a daily basis. However, amidst this abundance lies a challenge: the time and effort required to sift through lengthy articles to extract relevant information.Recognizing this pressing need for a more efficient and streamlined method of news consumption, we introduce our innovative News Summarization Bot. Powered by state-of-the-art natural language processing (NLP) models from Hugging Face and seamlessly integrated with Python Flask, our Bot revolutionizes the way users interact with news articles.

1.3 Application of News Summarization Bot

Our News Summarization Bot represents a paradigm shift in the way information is processed and presented. By leveraging advanced machine learning algorithms, the Bot is capable of analyzing the content of news articles across diverse domains, extracting key insights, and condensing them into concise, digestible summaries. This transformative approach enables users to stay informed and up-to-date with current events without the burden of information overload.

In addition to its practical utility, our News Summarization Bot embodies the spirit of innovation and progress in the field of artificial intelligence and natural language processing. As technology continues to evolve and new breakthroughs emerge, our Bot remains at the forefront of the digital revolution, empowering users to engage with news content in a more meaningful and efficient manner.Furthermore, our News Summarization Bot prioritizes user privacy and data security, ensuring that user interactions and data remain confidential and protected. We adhere to industry-standard security protocols and encryption methods to safeguard user information and uphold the highest standards of data privacy.

In summary, the News Summarization Bot represents a groundbreaking solution for navigating the vast landscape of online news and information. By harnessing the power of advanced NLP techniques and transformer models, our bot empowers users to stay informed, make informed decisions, and engage with news content more effectively. Join us on the journey to revolutionize the way we consume and interact with news in the digital age.

2. LITERATURE REVIEW

This chapter is the summary of all literature Surveys related to AI Based News Summarizaion Bot. We studied and reviewed all the relevant cases and discussed with the guide and identified the problem statement.

News summaries frequently include crisp information that provides a quick overview of the topic with all the necessary information. This lets readers quickly grasp the article's most important points, even if they don't have time to read it. People need a fast and efficient way to receive news, and as the amount of information available online grows, news summaries become increasingly important. Data from multiple regional online newspapers are automatically collected and summarized using a transformer-based model, and a user friendly application is proposed in this work. Performance is measured using the ROUGE metric.

Natural Language Processing is booming with its applications in the real world, one of which is Text Summarization for large texts including news articles. This research paper provides an extensive comparative evaluation of extractive and abstractive approaches for news text summarization, with an emphasis on the ROUGE score analysis. The study employs the CNN-Daily Mail dataset, which consists of news articles and human-generated reference summaries. The evaluation employs ROUGE scores to assess the efficacy and quality of generated summaries. After Evaluation, we integrate the best-performing models on a web application to assess their real-world capabilities and user experience Automatic Text Summarization (ATS) is becoming much more important because of the huge amount of textual content that grows exponentially on the Internet and the various archives of news articles,scientific papers, legal documents, etc. Manual text summarization consumes a lot of time, effort, cost, and even becomes impractical with the gigantic amount of textual content. Researchers have been trying to improve ATS techniques since the 1950s. ATS approaches are either extractive, abstractive, or hybrid. The extractive approach selects the most important sentences in the input document(s) then concatenates them to form the summary
.

3. METHODOLOGY PROPOSED
Detailed methodology for creating an AI-based news summarization bot

3.1. Data Collection:

- Gather a large dataset of news articles from various sources and domains. Ensure that the dataset covers a diverse range of topics and writing styles to train a robust summarization model.
- Collect both the full text of the articles and their corresponding human-generated summaries. This paired data will be used for supervised learning to train the summarization model.

3.2. Preprocessing:

- Clean and preprocess the raw text data to remove noise, such as HTML tags, special characters, and irrelevant content (e.g., advertisements, comments).
- Tokenize the text into words or subwords to prepare it for input into the summarization model. Consider using techniques like word tokenization or byte pair encoding (BPE) to handle out-of-vocabulary words and improve model performance.
- Perform additional preprocessing steps such as lowercasing, stemming, lemmatization, and removing stop words to further refine the text data.

3.3. Model Selection:

- Choose an appropriate architecture for the summarization model based on the nature of the task (extractive or abstractive) and the available computational resources.
- For extractive summarization, consider algorithms such as TextRank, LexRank, or graph-based methods.
- For abstractive summarization, consider deep learning architectures such as sequence-to-sequence models (e.g., LSTM, GRU) or transformer-based models (e.g., BERT, GPT, T5).
- Experiment with different model architectures and hyperparameters to find the optimal configuration for the task.

3.4. Training:

- Split the preprocessed data into training, validation, and test sets. The training set will be used to train the model, the validation set for hyperparameter tuning and model selection, and the test set for final evaluation.
- Train the summarization model using the training data and an appropriate optimization algorithm (e.g., Adam, SGD). Monitor the model's performance on the validation set and adjust hyperparameters as needed to prevent overfitting.
- Fine-tune pre-trained models on the news summarization task to leverage transfer learning and improve performance, especially for abstractive summarization.

3.5. Evaluation:

- Evaluate the performance of the trained summarization model using appropriate metrics such as ROUGE (Recall-Oriented Understudy for Gisting Evaluation) scores.
- Calculate ROUGE scores for various metrics (e.g., ROUGE-1, ROUGE-2, ROUGE-L) to assess the quality of the generated summaries in terms of overlap with the human-generated references.
- Conduct qualitative analysis by manually inspecting the generated summaries to assess their coherence, relevance, and informativeness.

3.6. Deployment:

- Deploy the trained summarization model as a service or API that accepts news articles as input and returns summarized versions as output.
- Integrate the summarization bot into news websites, mobile apps, or other platforms to make it accessible to users.
- Monitor the performance of the deployed model in real-world scenarios and collect user feedback to iteratively improve the summarization quality.

3.7. Maintenance and Updates:

- Continuously monitor the performance of the summarization bot and incorporate new data to keep the model up-to-date with the latest news trends and topics.
- Periodically retrain the model with fresh data to adapt to changes in the distribution of news articles and maintain optimal performance.
- Stay informed about advancements in natural language processing and machine learning research to incorporate new techniques and improvements into the summarization bot.

4. CONCLUSIONS

In conclusion, our project introduces a groundbreaking solution in the form of a new browser extension leveraging generative AI technology to summarize news articles effectively. This innovative tool aims to address the challenges associated with information overload by providing users with concise and accurate summaries of complex news content.Through the implementation of state-of-the-art natural language processing techniques, our browser extension can distill lengthy articles into digestible snippets, enabling users to quickly grasp the key points and essential information without having to sift through extensive texts. By harnessing the power of generative AI, our tool ensures that the summaries generated are not only concise but also coherent and contextually relevant.Furthermore, our project prioritizes user experience and accessibility, offering a seamless integration into existing web browsing workflows. With just a click of a button, users can access summarized versions of news articles, saving time and effort while staying informed about current events.Overall, our proposed browser extension represents a significant advancement in the field of information consumption, providing users with a valuable tool to navigate the vast sea of online news content efficiently. By harnessing the capabilities of generative AI, we aim to

empower individuals to make informed decisions and stay up-to-date with the latest developments in a fast-paced world.

In addition to its practical utility, our News Summarization Bot embodies the spirit of innovation and progress in the field of artificial intelligence and natural language processing. As technology continues to evolve and new breakthroughs emerge, our Bot remains at the forefront of the digital revolution, empowering users to engage with news content in a more meaningful and efficient manner.Furthermore, our News Summarization Bot prioritizes user privacy and data security, ensuring that user interactions and data remain confidential and protected. We adhere to industry-standard security protocols and encryption methods to safeguard user information and uphold the highest standards of data privacy.

In summary, the News Summarization Bot represents a groundbreaking solution for navigating the vast landscape of online news and information. By harnessing the power of advanced NLP techniques and transformer models, our bot empowers users to stay informed, make informed decisions, and engage with news content more effectively. Join us on the journey to revolutionize the way we consume and interact with news in the digital age.

5. REFERENCES

These papers cover a range of approaches and techniques used in the field of news summarization bot using AI, including summarization, reinforcement learning, neural network architectures, and more.

[1]Title: "Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond" Author: Ramesh Nallapati, Bowen Zhou, Cicero Nogueira dos Santos, Caglar Gulcehre, Bing Xiang Published Year: 2016

[2]Title: "A Neural Attention Model for Abstractive Sentence Summarization" Author: Alexander M. Rush, Sumit Chopra, Jason Weston Published Year: 2015

[3]Title: "Get To The Point: Summarization with Pointer-Generator Networks" Author: Abigail See, Peter J. Liu, Christopher D. Manning Published Year: 2017

[4]Title: "Deep Reinforcement Learning for Sequence to Sequence Models" Author: Jiatao Gu, James Bradbury, Caiming Xiong, Victor O.K. Li, Richard Socher Published Year: 2017

[5]Title: "Text Summarization with Pretrained Encoders" Author: Yang Liu, Mirella Lapata Published Year: 2019

[6]Title: "A Deep Reinforced Model for Abstractive Summariztion" Author: Romain Paulus, Caiming Xiong, Richard Socher Published Year: 2017

[7]Title: "A Survey of Data Augmentation Approaches for NLP" Author: Weiwei Guo, Shujian Huang, Xiaoyan Zhu, Xinyu Dai, Jiajun Chen Published Year: 2020

[8]Title: "Attention Is All You Need" Author: Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin Published Year: 2017

[9]Title: "Hierarchical Neural Story Generation" Author: Angela Fan, Mike Lewis, Yann Dauphin Published Year: 2018

[10]Title: "SummaRuNNer: A Recurrent Neural Network based Sequence Model for Extractive Summarization of Documents" Author: Ramesh Nallapati, Feifei Zhai, Bowen Zhou Published Year: 2017

[11]Title: "Fast Abstractive Summarization with Reinforce-Selected Sentence Rewriting" Author: Samuel Humeau, Kurt Shuster, Marie-Anne Lachaux, Jason Weston Published Year: 2020

[12]Title: "Query-Focused Summarization with User Feedback" Author: Nitish Shirish Keskar, Bryan McCann, Lav R. Varshney, Caiming Xiong, Richard Socher Published Year: 2019

[13]Title: "Reinforcement Learning for Bandit Neural Machine Translation with Simulated Human Feedback" Author: Marcin Junczys-Dowmunt, Roman Grundkiewicz, Tomasz Dwojak, Hieu Hoang, Kenneth Heafield, Tom Neckermann, Frank Seide, Ulrich Germann, Alham Fikri Aji, Nikolay Bogoychev, André F. T. Martins, Alexandra Birch Published Year: 2018

[14]Title: "Learning to Select Data for Transfer Learning with Bayesian Optimization" Author: Xuezhi Wang, Jiawei Han Published Year: 2020