"AI-Powered News Article Summarization for Efficient Information Consumption"

1ST Dr Jyothi A P, 2ND Shivaranjini C, 3RD Usha Shree B N

Ast prof dept of CSE, FET, Student. Dept of CSE, FET, Student. Dept of CSE, FETComputer Science and Engineering,
Ramaiah University of Applied Sciences, Bengaluru, Karnataka, India

Abstract: News summarising condenses a news report into a shorter version while retaining the crucial details. 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 summarised using a transformer-based model, and a user-friendly application is proposed in this work. Performance is measured using the ROUGE metric.

Index Terms - News summarization hybrid summarization; evaluation metrics; machine learning; natural language processing; problems; future prospects.

I. INTRODUCTION

The project's goal is to create a reliable system for news article summarizing that uses Transformer-based models and can manage a lot of pieces from different sources. The program will extract the article's main terms and plot summary. The goal is to offer concise summaries that faithfully communicate the theme of the original material. Using Natural Language Processing (NLP)[1] techniques like BERT (Bidirectional Encoder Representations from Transformers) and T5 (Text-to-Text Transfer Transformer)[2,3], news stories are automatically summarised. These models are initially improved on news-specific datasets after being trained on massive datasets to generate news- specific summaries. The system will be assessed, and the quality of the summaries will be gauged using ROUGE (Recall-Oriented Understudy for Gisting Evaluation)[6] a standard metric. This initiative has important practical implications because it can save news providers and viewers time and money. Additionally, it can assist in bridging the gap between individuals who face access or linguistic hurdles and need easy access to news articles. In general, using Transformer-based models to summarise news articles is an intricate but fascinating area of research that could change how consumers receive their news.

Calculations are used by machine learning (ML) to analyse information and make smart decisions based on what it has learned. ML's subfield of deep learning (DL) is known. While both can be categorised as forms of artificial intelligence, profound realising is what governs the most human-like aspect of artificial consciousness[12,13].

Research on the use of novel AI algorithms to enhance targeting and tracking for military applications is still scarce. Working with new research trends in AI, a robust response to cybersecurity concerns in the military must be developed[13]. One of the primary categories used in neural networks for image recognition and classification is the convolutional neural network. Convolutional neural networks are frequently utilised in a variety of applications, such as scene labelling, object detection, face recognition, and others.Farmers can make informed[14]decisions about the health of their plants because to the system's provision of weather forecasting data. Plant diseases are found via CNN. Systems for the Internet of Things (IoT) have evolved significantly in recent years, particularly with the use of machine learning (ML) and artificial intelligence (AI) tools. The provision of intricate and user-specific smart city services is mostly being supported by a large number of AIenabled IoT. devices.[15]. The ML and DL techniques that can be utilised to extract predictive and prescriptive insights from Big Data are explained in detail in Applied Learning techniques for Intelligent IoT. The availability of data processing and mining, analytics algorithms, platforms, and frameworks has increased the importance of converting raw data into information and pertinent knowledge.[16].

II.WHY NEWS SUMMARIZATION REQUIRED?

There is a need for news summaries because of the sheer volume of today's news coverage. The proliferation of digital media and online news outlets has resulted in an information overload. It takes time and effort to keep up with the news while taking care of daily responsibilities.

News summarising offers a remedy by condensing lengthy news reports into concise summaries focusing on the most relevant information. People can stay informed without devoting a lot of time to reading lengthy articles thanks to this.

Summaries are useful for folks with a limited attention span or who just want to get to know current events as quickly as possible. News summaries are useful in and of itself, but they also help make the news more accessible to people who may have trouble reading the full stories. More people can access and engage with news content byproviding more digestible summaries.

News summaries are essential because they help individuals keep informed without requiring them to devote excessive amounts of time or energy to the task.

III. AI IN NEWS SUMMARIZATION

In automated journalism, sometimes known as "robot journalism," natural language generation algorithms that are powered by artificial intelligence are used to create news stories, photographs, videos, and visualizations from data. Automated journalism is frequently referred to as "robot journalism." After that, we will be able to disseminate these pieces through various venues for automated journalism. Journalists now have access to AI as a tool that may be used when producing media. They are able to gather content, analyze data pools, write pieces, and send them out, all with the click of a button in what is becoming known as "automated journalism." On a very large scale, reports are generated by the use of algorithms.

IV. TRANSFORMER MODEL USED FOR NEWS SUMMARIZATION

The ability of transformers to produce summaries that are both more accurate and more applicable to real-world situations has fundamentally altered how the news is conveyed. A news summary is a condensed version of a news report that maintains the most critical elements while being shortened for readability. Traditional methods for news summaries frequently used histograms and rule-based approaches, both of which frequently led to inaccurate summaries. On the other hand, ever since the development of transformers, machine learning models have provided increasingly accurate summaries.

Transformers uses a technique known as attention in constructing a summary[4]. This strategy enables the model to focus on the most important parts of the text being summarised. Because of this attention mechanism, transformers are particularly skilled at providing concise summaries packed with relevant information.

When it comes to the process of summarising news, the architecture of a transformer often consists of an encoder decoder structure with a significant number of feedforward and self-attentional neural network layers. In order to generate a collection of representations that are not visible, the encoder component analyses each word and token that is included in the input news item. By employing a wide variety of levels of self-attention, the model can pay attention to different parts of the input text and determine the context-specific connections between the words.
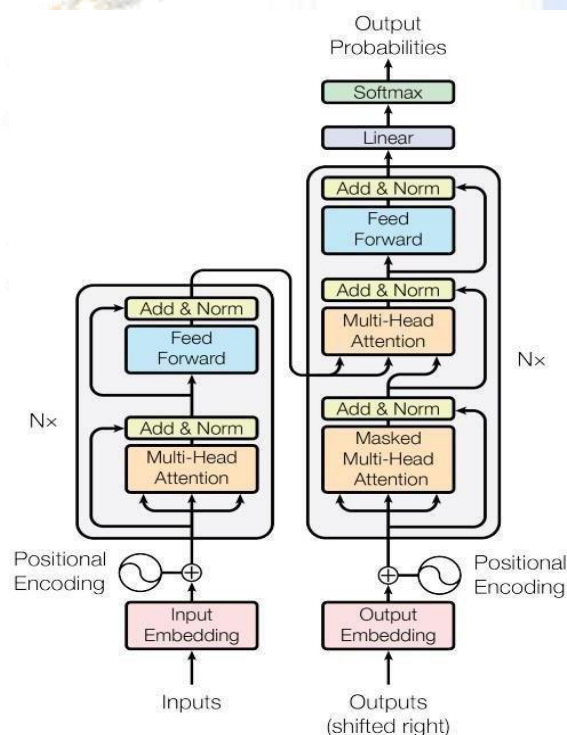

Fig : architecture of the transformer model

V. HUGGING FACE MODELS

Natural Language Processing (NLP)[3] is a popular library that provides easy access to state-of-the-art pre-trained models for various NLP activities. One of these NLP tasks is text summarization. Hugging Face offers easy access to these models[9]. The following is a selection of the text summary models that are accessible through Hugging Face:

Bart is a pre-trained sequence-to-sequence model that may be fine-tuned for text summarization. Bart was named after the fictional character Bart Simpson. It has achieved state-of-the-art performance on a wide variety of text summarization benchmarks and is based on an architecture called a transformer. T5: T5 is a pre-trained language model built on transformers that may be fine-tuned for

text summarization. Text summarization is just one of the many applications this highly adaptable model can have in the natural language processing (NLP) industry.

Pegasus is a transformer-based pre-trained model that was specially developed for text summarization. Pegasus was given the name "Pegasus." The CNN/Daily Mail dataset is one of the several text summarization benchmarks on which it has achieved state-of-the-art performance. For the aim of text summarization, the GPT-2 pre-trained language model can be adjusted. Google is responsible for creating GPT-2. It is a highly effective model that has surpassed the state-of-the-art on several natural language processing (NLP) tasks, such as text generation and information summarization.

VI. CHALLENGES IN NEWS SUMMARIZATION

Dealing with bias and subjective perception is an issue in news summaries. As a result, the summaries produced by various models or systems may differ depending on the individual's perspective on what is significant or pertinent in a news piece.

Making accessible, coherent, and understandable summaries is another difficult task. However, the quality of abstractive summarization has increased because, with transformer-based models, there is still potential for advancement in producing summaries as fluid and well-written as those written by humans.

Advantages of news summarization

● Time-saving: Summarized news articles can be read quickly and efficiently, allowing readers to stay upto-date with the latest news in less time.
● Concise: Summarized news articles are typically shorter than the original, making them easier to read and understand.
● Accessibility: Summarized news articles can be helpful for individuals with disabilities or who have difficulty reading long articles.
● Variety: Summarized news articles from different sources can provide a variety of perspectives on a particular topic.
● Efficiency: Machine learning models can automatically generate summarized news articles, allowing for the efficient processing of large volumes of text.

Disadvantages of news summarization

● Bias: The selection of essential information for a summary may be subjective and influenced by the bias of the summarizer or the model used.
● Inaccuracy: The summarization process may result in important information being included or misinterpreted, leading to inaccuracies in the summary.
● Lack of context: Summarized news articles may need to provide more context for readers to understand the topic or issue being discussed fully.
● Loss of detail: Summarized news articles may omit important details or nuances that are present in the original article.
● Quality: Summarized news articles may be lower quality than the original article, particularly in writing style and language.

VII. Future of News Summarization

● **Personalization:** As AI technology advances, news summaries could become more personalized, tailored to the reader's interests, preferences, and reading level. This could help readers stay informed on the topics that matter most to them while making the information more accessible and engaging.
● **Multimodal summaries:** To make summaries more engaging, they could include multimedia elements like movies, photos, and audio samples. This has the potential to facilitate new forms of reader engagement and comprehension of challenging news items.
● **Real-time updates:** With the increasing speed of news dissemination, news summarization could evolve to provide real-time updates on breaking news stories. This could help readers stay informed of developing events without reading lengthy articles.
● **Fact-checking:** With the rise of misinformation and fake news, news summarization could incorporate factchecking features to verify the accuracy of the information provided. This could help readers discern between reliable and unreliable sources of information, promoting greater transparency and accountability in journalism.

Overall, the future of news summarization looks bright as new technologies and innovations continue to transform how we consume and engage with news content.

VIII. Suggestion to Future Work

The quality of news summaries can be improved by the development of more complex models that can better understand the content and context of the text, the reduction of bias, the promotion of readability, the allowance of customisation, and the improvement of assessment measures.

One method to improve the precision of news summaries is to develop models that can better understand the context and meaning of the original content. This can be achieved by using deeper-learning techniques on larger training datasets. Providing users with the flexibility to tailor their news summaries has been shown to increase both satisfaction and participation. Options for the level of detail and tone of the summary, as well as the selection of news categories of interest, may be provided to the customer.

IX. Conclusion

Summaries of the news can be a useful tool for quickly and efficiently processing massive amounts of data. However, there are serious drawbacks to relying just on summaries of news items, such as bias, inaccuracy, and a lack of background information. Whether you read the news in its entirety or in a simplified form, you should always use your critical thinking skills. The trend in news summarization is towards more accurate and effective summaries, with more customization and personalisation, to fulfil the needs of a wide audience. In recent years, news summaries have become increasingly valuable as consumers seek out more efficient means of consuming the news. Due to the rise of digital media and the abundance of news outlets, it is more difficult for the general public to keep up with current events in a timely and efficient manner. News summaries provide a simple and efficient solution to this problem by highlighting the most important points from each news articles. Improvements in deep learning and natural language processing have led to far better news summarization models. There are many types of summarising methods, such as extractive and abstract summarising, as well as summarising that is both domain-specific and specialised. Despite these advancements, there is still much work to be done to address issues like reducing prejudice and improving the accuracy and readability of summaries.

X. References

[1] McMillan-Major, A., Osei, S., Rodriguez, J.D., Ammanamanchi, P.S., Gehrmann, S. and Jernite, Y., 2021. Reusable templates and guides for documenting datasets and models for natural language processing and generation: A case study of the HuggingFace and GEM data and model cards. *arXiv preprint arXiv:2108.07374*.

[2] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin,, 2017. Attention is all you need. Advances in neural information processing systems, 30

[3] Jain, S.M. (2022). Hugging Face. In: Introduction to Transformers for NLP. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-8844-3_4

[4] Jain, S.M. (2022). Hugging Face. In: Introduction to Transformers for NLP. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-8844-3_4

[5] Liu, Y. and Lapata, M., 2019. Text summarization with pretrained encoders. arXiv preprint arXiv:1908.08345.

[6] Grail, Q., Perez, J. and Gaussier, E., 2021, April. Globalizing BERT-based transformer architectures for long document summarization. In Proceedings of the 16th conference of the European chapter of the Association for Computational Linguistics: Main volume (pp. 1792-1810).

[7] Hayashi, T., Yoshimura, T., Inuzuka, M., Kuroyanagi, I. and Segawa, O., 2021, August. Spontaneous Speech Summarization: Transformers All The Way Through. In 2021 29th European Signal Processing Conference (EUSIPCO) (pp. 456-460). IEEE.

[8] Subramanian, S., Li, R., Pilault, J. and Pal, C., 2019. On extractive and abstractive neural document summarization with transformer language models. arXiv preprint arXiv:1909.03186.

[9] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M. and Davison, J., 2020, October. Transformers: State-of-the-art natural language processing. In Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations (pp. 38-45).

[10] Chouikhi, H. and Alsuhaibani, M., 2022. Deep Transformer Language Models for Arabic Text Summarization: A Comparison Study. Applied Sciences, 12(23), p.11944.

[11] Kalra, S., Pathak, A., Agarwal, A., Sharma, Y. and Chauhan, G.S., 2022. Comparative analysis of various text summarization techniques via leveraging transformer model for the fake news detection. In Applications of Machine Intelligence in Engineering (pp. 511-520). CRC Press.

[12] Jyothi, A.P., Usha, S. and Archana, H.R., 2022. Adoption of Machine/Deep Learning in Cloud With a Case Study on Discernment of Cervical Cancer. Advanced Analytics and Deep Learning Models, pp.65- 109.

[13] Jyothi, A.P., Shankar, A., Narayan, A.J., Bhavya, K., Reddy, S.S.M.S. and Yashwanth, A., 2022, December. AI Methodologies in Upcoming Modern Warfare Systems. In *2022 IEEE International Conference on Current Development in Engineering and Technology (CCET)* (pp. 1-7). IEEE.

[14] Pawar, S., Thorave, P., Panigrahi, N., Jyothi, A.P., Shedge, S. and Sayyad, S., 2022, October. Detection of Leaf Diseases for Multiple Plants using Convolutional Neural Network. In 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC) (pp. 1162-1168). IEEE.

[15] Jyothi, A.P. and Usha, S., 2022. Fog and edge computing paradigms for emergency vehicle movement in smart city. Streaming Analytics: Concepts, Architectures, Platforms, Use Cases and Applications, p.77.

[16] Sakthivel, U., Jyothi, A.P., Susila, N. and Sheela, T., 2021. Conspectus of k-Means Clustering Algorithm. In Applied Learning Algorithms for Intelligent IoT (pp. 193-213). Auerbach Publications.

Funtowicz, M. and Davison, J., 2020, October. Transformers: State-of-the-art natural language processing. In Proceedings of the 2020 conference on empirical methods in natural language processing: systemdemonstrations (pp. 38-45).

[17] Chouikhi, H. and Alsuhaibani, M., 2022. Deep Transformer Language Models for Arabic Text Summarization: A Comparison Study. Applied Sciences, 12(23), p.11944.

[18] Kalra, S., Pathak, A., Agarwal, A., Sharma, Y. and Chauhan, G.S., 2022. Comparative analysis of various text summarization techniques via leveraging transformer model for the fake news detection. In Applications of Machine Intelligence in Engineering (pp. 511-520). CRC Press.

[19] Jyothi, A.P., Usha, S. and Archana, H.R., 2022. Adoption of Machine/Deep Learning in Cloud With a Case Study on Discernment of Cervical Cancer. Advanced Analytics and Deep Learning Models, pp.65- 109.

[20] Jyothi, A.P., Shankar, A., Narayan, A.J., Bhavya, K., Reddy, S.S.M.S. and Yashwanth, A., 2022, December. AI Methodologies in Upcoming Modern Warfare Systems. In *2022 IEEE International Conference on Current Development in Engineering and Technology (CCET)* (pp. 1-7). IEEE.

[21] Pawar, S., Thorave, P., Panigrahi, N., Jyothi, A.P., Shedge, S. and Sayyad, S., 2022, October. Detection of Leaf Diseases for Multiple Plants using Convolutional Neural Network. In 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC) (pp. 1162-1168). IEEE.

[22] Jyothi, A.P. and Usha, S., 2022. Fog and edge computing paradigms for emergency vehicle movement in smart city. Streaming Analytics: Concepts, Architectures, Platforms, Use Cases and Applications, p.77.

[23] Sakthivel, U., Jyothi, A.P., Susila, N. and Sheela, T., 2021. Conspectus of k-Means Clustering Algorithm. In Applied Learning Algorithms for Intelligent IoT (pp. 193-213). Auerbach Publications.