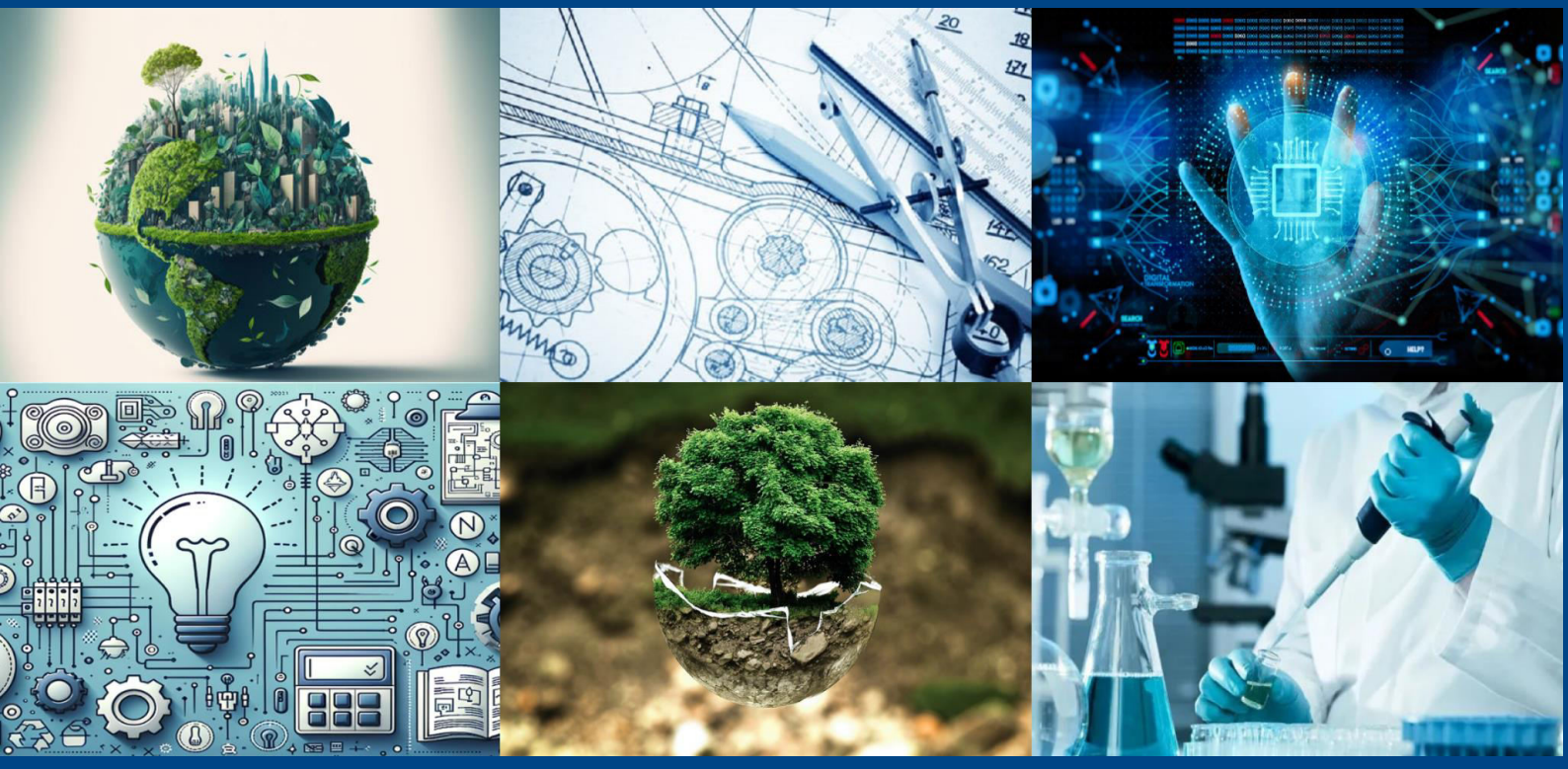
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

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



Impact Factor: 8.206 **Volume 8, Issue 5, May 2025**

© 2025 IJMRSET | Volume 8, Issue 5, May 2025| DOI:10.15680/IJMRSET.2025.0805056

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

Text Summary of Scientific Articles

Yugandhara Jagtap, Prof. Vishwatej Pisal

PG Student, Dept. of MCA, Anantrao Pawar College of Engineering and Research, Pune, India

Assistant Professor, Dept. of MCA, Anantrao Pawar College of Engineering and Research, Pune, India

ABSTRACT: Summarizing text is focuses on producing brief yet informative summaries from extensive text. Early methods used extractive techniques, picking important sentences from the original text, but they often sounded awkward. Machine learning, especially Transformer models have made text summarization much better text summarization has enabled abstractive summarization, which generates more natural and coherent summaries. While powerful, these models face challenges like computational intensity and occasional factual errors. This paper explores the evolution of summarization methods, comparing Extractive and Abstractive approaches and highlighting recent advances.

I. INTRODUCTION

Text summarization in NLP(Natural Language Processing) aims to condense lengthy texts into brief, As digital content grows, summarization helps create useful summaries. whether extractive or abstractive has become vital for improving information access, supporting decision-making, and saving time across fields like academia, media, and business..

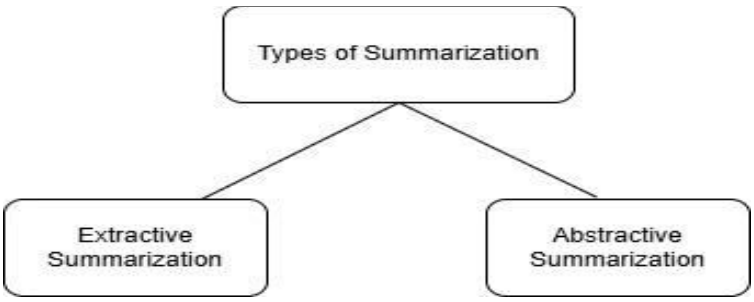


Figure 1: Summarizations Types

Text summarization has evolved from picking important sentences to creating new, shorter versions of the text using statistical methods like TF-IDF and graph-based algorithms to more sophisticated abstractive techniques. The latter generate new phrasings and aim for human-like coherence It makes

use of deep learning techniques like RNNs, LSTMs, and newer methods, Transformers such as GPT, BERT and PEGASUS . While abstractive methods offer improved fluency and contextual understanding, They require a lot of computing power. resources and may risk introducing inaccuracies.
Even with these challenges, improvements in NLP are making automatic summaries better and more reliable

II. REVIEW OF PREVIOUS WORK

Content summarization shortens long texts by keeping important information. It has two types: extractive, which picks key sentences, and abstractive, which creates new phrases to highlight the essential ideas

Extractive Summarization
Extractive summarization means picking important Exact lines taken from the source to make a summary. This pproach relies on identifying content with high relevance, often through metrics such as word frequency, sentence position, and thematic similarity. Algorithms like TextRank shows sentences as points in a graph, ranking them asedon interconnectivity to identify the most informative ones. Additionally, machine learning models such We can train SVM

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

and neural networks to do the task and find important by examining sentence features . While extractive summaries preserve original phrasing, they may lack cohesion because of the direct sentence extraction method.
Content summarization shortens long texts by keeping important information. It has two types: extractive, which picks key sentences, and abstractive, which creates new phrases to highlight the essential ideas

Extractive Summarization
Extractive summarization means picking important Exact lines taken from the source to make a summary. This approach relies on identifying content with high relevance, often through metrics such as word frequency, sentence position, and thematic similarity. Algorithms like TextRank shows sentences as points in a graph, ranking them based on interconnectivity to identify the most informative ones. Additionally, machine learning models such We can train SVM and neural networks to do the task and find important by examining sentence features . While extractive summaries preserve original phrasing, they may lack cohesion because of the direct sentence extraction method.

Abstractive Summarization
Abstractive summarization shortens a text by creating new sentences that capture the main points express the main ideas, rather than directly pulling out phrases. This approach often uses encoder-decoder architectures, where the encoder processes the data and the decoder creates the summary. Transformer models like GPT, BERT and PEGASUS enhance this process By using attention mechanisms, the model focuses on important parts, making the summary more accurate and clear. While powerful for summarizing complex texts like research papers and abstractive summary can sometimes make mistakes by creating incorrect information.

III. DATASETS FOR SUMMARIZING SCIENTIFIC ARTICLES

Data collections are important for training and testing summarization models because they provide examples of complex research papers. PubMed and PMC open-access dataset, managed by NCBI, is a key resource with full-text biomedical texts and their summaries. It's helpful for extractive and abstractive summarization. Another helpful dataset includes ArXiv and PubMed collection, which includes full-text articles and abstracts from various scientific fields. This dataset contains challenging because of the different Writing approaches and technical content, making it great for training summarization models.

IV. TRANSFORMER MODELS USED TO SUMMARIZE SCIENTIFIC TEXTS

In language processing, transformers are important deep learning models. They can understand long-range Connections between words and sentences, helping with tasks like translation, summarizing, and question-answering. The first step for transformers is to convert the input sentence into a series of vectors using a technique known as self-attention. This helps the model understand the relationships between words. After encoding the input, The model converts it into an output sequence, also using self-attention.

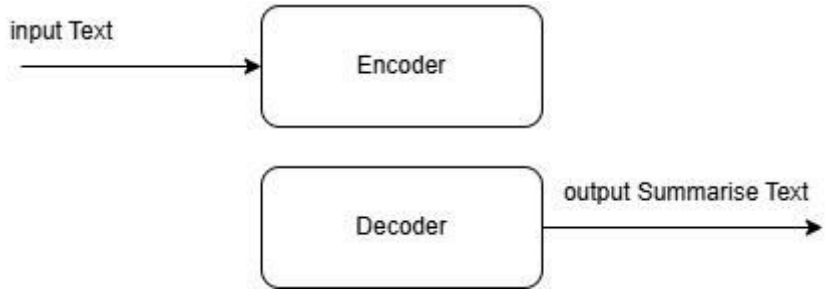


Figure 2: Transformer Architecture

BART
BERT (a type of transformer model) is a powerful language model created by Google. Unlike previous models, BERT processes text in both directions, understanding the context on both sides of a word, enhancing its understanding of

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

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

language complexity. It uses masked language modeling to predict missing words in a sentence, and next-sentence prediction BERT works well in tasks like analyzing sentiment and answering questions because it looks at both directions of a sentence to understand the relationship between two sentences, as it can better grasp contextual nuances and sentence flow. The model's pre-trained versions can be adjusted for specific NLP tasks

PEGASUS

Pegasus, created by Google Research, is a special NLP model for abstractive text summarization. It introduces gap-sentence generation during pre-training, where entire sentences are masked, and the model reconstructs them, mimicking the summarization process focuses on important points. It is trained using datasets such as C4 and Huge News, it excels in various summarization tasks, including news, legal, and academic article compression. Pegasus outperforms general models like BERT and T5 in summarization benchmarks and is highly adaptable It can be found on platforms like Hugging Face for easy use or customization

V. SUMMARY EVALUATION METHOD

Rouge is a common method for evaluating machine-generated summaries by comparing them to human-written reference summaries. It measures the overlap of n-grams or phrases between the generated summary and reference summaries. Key variations include ROUGE-N (for n-gram overlaps), ROUGE-L (which evaluates the longest common subsequence for fluency), and ROUGE-W (which emphasizes longer matches). The ROUGE metric focuses on recall helps capture important information from the reference summary and calculates precision and F1 measure for full evaluation. Its ease and efficiency have made it a standard for summarization Jobs in language processing

VI. RESULT

BART and **PEGASUS** have proven effective for abstractive text summarization, generating high-quality summaries with significant similarity to human-written content. The ROUGE metric is commonly used To measure how well automatic summarization systems work. Both the PubMed and arXiv datasets are frequently employed in such tasks, with BART and PEGASUS achieving comparable results. However, PubMed tends to yield slightly higher scores when using ready-made models trained.

VII. CONCLUSION

This study looks at using transformer models to summarize scientific articles, with a focus on arXiv and PubMed datasets. BART is highlighted for its strong performance, excelling in handling lengthy texts and achieving high ROUGE scores. Future studies can look at other deep learning models and optimization methods and preprocessing methods to enhance performance. The findings underscore deep learning can do a lot to improve scientific text summarization.

VIII. ACKNOWLEDGEMENTS

I sincerely thank my guide, Prof. Vishwatej Pisal, for their support, guidance, and encouragement throughout my research. Their help and guidance have been crucial in building this work, and I truly appreciate their dedication to my achievement.

REFERENCES

1. Sung-Guk Jo, Seung-Hyeok Park, Jeong-Jae Kim, and Byung-Won On ”Learning Cluster Patterns for Abstractive Summarization”, @ IEEE 2023
2. Wenfeng Liu, Yaling Gao, Jinming Li, and Yuzhen Yang ”A Combined Extractive with Abstractive Model for Summarization”, @ IEEE 2021
3. Mandar Bakshi, Ashish Tak, Omkar Tendolkar, Aayush Yadav, Prof. Neelam Phadnis on” Quick Reads- Text Summarization For News And Science Articles” Volume:05/Issue:04/April-2023
4. Heewon Jang, and Wooju Kim ”Reinforced Abstractive Text Summarization with Semantic Added Reward”, @IEEE 2021
5. Nitte, Udupi, on “Abstractive Text Summarization”,@IJEMETS 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com