# AUTOMATIC TEXT SUMMARIZATION OF EDUCATIONAL BOOKS BY FINE-TUNING OF ABSTRACTIVE MODELS

A blue and white logo

Description automatically generated

**By**

## NABILA ANUM

**Ms220400077**

**DEPARTMENT OF COMPUTER SCIENCE**

**FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY  
VIRTUAL UNIVERSITY OF PAKISTAN  
ISLAMABAD, PAKISTAN**

**202****5**

**AUTOMATIC TEXT SUMMARIZATION OF EDUCATIONAL BOOKS BY FINE-TUNING OF ABSTRACTIVE MODELS**

**BY**

**NABILA ANUM**

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF REQUIREMENTS FOR THE DEGREE**

**OF**

**MASTER OF SCIENCE   
IN  
COMPUTER SCIENCE**

**FACULTY OF COMPUTER SCIENCE & INFORMATION TECHNOLOGY  
VIRTUAL UNIVERSITY OF PAKISTAN  
ISAMABAD, PAKISTAN**

**2025**

# DECLARATION

I hereby declare that the contents of the thesis "Automatic Text Summarization of Educational Books by Fine-Tuning of Abstractive Models" are the creation of my research and no part has been copied from any published source (except the references, standard mathematical or geometrical models/equations/formulae/protocols, etc.). I further declare that this work has not been submitted for the award of any other diploma/degree. The University may take action if the information provided is found inaccurate at any stage. (In case of default, the scholar will be proceeded against as per HEC plagiarism policy).

****

**\_\_\_\_\_\_\_\_\_\_\_\_\_**

NABILA ANUM

(ms220400077)

**To**

**The Controller of Examinations,**

**Virtual University of Pakistan,**

**Islamabad.**

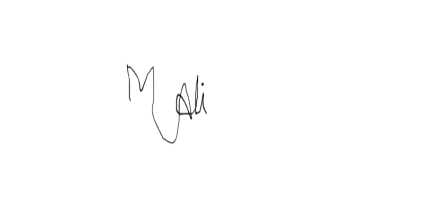
We, the supervisory committee, certify that the contents and forms of the thesis submitted by **NABILA ANUM**, VUID **ms220400077** have been found satisfactory and recommend that it be processed for evaluation by the External Examiner(s) for the award of degree.

**SUPERVISORY COMMITTEE:**

1. Dr. Mushtaq Hussain (Supervisor) Signature:

****

1. Dr. Said Nabi Signature:

****

1. Mehboob Ali Signature:

# DEDICATION

I would like to dedicate my thesis

To

My dear parents and my beloved elder brother.

# ACKNOWLEDGEMENTS

In the name of Allah, the most beneficent and the most merciful; I would like to thank my thesis supervisor, Dr. Mushtaq Hussain for guiding and assisting me with my thesis. I am also thankful to him for being very helpful and providing me with valuable ideas and crucial technical information that was required during the course of my thesis.

Finally, I am thankful to my family and friends for their great support, encouragement during tough times, and guidance during all the time for completion of my thesis.

Nabila Anum

## TABLE OF CONTENTS

[TITLE I](#_Toc24414)

[DECLARATION III](#_Toc31698)

[DEDICATION V](#_Toc30758)

[ACKNOWLEDGEMENTS VI](#_Toc22768)

[TABLE OF CONTENTS VII](#_Toc19118)

[LIST OF FIGURES X](#_Toc14544)

[LIST OF TABLES XI](#_Toc25927)

[LIST OF ABBREVIATIONS XII](#_Toc30841)

[ABSTRACT XIII](#_Toc15757)

[CHAPTER 1 1](#_Toc22787)

[INTRODUCTION 1](#_Toc21115)

[1.1 Text Summarization System 2](#_Toc7863)

[1.2 Significance of Text Summarization System 6](#_Toc31624)

[1.3 Innovation of the Current Study 6](#_Toc7736)

[1.4 Research questions 7](#_Toc20906)

[1.5 Research Contributions 7](#_Toc20483)

[1.6 Thesis outline 8](#_Toc2400)

[CHAPTER 2 10](#_Toc1212)

[REVIEW OF LITERATURE 10](#_Toc3646)

[CHAPTER 3 27](#_Toc26920)

[MATERIALS AND METHODS 27](#_Toc29574)

[3.1 Data Description 27](#_Toc16075)

[3.2 Proposed Framework 27](#_Toc10959)

[3.2.1 Data Collection 28](#_Toc5749)

[3.2.2 Data Preprocessing 29](#_Toc14557)

[3.2.2.1 Tokenization: 29](#_Toc1099)

[3.2.2.2 Stop words Removal: 29](#_Toc8704)

[3.2.2.3 Word Frequency Calculation: 29](#_Toc16657)

[3.2.3 Word Frequency Algorithm 29](#_Toc21184)

[3.2.3.1 Tokenization of sentences 29](#_Toc2747)

[3.2.3.2 Sentence Scoring 30](#_Toc11940)

[3.2.3.3 Retrieving top score sentences 30](#_Toc26255)

[3.2.3.4 Formation of Summary 30](#_Toc32327)

[3.2.4 Model Training 30](#_Toc14694)

[3.2.4.1 Data Preparation: 30](#_Toc3778)

[3.2.4.2 Model Initialization: 30](#_Toc10950)

[3.2.4.3 Fine-tuning: 31](#_Toc28124)

[3.2.4.4 Optimization: 31](#_Toc14375)

[3.2.5 Model Evaluation 31](#_Toc21459)

[3.2.5.1 ROUGE 31](#_Toc25762)

[3.2.6 Model Deployment 32](#_Toc2664)

[3.2.7 Deep Learning and NLP Techniques 32](#_Toc12432)

[3.2.7.1 Word Frequency Algorithm 32](#_Toc15130)

[3.2.7.2 T5 model 33](#_Toc6697)

[3.2.7.3 BART model 34](#_Toc24555)

[CHAPTER 4 35](#_Toc2311)

[RESULTS AND DISCUSSION 35](#_Toc29945)

[4.1 Visualization 35](#_Toc14333)

[4.2 Experimental setup 38](#_Toc25685)

[4.2 Results 39](#_Toc8532)

[4.4 Discussion 45](#_Toc32177)

[4.4.1 Model Performance Improvement Over Epochs 45](#_Toc25151)

[4.5 Deployment 47](#_Toc27799)

[CHAPTER 5 51](#_Toc30391)

[SUMMARY 51](#_Toc31061)

[5.1 Summary 51](#_Toc461)

[5.2 Thesis Contributions 51](#_Toc12010)

[5.2 Limitation of current study 52](#_Toc885)

[5.3 Future Work 53](#_Toc9184)

[LITERATURE CITED 54](#_Toc4966)

## LIST OF FIGURES

**FIGURE DESCRIPTION PAGE**

Figure 3.1 Framework of current study 28

Figure 3.2 Flowchart of Extractive Text Summarization 33

Figure 3.3 Architecture of T5 model 34

Figure 4.1 Graph of Word’s Frequencies in Dataset 37

Figure 4.2 Graph of Sentence Count Comparison of Original text vs Extractive Summary 38

Figure 4.3 Graph of Sentence’s Lengths in dataset 39

Figure 4.4 Graphical representation of results of T5 and BART 43

Figure 4.5 Graphical representation of Accuracy Distribution of T5

and BART 44

Figure 4.6 Graphical representation of results of Accuracy Confusion Matrix of T5 and BART 45

Figure 4.7 Text summarizer and important points generator app 49

Figure 4.8 Input text by the user 50

Figure 4.9 Output summary and highlighted important points by the app 51

## LIST OF TABLES

TABLE DESCRIPTION PAGE

Table 2.1 Summarized Literature Review 21

Table 4.1 Results of T5 and BART after fine-tuning for text summarization 42

## 

## LIST OF ABBREVIATIONS

NLP Natural Language Processing

WFA Word Frequency Algorithm

T5 Text-To-Text Transfer Transformer

BART Bidirectional and Auto-Regressive Transformers

ROUGE Recall-Oriented Understudy for Gisting Evaluation

LSA Latent Semantic Analysis

LDA Latent Dirichlet Allocation

seq-to-seq Sequence-to-Sequence

RNN Recurrent Neural Network

CNN Convolutional Neural Network

RoBERT Robustly Optimized BERT Approach

BLEU Bilingual Evaluation Understudy

SA model Sentiment Analysis Model

T5QG T5 Question Generation

Xsum Extreme Summarization

Ptr-Net Pointer Network

PGN Pointer Generator Network

RRF Review-related Features

ARF Aspect-related Features

HF Hybrid Feature Vector

CHIR Common Human Information Retrieval

SIM Semantic Information Metric

BBC British Broadcasting Corporation

EXABSUM Extractive and Abstractive Summarization Dataset

ATS Automatic Text Summarization

ISF Important Sentence Finder

SLR Systematic Literature Review

LSTM Long Short-Term Memory

# ABSTRACT

Text summarization is a process of getting useful and relevant information from the original text. There are two types of text summarization: Extractive summarization, which extracts some important sentences from the original text and the other is Abstractive summarization which requires a sophisticated intermediate representation of text. In this research, a hybrid approach is proposed for text summarization. Several challenges are associated with text summarization systems in educational institutions. Firstly, there is an under-representation of the research performed in this field. Secondly, most of the models that are used for text summarization have shown results with very low accuracy. These include either extractive techniques or transformer models. An efficient deep-learning technique is needed to predict the summary of educational books with enhanced accuracy. In the current study, we used the T5(Text-To-Text Transfer Transformer) and BART(Bidirectional and Auto-Regressive Transformers) models to predict the summary of educational books. To enhance the performance of these techniques we used extractive and abstractive models. The main purpose of this research is to fine-tune the T5 and Bart models that are used for text summarization. The dataset to be used for the implementation of hybrid models contains the topics from educational books. The results are evaluated through the well-known ROUGE (Recall-Oriented Understudy for Gisting Evaluation) metric. Our best model achieved R-1 score of 0.73, R-2 score of 0.63 and R-L score of 0.42, as the F score calculated by ROUGE is evident to be the best as compared to the already developed techniques shown in the literature review. The significance of this study is that it can turn large amounts of data into easy-to-use summary information, and it can aid students at higher institutions in comprehending complex concepts by providing concise and accessible explanations. Similarly, teachers can take advantage of summarized material by making notes from different sources for students or taking important highlights, which is helpful for students.

# CHAPTER 1

## INTRODUCTION

Text summarization leads to searching for important information from the source document and then converting this information into a precise format without updating the importance of the original text Zaware et al., (2021). Similarly, the summarization of textual data is the process of reducing the number of texts to retrieve the important information from the input text and presenting it to users Elbarougy et al., (2020). Automatic text summaries run summaries automatically. Hence, summarization plays a vital role and is the most efficient solution in the case of literature review, for getting the gist of information covered in newspapers, covering the maximum content from story books Zhao et al., (2022), or from course content materials for prior exam preparation. Similarly, while reading a newspaper, it is ultimately necessary to get the gist of important news in the form of outlines. Furthermore, while giving some useful information to students from books available in the school library, it is sometimes necessary to cover all important details in a limited time duration. Moreover, students at the college or university level may need to consult multiple books for a single course, and everyone wants to cover a maximum of the respective books for exam preparation. Furthermore, while telling some useful information to students from books available in the school library, It is sometimes necessary to cover all important details in a limited time duration.

Hence, summarization is a vital and most efficient solution in the case of a literature review, for getting the gist of information covered in a newspaper, covering the maximum content from story books or course content materials for prior exam preparation. Although summarization supports the solution of many text compilation-related problems. Still, some challenges must be sorted out in an efficient way to produce a precise summary. The challenges in automated text summarization include exponential growth of data, complexity of text structures, language, and domain-specific considerations, to improve the performance of deep learning models through integration of multiple data sources. Gunuputi et al., (2022) have also discussed challenges in text summarization, such as maintaining information content, ambiguity in natural language, and the need for coherent sentence construction. So, an advanced technique is needed to provide a solution to all these challenges.

## 1.1 Text Summarization System

In our current research, we intend to develop a text summarization system for producing summaries of a given text input by a user. Text summarization is a program or computational method created to shorten lengthy text content into more precise text without losing important information or main ideas. This system employs different natural language processing (NLP) methods, like extraction and abstraction, to produce summaries.

Moreover, the extractive approach to text summarization involves the selection of sentences and phrases from the original text to form a summary. It has multiple phases before it generates a summary that includes preprocessing (removal of stop words, stemmatization, tokenization), calculating scores for sentences, and then, last but not least, the generation of important sentences based on the calculated scores of these sentences.

While, on the other hand, the abstractive approach includes generating new sentences to convey the main ideas of the original text concisely and coherently. Unlike extractive methods, which select and combine existing sentences, abstractive methods aim to produce summaries that may not exist word-for-word in the original text. Abstractive summarization often employs encoder-decoder architectures, such as sequence-to-sequence (Seq2Seq) models, which consist of two main components: encoder and decoder. To implement an abstractive approach, the model is trained using suitable parameters, and then it is evaluated using Rouge(Recall-Oriented Understudy for Gisting Evaluation). With all this implementation, the challenges of text summarization are also addressed and are sorted out to achieve the required performance.

In this research, the primary problem being addressed is the efficient and effective summarization of educational texts to aid learners in understanding key concepts without reading the entire content. To achieve this, a hybrid approach is proposed, combining extractive summarization through a word frequency algorithm to retrieve important sentences, followed by abstractive summarization using fine-tuned T5 and BART models to generate coherent summaries. However, this approach presents several challenges that need to be addressed.

One of the first challenges is information overload and selection. In summary, it's crucial to select the most important information from the text, avoiding unnecessary details. The proposed solution in this research is the use of an extractive technique, the Word Frequency Algorithm. The word frequency algorithm (WFA) helps select important sentences based on word occurrence, which is an effective method for ensuring key points are included. However, this may sometimes prioritize common but less informative words. The system might struggle with content that has important low-frequency terms.

Context understanding presents another challenge. Abstractive summarization requires a model to understand the deeper context of the text, which can be difficult, especially in complex documents. The fine-tuning of T5 and BART models on the extracted summaries generated by the WFA provides a level of abstraction that can rephrase or combine information from the source text.

Similarly, coherence and fluency can be a challenge in the way of automated text summarization. Abstractive summarization requires a model to understand the deeper context of the text, which can be difficult, especially in complex documents. The fine-tuning of T5 and BART on the extracted summaries generated by the WFA provides a level of abstraction that can rephrase or combine information from the source text.

Likewise, redundant information in summarized text may be another major challenge. Extractive methods often result in repetitive or redundant information because of sentence selection based on word frequencies. The word frequency algorithm may introduce redundancy by selecting multiple sentences with overlapping information. However, the T5 model can reduce this by paraphrasing or condensing similar ideas.

Furthermore, the summarization of Long documents pose a challenge for models like T5, which have input size limitations. The WFA reduces the text size by selecting key sentences, making the task more manageable for the T5 model.

Moreover, the evaluation of produced summaries can be a big challenge. Assessing the quality of summaries can be complex, requiring alignment with human judgment. To deal with this challenge our proposed system includes a ROUGE score calculation to evaluate the quality of the generated summaries. ROUGE provides a way to compare the summaries against reference texts, ensuring some level of validation for the summarization quality.

Lastly, there is the issue of generalization and scalability. While the approach might work well for educational books, it may not generalize to other types of texts without substantial modification. Additionally, scaling the system to handle larger datasets or texts from diverse educational domains could be difficult. This problem is solved by the inclusion of some input text from newspapers and articles.

In this research, text summarization will be performed using an extractive abstractive approach. This involves using a technique called a Word Frequency Algorithm to extract information. A Word Frequency Algorithm retrieves important sentences by calculating their scores, and then these retrieved sentences are passed through a seq-to-seq transformer that is fine-tuned to improve its accuracy. In this research, the accuracy of the T5 and BART models will be improved using a novel technique called a hybrid approach.

Most of the literature on text summarization includes either the extractive or the abstractive approach. For instance, Rani et al., (2021) demonstrated the comparative analysis of different extractive methods for text summarization. These methods include tf-idf, text rank, LDA, etc. Similarly, Mandal et al., (2020) also performed text summarization using LSA(Latent Semantic Analysis) which lies under the category of extractive approach only. Other papers mostly depict the summarization through an abstractive approach. As, Wang et al., (2023) trained the T5 model, which lies under the abstractive approach. They have trained T5 for text summarization on multiple datasets. And the results represent below-average accuracy. To address the issue of low accuracy in text summarization, this research focuses on summarizing topics from educational books, which will be particularly beneficial for students in higher education. By condensing key content from multiple textbooks, the summaries will help students prepare for exams more efficiently. This research leverages sequence-to-sequence (seq-to-seq) transformer models to generate concise summaries of textbook topics, contributing to the improvement of educational outcomes. These summaries provide students with accessible explanations of complex concepts, enhancing comprehension and learning. Additionally, summaries generated by these models can guide educators in prioritizing content and designing instructional materials more effectively.

All these challenges are sorted out by the adoption of T5 and BART, as previous studies also support the adoption of T5 and BART for text summarization. Ahuir et al. (2022) identified BART, T5, and PEGASUS as the top-performing abstractive summarization models. Similarly, Goloviznina and Kotelnikov (2022) demonstrated that T5 and BART exhibit strong performance in text summarization tasks, further justifying their use in the current study.

## 1.2 Significance of Text Summarization System

Text summarization systems are utilized in different fields like news aggregation, document summarization, email summarization, and generating abstracts for research papers. These systems aid users in efficiently understanding the key points from a substantial amount of text, thus reducing the time and effort required for information processing. Most of the work done in the field has mostly focused on the text summarization of news articles, online textual data, and the data from the corpus from Kaggle or similar platforms. In this research, textual data from educational books is used as a dataset. By developing summarization models tailored to this educational context, the research aims to enhance students' understanding of complex concepts presented in textbooks.

Effective summarization can significantly improve learning efficiency by obtaining key concepts from lengthy texts. This research contributes to the creation of tools that enable students to study more efficiently and educators to deliver content more effectively.

Furthermore, by providing concise summaries of educational topics, the developed models can serve as valuable pedagogical aids for both students and teachers. They can help students grasp essential concepts more easily and assist educators in designing lesson plans and assessments.

## 1.3 Innovation of the Current Study

The novelty of this research work is the enhancement of performance or accuracy of T5 and BART models along with word frequency algorithm represents an innovative approach to text summarization in the educational domain. By adapting state-of-the-art pre-trained models to the specific characteristics of educational textbooks, the research pushes the boundaries of what is achievable in educational content summarization.

In this research, the accuracy of seq-to-seq models is improved using a novel technique that is a hybrid approach. Most of the literature on text summarization includes either the extractive or the abstractive approach. Moreover, Gurusamy et al., (2023) also used a hybrid approach in their research paper and their results showed that this mixture approach yielded better results as compared to the extractive approach alone. Moreover, the current study will also be a first step to asking questions and creating MCQs shortly, as stated in the text provided. It also provides an opportunity for further research in the literature review, which may include further development of the techniques used in this study.

## 1.4 Research questions

Q#1: How to implement a hybrid method for text summarization that combines the Word Frequency Algorithm with transformer models like T5 and BART?

Q#2: How to enhance the accuracy of abstractive models T5 and BART by fine-tuning on a self-created dataset?

## 1.5 Research Contributions

In the field of education, this has led to the rise of e-learning in many higher education institutions, as students are encouraged to access research and learning materials online rather than visiting physical libraries. This can lead to more opportunities for research and learning when students use a variety of online learning tools. This requires a solution that can help transform large amounts of data into user-friendly data summaries, as suggested by Karanja and Matheka (2022). Likewise, the materials on the internet change rapidly, and many repetitions or inconsistencies are encountered in this information. Hence, the need for summarizing and organizing written content becomes pressing and significant. Manual summarization proves to be costly and time-intensive. Text summarization emerges as a pivotal solution to address this challenge, as highlighted by El-Kassas et al., (2021). Current research provides a solution to all these challenges. To deal with these challenges following main contributions in my research work are given.

The research tailors summarization models to the unique vocabulary, syntax, and content structure of textbooks by using a dataset from the educational field.

The research introduces novel evaluation metrics tailored to educational text summarization, assessing the models' ability to capture key concepts and maintain coherence, providing valuable performance insights.

Advancements in NLP (Natural Language Processing) summarization techniques to make it effective for educational purposes.

Educators can use the proposed system to create concise notes for students while seq-to-seq model summaries help prioritize content and instructional design. It can be used in the enhancement of learning outcomes.

Moreover, the current research will also serve as the initial step, shortly, for question generation and the generation of MCQs as well from the given text. It paves a path towards further research in text summarization that may include further refinement of the techniques used in this research. By using a hybrid approach, fine-tuning with enhanced accuracy is achieved.

## 1.6 Thesis outline

The thesis consists of the following chapters:

Chapter 1:

The first chapter focuses on the problem statement, Definition and types of Summarization, the significance of Automated Text Summarization, and the innovation of the current study.

Chapter 2:

The second chapter consists of a literature review. Primarily focuses on the text summarization techniques and algorithms that have been implemented in the past.

Chapter 3:

The third chapter consists of material and methods. It contains different techniques to be used in the implementation of the project.

Chapter 4:

This chapter is about the system implementation of techniques in the project and the results produced after implementation and execution.

Chapter 5:

The sixth chapter includes the conclusion of the project and the future work that needs to be done related to the current study.

Chapter 6:

The last chapter includes references that are related to the citations in the text.

# CHAPTER 2

## REVIEW OF LITERATURE

Text summarizing has been the subject of much research utilizing a variety of methods and strategies. Various methods have been used in previous research to investigate the efficacy of summarization models on a variety of datasets with a wide range of input attributes. Previous research papers have investigated strategies for extracting and producing summaries from published content, such as newspapers and articles, to enhance learning results. The following highlights some of the pertinent work being done on the subject of automatic text summarization, especially about improving abstractive models like T5 and BART:

An extractive or abstractive strategy is used to accomplish work in the context of text summarization, according to a survey of the literature on the subject.

In a similar vein, Zhao et al., (2022) have introduced a unique approach that uses prior summarization to produce questions derived from narrative books. Three modules make up the methodology that is described: instructional question creation, event-centric summary generation, and question-type distribution learning. This research uses a double-fine-tuning technique for BART. The first time was for event-centric summarization, and the second time was for questions to be automatically generated from the summarized text. By comparing the produced questions from the proposed model with the questions from the Fairytale QA dataset, the effectiveness of the suggested mechanism is assessed. On F1, the computed ROUGE L is 41.78/38.29.

Etemad et al. (2021) provided similar details regarding previously utilized models, such as RNN, CNN, etc. However, these are being replaced by more modern transformers, such as BERT and other seq to seq and text to text transformers, because of problems including gradient disappearing, exploding, long-term dependency, parallelization, and being computationally restricted.

Furthermore, Deokar & Shah (2021) have compared two abstractive transformer models, t5 and Bert. They have made use of the BBC News Summary dataset, also available on Kaggle. This dataset has news articles grouped into 5 classes: 'business', 'entertainment', 'politics', 'sport', and 'tech'. There are also summaries provided for each article in each class. The authors have mentioned in the newspaper that they have also introduced a method of scraping news articles. The accuracy of BART and T5 is compared using ROUGE F1 scores which are 33 % and 26%, respectively.

Similarly, Karanja & Matheka (2022) have described two approaches to text summarization: extractive and abstractive. To have good precision and recall, the authors have implemented a hybrid approach for text summarization. This hybrid approach means that the test is first summarized through an extractive approach, and then the output of this will act as the input for abstractive summarization. The improved scores of precisions and recall are shown in tabular and graphical form. And the accuracy of their proposed approach is given as F1 SCORE: ROUGE 1=0.23, ROUGE2=0.04, ROUGE=0.14.

Moreover, Guoy et al., (2022) have presented a new Transformer-based neural model called LongT5, with which the effects of scaling both input length and model size can be explored at the same time. Through experimentation in various complex summarization and question-answering datasets, the authors have explored the performance gains that can be achieved by scaling both input length and model size, resulting in state-of-the-art results on multiple datasets. The results to show the accuracy are computed on arXiv, PubMed, BigPatent, MediaSum, and TriviaQA. The main conclusion on the result is LongT5 (xl - 16k input) =48.35, 21.92 and 44.27.

In the same vein, the authors Mengi & Kakade (2023) have used a pre-trained model proposed by Colin Raffel which is fine-tuned in the Xsum and Gigaword datasets and produces state-of-the-art performance and abstractive text summarization. Xsum is a collection of BBC articles with a summary written by the author of the article. The dataset consists of a huge collection of BBC News articles, Amazon product reviews, IMDB movie reviews, and English sentiment data, amounting to a significant 2.5GB in size and spanning several million rows. The data set is first prepared for training and passed through multiple stages of preprocessing which are cleaning, lemmatization, and appoint title. The researchers have tried to achieve an improvement in the summarization model by increasing ROUGE from 13.6168 to 21.4997. Conversely, the RoBERT model excelled in sentiment analysis, achieving a remarkable accuracy of 94.71% and an F1 score of 94.60%.

Additionally, the use of the T5 model is observed by Kumar et al., (2022). This research needs a fine-tuning of the t5 model on a small dataset, and then the fine-tuning of T5-eQA is needed to generate the precise questions. The accuracy is computed by BLEU and ROUGE, which are given by T5QG =40.96, 17.54, 19.21, 42.36 and Contrastive\_T5QG = 42.04, 19.11, 20.07, 48.50.

Similarly, to shed light on text summarization, Kaur & Sharma (2023), have proposed a novel framework called the SA model implementing deep learning for CRS (SADL-CRS) based on the effective approach of SA and review summarization (RS). There are three stages mentioned in this paper: the input stage, the tokenization stage, and the summarization section. In the Input section, data is pre-processed, whereas, in the tokenization section, the pre-processed data is tokenized, marking the review as positive, negative, or neutral. Moreover, in the last summarization section, the final summary of the given review is retrieved. The performance is evaluated on a ROUGE score concerning multiple datasets, for example, the SemEval-2014 dataset and the Sentiment140 dataset. The accuracy is determined as ROUGE L OF F1 CRS 18.71.

Additionally, to achieve efficiency in text summarization, Liu et al., (2021) have presented a novel framework. The proposed model, Refactor, can be used either as a base system or a meta-system, effectively coping with the learning gaps introduced in two-stage learning. The main techniques BART, GSum, and PEGASUS are used as the base systems, keeping in view their capability of achieving improved performance on at least one dataset. The datasets that are specifically used in this research paper are Xsum, Pubmed, and wikiHow.And the measure used to evaluate the performance is Rouge. The maximum accuracy achieved on the Pubmed dataset by fine-tuning the model is 43.72.

Moreover, the analysis revealed several unexpected trends in the text summarization. Rehman et al., (2023) have presented the analysis of three pretrained models named as google/pegasus-cnn-dailymail, T5-base, facebook/bart-largecnn on CNN-dailymail on three different datasets that are Samsum, Billsum and CNN-dailymail.The highest accuracy in the form of f1 score is achieved by BART model on Billsum dataset that is 42.84.

Furthermore, the authors Merrouni et al., (2023) have focused on the challenges that persist in the development phases of text summarization. These include the detection of text relevancy, identification of redundancy and coherency, and the third challenge pertains to abstractive and hybrid summarization. In this paper, they introduce EXABSUM, an ATS SYSTEM equipped to generate two distinct summary categories that follow an extractive and abstractive approach, respectively. EXABSUM Extractive includes multiple preprocessing phases such as lemmatization, calculation of TF/IDF for each term, and calculation of the relative position of sentences in the summary. Similarly, EXABSUM Abstractive includes substages that are Word graph generation, path filtering, re-ranking using key phrases, and then finally the abstractive summary generation. The maximum calculated Rouge score using scoring techniques ISF, sentence's length, and sentence's resemblance to the title is 0.493 on the DUC2002 dataset.

Text summarization is an interesting research topic in the NLP community because it helps produce concise information. For new research to be conducted, it is very important to have advance knowledge of the topic on which research is to be conducted, along with a review of previous work done on that topic. For this purpose, Widyassaria et al., (2022) have presented the latest research and progress in the field of text summarization through the systematic literature review (SLR) method. The SLR method is proven to provide up-to-date and well-equipped knowledge about the use of Text summarization techniques, datasets, preprocessing techniques, feature extraction techniques, and the problems and methods of evaluation that were used in previous research regarding text summarization. It also paves the path for future research.

To highlight the importance of text summarization in English and other text mining courses of computer science, Abidin et al., (2022) have implemented an extractive approach that is TF-IDF with the achievement of a recall value of 0.675, precision value of 1.00 and f-measure test value of 0.8059 which indicates that the summarizing system has vital importance and can be used as a medium for learning English courses and text mining. A summary of a given text is generated by passing from multiple sequential stages that include counting the words in the text, the removal of stop words, the stemming process, the calculation of terms and document frequencies, and then the formulation of a summary using these frequencies. Evaluation of the used technique is performed by making a comparison between a system-generated summary and an online summary using some online tool, for example, tool noob.

Text summarization can be achieved through extractive approaches. One of these approaches was mentioned by Christian et al., (2016). In their paper, the authors have implemented the TF-IDF method for summary generation. The generated summary by the program is compared with the human-generated summaries and the online tools. The accuracy achieved is calculated as a F1-score of 67%.

Moreover, financial documents and annual reports may contain some useful but lengthy information that needs to be summarized efficiently. To tackle this problem, an approach is developed by Litvak et al., (2020). The research aims to address the challenge of efficiently summarizing these documents by utilizing TF-IDF weighting of both single-word and multi-word expressions to identify key terms and sequences that capture the most important information in the documents. The F score is used to measure the accuracy of a newly developed approach, which is 0.433.

Similarly, to generate high-quality and information-rich summaries of numerically-oriented commodities reports, the transformer models of Natural Language Processing that are Bart and T5, have been fine-tuned by Zeng (2024). The goal is to assess how effectively these models can relay numerical information in summaries, with a focus on aiding farmers, producers, and small businesses in the agricultural industry. The accuracy achieved is Flan T5 vs Numerical = 56% Flan T5 vs Compression = 62% and Bert vs Numerical = 17.25% Bart vs compression=22%.

Moreover, Lubis et al.,(2023) have thrown light on the importance of text summarization by performing text summarization to enhance the performance of state-of-art model T5. In the Bayesian optimization task, they have used Bayesian probability theory for an iterative model so that it can have the advantage of updating initial knowledge. To evaluate the performance of T5, ROUGE was used, and the accuracy recorded in the paper was F SCORE ROUGE1=0.65, ROUGE2 = 0.67, ROUGEL= 0.69.

In the same vein, some others have used a hybrid approach, which means combining the characteristics of two or more abstractive models to achieve efficiency in text summarization. Chaurasia et al.,(2023) also proposed an innovative approach in which the encoder-decoder model was combined with a stacked LSTM layer with an attention mechanism and a T5 transformer model pre-processor. Therefore, a proper hybrid model, T5LSTM-RNN, is implemented to generate the summarized data. Hence, the recorded accuracy of this proposed technique is T5LSTM-RNN with Attention Mechanism = 0.2834, 0.1132, and 0.2629.

Furthermore, some others have used different NLP techniques to make a comparative analysis of these. The same kind of approach is followed by Gopikakrishna et al., (2021), where the authors have conducted multiple experiments to make comparative analyses of RNN and NLP for text summarization. Three factors were included in the comparative analysis: time, quality, and length. After performing preprocessing, the experiments were performed, and the required results were obtained. The results were observed according to the similarity index between system-generated and human-generated summaries. The average time of NLP on 112 words is 0.008 with a ratio of match with a human summary of 92%, and the average time of RNN on 69 words is 17.336 with a ratio of match with a human summary of 75%.

Moreover, a comparative analysis is performed in the text-to-text T5 model by Borah et al., (2022). In the research work, the authors have worked on multiple data types to take a look at the performance of T5. Bilingual Evaluation Understudy (BLEU) and Recall-Oriented Understudy for Gisting Evaluation (ROUGE) were used to measure the accuracy of

T5 model on different datasets. The highest scores observed in ROUGE 1 were 40.791, 42.287 and 35.063 on the datasets CNNDM, MSMO and XSUM respectively.

Another approach was tried by Tsonkov et al., (2021) where combined models of extractive abstractive were used for summarization. A naïve approach was used in which similarity between the conclusion part of the research papers and the abstract part was founded. And the best score that was recorded in the result portion was on ROUGE 1 that was 0.4182.

Some performances of the state-of-the-art models on a particular dataset were described by Aswani et al., (2024) where the definitions of these models along their uses were defined. The state-of-the-art models that were used in the research work include PEGASUS-Large, PEGASUS-X, BART-LS, Long T5, and BigBird. One of the important things that was observed in the research paper was that datasets only include the text content of the scientific article and ignore the mathematical equations, images, graphs, tables of results, and comparisons that hold the most important information about the scientific article. The highest performance was recorded by BART-LS, which was 50.3.

Karnik, M. P., (2023) utilized different seq- to-seq transformer-based models to have an outlook on the performance of the outperformed model. Four different models, namely Transformer, Fast Former, Transformer with Pointer Generator Networks and Coverage, and Fast Former with Pointer generator networks and coverage, were implemented in the paper presented here. Rouge was used to make a comparative analysis of all four models and the best performance was seen by the model Fastformer+ PGN + Coverage with scores mentioned as 47.22, 35.87, 31.17, and 51.06.

Moreover, Muia et al.,(2024) have demonstrated in their research the difficulty of summarizing scientific documents and highlight the need for further development in both extractive and abstractive summarization techniques. It also points out that simple, intuitive approaches can sometimes yield surprisingly effective results. This paper also contributes to the field by providing a detailed comparison of different summarization techniques on a dataset of scientific papers and by suggesting future research directions to improve summarization performance for long and complex texts​. According to the findings in this research paper, the highest ROUGE score is achieved by the T5 model which is ROUGE-1: 35.12, ROUGE-2: 22.75, ROUGE-L: 32.82, ROUGE-Lsum: 28.59​.

Likewise, Talukder et al., (2020) have provided a comprehensive overview of the research work conducted in the field of abstractive text summarization. The paper aims to compare and analyze different methodologies and approaches used in generating abstractive summaries from large text documents. By reviewing prominent works in the field, the paper seeks to highlight the strengths and weaknesses of various models, algorithms, and techniques employed in abstractive text summarization. The evaluation is performed using ROUGE and the highest achievement ROUGE-1 F-measure score of 0.4674 and a ROUGE-2 F-measure score of 0.2561.

Similarly, Odeyemi, G. (2025) has explored the effectiveness of the pause approach in encoder-decoder transformer models for automatic text summarization. In this paper,  the performances of transformer models using the pause token approach against the standard fine-tuning methods are compared. The optimal ROUGE scores were observed around 0.5631 for ROUGE-1 and similarly enhanced scores for ROUGE-2 and ROUGE-L, particularly when utilizing 30-50 pause tokens. While, Bart achieved a maximum ROUGE-1 of about 0.6501, with consistent improvements across datasets when using pause tokens, primarily peaking when around 30 tokens were used.

Furthermore, Peddarapu et al., (2025) focused on a comparative study of text summarization based on RNN and NLP-based techniques. This paper aims to assess the effectiveness of word embedding and explore their potential enhancement in PDF summarization. It draws insights from recent research on the evaluation and integration of word embedding for NLP tasks in document management.The evaluation in terms of ROUGE or any other matrix is not mentioned in this paper.

Fendji et al., (2025) explained how text summarization and user profiling work and introduced a model that combines both. In this paper, the system collects web pages based on a user’s search and creates short summaries that match the user’s interests and preferences. The maximum accuracy achieved in terms of ROUGE is 0.43.

Bharti et al., (2025) introduced a novel method for text summarization using Generative Adversarial Networks (GANs). This approach uses two components—a generator and a discriminator—to produce meaningful and coherent summaries. The maximum ROUGE score is 0.75. The model's performance was assessed using evaluation metrics like ROUGE and BLEU. However, since the generated summaries were compared to human-written ones, there is a possibility of bias in the evaluation.

Last but not least, Ravichandran et al., (2023) also described the effectiveness and applications of text summarizers and shed light on their potential to enhance efficiency, comprehension, and decision-making processes across various domains. The successful project focused on abstractive text summarization, introduced a system powered by the T5 transformer language model. The project highlighted the utility of abstractive summarization in automating data extraction and elevating decision-making processes.

Despite the growing interest in Text Summarization using an abstractive approach, there is a notable under-representation of studies focusing on hybrid approaches to text summarization that include both extractive and abstractive methods for summarizing text efficiently. The literature reveals a lack of consensus regarding the fine-tuning of T5 on a dataset that has already passed through some extractive technique, such as the word frequency algorithm or Term Frequency-Inverse Document Frequency (TF-IDF). Moreover, the scope of existing research is limited in its coverage of datasets from students' textbooks. Similarly, in previous studies, the performance of abstractive approaches was very low. It means that they produce irrelevant or lengthy summaries. These summaries may contain repeated sentences. In the same vein, some works recognize the value of personalized summarization, but models tailored to audience-specific complexity or educational levels remain largely unexplored. Moreover, transformer models like BART and T5 often struggle with long documents due to memory constraints, a limitation that is widely acknowledged but inadequately addressed in the literature. Lastly, most research is English-centric, lacking advancements in multilingual summarization or low-resource language support, and there is little progress on models that integrate external knowledge to improve contextual understanding in summaries. These gaps suggest opportunities for future research to develop more adaptable, interpretable, and contextually aware summarization models or a mixture of these.

Likewise, some of the related research work in the field of text summarization has shown in table 2.1. It depicts the purpose of the research with input features and achieved accuracy.

**Table 2.1: Table of Research Studies on Text Summarization using Natural Language Processing Techniques**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Ref. No.** | **Authors** | **Purpose** | **Model accuracy** | **Input Features** | **Year** | **Journal** |
| 1 | Abdul GhafoorEtemad, Ali Imam Abidi, &Megha Chhabra | Comparison of two abstractive transformer models that are t5 and Bert by the use of the BBC News Summary dataset. | FT5 on Gigaword daraset  ROUGE1=43.02  ROUGE2=14.50  ROUGEL=37.43  ROUGELSUM=37.49 FT5 on Xsum dataset ROUGE1=30.91  ROUGE2=5.26  ROUGEL=20.85  ROUGELSUM=20.85 | Textual content of the articles from the Xsum dataset, which consist of articles from the BBC along with their corresponding summaries | 2021 | International Journal of Performability Engineering, vol. 17, no. 10, October 2021, pp. 900-906, DOI: 10.23940/ijpe.21.10.p8.900906 |
| 2 | Varun Deokar&Kanishk Shah | Comparison of Bart and T5 model for text-summarization of news articles was performed and a method for the automatic scraping of articles from google news using just keywords related to the article was proposed. | BART F1 score of about  33% and T5 model average F1 score of  about 26%. | requests, bs4[5], and newspaper packages to  automatically scrape the web and store the most relevant  article without any of the ads or irrelevant links. | 2021 | International Research Journal of Engineering and Technology (IRJET), Volume: 08 , e-ISSN: 2395-0056 , p-ISSN: 2395-0072, www.irjet.net |
| 3 | James MugiKaranja& Abraham Matheka | The authors aim to address the increasing challenge of information overload by developing a hybrid model that combines extractive and abstractive summarization techniques. | F1 SCORE:ROUGE 1=0.232166667,ROUGE2=0.048533333, ROUGE=0.1416 | Text Content, Domain Information, Query, the structure of the document, including headers, figures, tables, and metadata, to aid in the analysis and summarization process, Language: | 2022 | Open Journal for Information Technology, 5(2), 65-80. ISSN (Online) 2620-0627 ▪ <https://doi.org/10.32591/coas.ojit.0502.03065k> |
| 4 | Mandy Guoy, Joshua Ainslie, David Uthus, Santiago OntañónJianmo Ni, Yun-Hsuan Sung &Yinfei Yang | The main goal is to explore the effects of this scaling on various natural language processing tasks, such as summarization and question answering. | LongT5 (xl - 16k input) =48.35 21.92 44.27 | Scalability, Transient Global Attention Mechanism LongT5 adopts the pre-training strategy from PEGASUS, | 2022 | International Conference on Learning Representations (ICLR) |
| 5 | Shobhan Kumar, Arun Chauhan &Arun Chauhan | to summarize the input text by fine-tuned t5 model and also questions from this summarize output were generated | T5QG =40.96, 17.54, 19.21, 42.36  Contrastive\_T5QG= 42.04, 19.11, 20.07, 48.50 | Transformer Model, Pointer Net (Ptr-Net) Multi-Head Self-Attention Mechanism, Pointer Generator Network (PGN), Coverage Mechanism | 2022 | Proceedings of the 19th International Conference on Natural Language Processing (ICON) |
| 6 | Gagandeep Kaur &Amit Sharma | This framework aims to enhance the efficiency of customer review summarization by leveraging sentiment analysis and review summarization techniques. | ROUGE L OF F1 CRS 19.31, 7.67, 18.71 | Review-related features (RRF), Aspect-related features (ARF), Hybrid feature vector (HF) | 2023 | International Journal of Electrical and Computer Engineering (IJECE) Vol. 14, No. 2, April 2024 |
| 7 | ZakariaeAlamiMerrouni , BouchraFrikh&BrahimOuhbi | The primary goal of the paper is to enhance the effectiveness of automatic text summarization (ATS) systems by introducing novel techniques for both extractive and abstractive summarization. | EXABSUMExtractive system achieved ROUGE-1 and ROUGE-2 scores of 0.601 and 0.451, respectively | Weighting Parameter (α)​​.  Scoring Techniques: Sentence Resemblance to Title​ | 2023 | Journal of Big Data |
| 8 | AdhikaPramitaWidyassaria,b, SupriadiRustad a, GuruhFajarShidika,EdiNoersasongko Abdul Syukur a, AffandyAffandy a & De Rosal Ignatius Moses Setiadi | The goal is to present a novel approach to Automatic Text Summarization (ATS) that can generate both extractive and abstractive summaries. | performance on the DUC 2001 dataset with ROUGE-1 and ROUGE-2= 0.480 and 0.208, respectively. For the DUC 2002 dataset, EXABSUMExtractive achieved ROUGE-1 and ROUGE-2 scores of 0.493 and 0.257, | Weighting parameter (α): weight assigned to both the statistical feature (CHIR) and the semantic feature (SIM) within the hybrid weighting model.  TR-ISF measure: | 2022 | Journal of King Saud University – Computer and Information Sciences 34 (2022) 1029–1046 |
| 9 | Aa ZezenZaenalAbidin, YuliMurdianingsih, UsepTatangSuryadi&DidikSetiyadi | The goal is to explore the use of text summarization techniques, specifically the TF-IDF method, to enhance students' interest in learning English, particularly through news articles. | The F-measure test value obtained for the system is 0.814 for testing by English teachers and 0.8059 for testing by online text summarizing systems | Text Data, Term Frequency (TF), Inverse Document Frequency (IDF) | 2020 | Journal of Critical Reviews ISSN- 2394-5125 Vol 7, Issue 5, 2020 |
| 10 | Hans Christian, MikhaelPramodanaAgus&DerwinSuhartono | an extractive text summarization with TF-IDF method is used to build the summary | F-measure=0.666 | sentence based on its importance and given the value between zero and one | 2016 | Article in ComTech Computer Mathematics and Engineering Applications. |
| 11 | Winston C Zeng | The aim is to enhance the summarization of documents containing a significant amount of numerical information while maintaining grammatical structure and word flow. | Flan T5.Numerical=56%compression=62%....Bert .Numerical=17.25%compression=22% | textual data, pre-trained embeddings, attention mechanisms, and encoder-decoder architectures | 2024 | http://hdl.handle.net/10150/668771 |
| 12 | Arif Ridho Lubis , Habibi Ramdani Safitri , Irvan , Muharman Lubis , Muhammad Luthfi Hamzah ,  Al-Khowarizmi Al-Khowarizmi , Okvi Nugroho | This study uses Bayesian optimization techniques that  performs a task in increasing the performance of the T5 model. | F SCORE ROUGE1=0.65, ROUGE2= 0.67, ROUGEL= 0.69 | It evaluates words that are not valuable and then  performs phrases from these words which will then be  included in the text summary. | 2023 | Revue d'Intelligence Artificielle |
| 13 | Shivangi Chaurasiaa, Debalay Dasguptab, Rajeshkhannan Regunathanc | In this proper hybrid model, T5LSTM-RNN, is implemented to generate the summarized data. | T5LSTM-RNN with Attention mechanism= 0.2834, 0.1132, 0.2629  Baseline Seq2Seq RNN =model 0.2541 ,0.1066, 0.2354 | LSTM layer with an attention mechanism and a T5 transformer model pre-processor | 2023 | International Conference on Machine Learning and Data Engineering |
| 14 | Todor Tsonkov, Gergana Lazarova, Valentin Zmiycharov, Ivan Koychev | The paper compares modern extractive and abstractive approaches and also proposes a naïve approach based on the idea that the conclusion part of a paper is often similar to its abstract. | T5:  ROUGE-1 (F1): 0.2649  ROUGE-2 (F1): 0.1020  ROUGE-3 (F1): 0.0622  Pegasus:  ROUGE-1 (F1): 0.2255  ROUGE-2 (F1): 0.1463  ROUGE-3 (F1): 0.0660 | BERT Embeddings: Clustering Algorithms, Encoder-Decoder Models: Google Pegasus: | 2021 | proceedings of the conference "Education and Research in the Information Society," held on September 27-28, 2021, |
| 15 | Seema Aswani  , Kabita Choudhary  , Sujala Shetty  , Nasheen Nur | to provide a comprehensive overview of transformer-based approaches used for text summarization of scientific research articles. The authors aim to highlight the significance of automatic summarization in aiding researchers in navigating through the vast amount of published articles efficiently. | For the arXiv dataset:  BART-LS: R-1: 50.2, R-2: 22.1, R-L: 45.4  Long T5: R-1: 48.3, R-2: 21.9, R-L: 44.2  For the PubMed dataset: BART-LS: R-1: 50.3, R-2: 24.3, R-L: 46.3  Long T5: R-1: 50.2, R-2: 24.7, R-L: 46.6 | Query (Q): Represents a matrix containing the query information.  Key (K): Contains the key, which is the vector representation of the input sequence in words.  Value (V): Represents the values associated with the input sequence. | 2024 | Journal of Autonomous Intelligence |
| 16 | Madhuri P. Karnik, Dr. D.V. Kodavade | Addressing the challenge of summarizing information effectively in order to extract essential knowledge from large volumes of data. | Transformer with PGN and Coverage:  ROUGE-1 (R-1): 41.74  ROUGE-2 (R-2): 29.32  ROUGE-3 (R-3): 23.77  ROUGE-L (R-L): 46.10  Fastformer with PGN and Coverage:  ROUGE-1 (R-1): 47.22  ROUGE-2 (R-2): 35.87  ROUGE-3 (R-3): 31.17  ROUGE-L (R-L): 51.06 | Text Corpus, Word Embeddings, Transformer Model, Pointer Generator Network (PGN), Coverage Mechanism, Self-Attention | 2023 | International Journal on Recent and Innovation Trends in Computing and Communication |
| 17 | Charles Munyao Muia  ,Aaron Mogeni Oirere ,Rachael Njeri Ndung’ | This research contributes to the field by providing a detailed comparison of different summarization techniques on a dataset of scientific papers and by suggesting future research directions to improve summarization performance for long and complex texts. | ROUGE-1: 35.12  ROUGE-2: 22.75  ROUGE-L: 32.82  ROUGE-Lsum: 28.59​ | pretrained models with encoder decoder attention layers | 2024 | International Journal of Advanced Trends in Computer Science and Engineering |
| 18 | Md Ashraful Islam Talukder, Sheikh Abujar, Abu Kaisar Mohammad Masum, Sharmin Akter, Syed Akhter Hossain | the paper aims to contribute to the understanding of the challenges and advancements in the area of text summarization, particularly focusing on the abstractive summarization process. | DUC-2002 dataset and achieved a ROUGE-1 F-measure score of 0.4674 and a ROUGE-2 F-measure score of 0.2561 | Applied techniques: Discourse rules, syntactic constraints, word graph, Input text portrayal: Word graph, utilized method, input text portrayal, utilized method for eliminating redundancy | 2020 | 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) |
| 19 | Ratan Ravichandran1, Sri Bharath Sharma P2, Shriyans Shriniwas Arkal3, Shubhangee Das4, | The purpose of the research presented in the PDF file "t5 tuner.pdf" is to explore the effectiveness and applications of text summarizers, by using the T5 transformer model in natural language processing. | ROUGE score: 0.23 on 10 epochs and 0.25 on 25 epochs | Data preprocessing, tokenization, training data and model architecture | 2023 | International Research Journal of Engineering and Technology (IRJET) |

# CHAPTER 3

# MATERIALS AND METHODS

In this chapter, the details of the techniques and models that are used during the implementation of the system are discussed. A brief overview of the proposed framework and the stages under the proposed framework is given. Last but not least, a description of the T5 and Bart models is given.

## 3.1 Data Description

The data used in this research for implementation purposes consists of educational content taken from student textbooks. This data was used to facilitate a detailed understanding of the model in the summarization task. The dataset, titled 'Automatic Text Summarization Dataset for Educational Books: Fine-tuning T5 and BART Models,' is publicly available in the Harvard Dataverse [Anum, Nabila, 2024, https://doi.org/10.7910/DVN/AQDTPF].

## 3.2 Proposed Framework

In this section, the details of the stepwise procedure of the proposed framework are given. The goal of this research is to deploy an app for text summarization and generation of important highlights from the given text, fine-tuning T5 and Bart models for working on text summarization tasks. There are multiple steps included in this framework.

The first step is the data collection. To collect the data from any dataset we must keep in mind that we want a diverse dataset. The diversity of the dataset makes the whole summarization process more effective and general, the dataset is first created by taking some texts from student's books, some articles from English newspapers, and some headlines from newspapers. Hence after the creation of the dataset, the dataset is then uploaded. The second step is data preprocessing, in this, the study will use several techniques including tokenization, removal of stop words, and calculation of word frequencies in the dataset to remove the unimportant details in the dataset. The third step is feature extraction, the study intended to use the Word frequency Algorithm to find out the most important sentences from the text. Next, the fourth step is data splitting and model training, in this step, the study will develop several ML-based predictor models that will predict the summary. The fifth step consists of model evaluation, in this step the developed models that are T5 and BART will be fine-tuned on the training data for text summarization and then the performance is observed in some performance measure that is ROUGE. Lastly, the sixth step includes the model deployment using the Streamlit app. The following diagram provides a detailed demonstration of the whole framework.

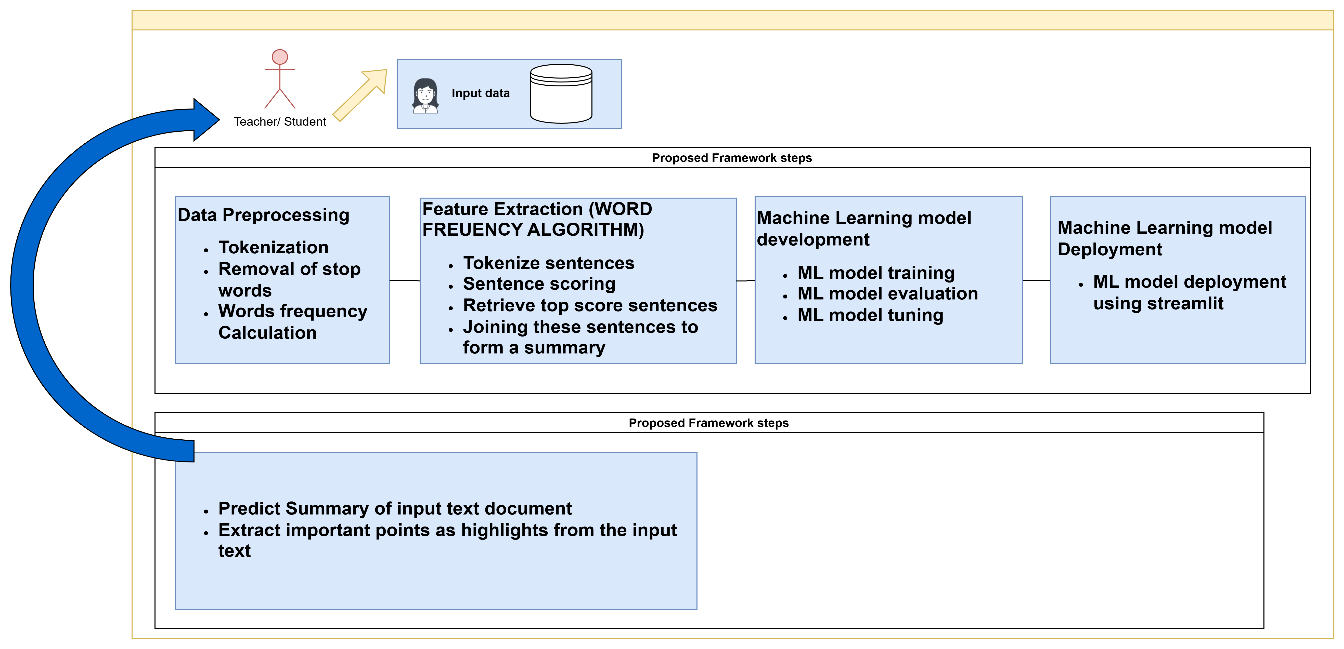


Figure 3.1: Framework of current study

**Figure 3.1: Framework of current study**

### 

### 3.2.1 Data Collection

The collection of accurate and reliable data is a very crucial task in any study. If the data is not accurate then the results of the study would be faulty and misleading. There is always a need for special attention in understanding the contents and structure of the dataset. The dataset, titled 'Automatic Text Summarization Dataset for Educational Books: Fine-tuning T5 and BART Models,' is publicly available in the Harvard Dataverse [Anum, Nabila, 2024, https://doi.org/10.7910/DVN/AQDTPF].

### 3.2.2 Data Preprocessing

The following preprocessing steps have been used during this stage: -

#### 3.2.2.1 Tokenization:

In this step, the splitting of text into words and sentences is performed. The text is tokenized into words using word\_tokenize and into sentences using sent\_tokenize.

#### 3.2.2.2 Stop words Removal:

In this step the common stop words of English are filtered out. Common English stop words are removed using the stop words corpus from NLTK.

#### 3.2.2.3 Word Frequency Calculation:

In this step the word frequencies are counted excluding stop words. A frequency distribution of the remaining words (after removing stop words) is calculated.

### 3.2.3 Word Frequency Algorithm

In the second phase after preprocessing, extractive text summarization is performed by applying the Word Frequency Algorithm using the following steps.

#### 3.2.3.1 Tokenization of sentences

Sentence Tokenization is the process of dividing a text into its constituent sentences. This is done using a tool or function that recognizes sentence boundaries, typically punctuation marks like periods, question marks, and exclamation marks. It allows the code to analyze and score each sentence individually based on the word frequencies calculated earlier.

#### 3.2.3.2 Sentence Scoring

In this step the frequency of the word that was calculated in preprocessing step is added to the score of the current sentence in the sentence\_scores dictionary. This means that sentences containing frequently occurring words will have higher scores.

#### 3.2.3.3 Retrieving top score sentences

The sentences having high scores will be retrieved in this step.

#### 3.2.3.4 Formation of Summary

This step will join the selected top sentences into a single string and returns it as the summary.

### 3.2.4 Model Training

#### In this phase, the model training is performed for the sake of implementation of abstractive summarization. Firstly, extractive summarization is performed using the Word Frequency Algorithm as mentioned above. Then, the models T5 and Bart are trained one after the other and fine-tuned for text summarization. Following are the sub-steps that are required in the training phase:

#### 3.2.4.1 Data Preparation:

The text data is first prepared by performing extractive summarization (selecting key sentences) and storing them in the dataset. Then, this data is split into training and testing sets.

#### 3.2.4.2 Model Initialization:

A pre-trained T5 model (specifically for text generation tasks) is loaded, along with its tokenizer. The tokenizer converts the text into token IDs that the model can understand.

#### 3.2.4.3 Fine-tuning:

The model is fine-tuned on the training dataset. In each epoch (a full pass over the training data), for each training example, the source text (input) and the target text (summary) are tokenized and passed to the model. The model generates predictions and computes a loss (a measure of error) between the predicted summary and the actual summary. The model's weights are adjusted using an optimizer to minimize this loss.

#### 3.2.4.4 Optimization:

After calculating the loss, the optimizer updates the model's parameters to improve performance in subsequent iterations. A learning rate scheduler is also used to gradually adjust the learning rate during training, which helps the model converge better.

### 3.2.5 Model Evaluation

#### After each epoch, the model's performance is evaluated on the test set. It generates summaries for unseen data, and the ROUGE metric is computed to compare the quality of generated summaries with the target summaries. In this phase, the generated summaries are compared to the reference summaries (target texts) using ROUGE scores. The average ROUGE-1, ROUGE-2, and ROUGE-L scores are computed and printed after each epoch to show how well the model is learning over time.

#### 3.2.5.1 ROUGE

ROUGE(Recall-Oriented Understudy for Gisting Evaluation) is a set of metrics commonly used to evaluate the quality of summaries produced by automatic text summarization models. It compares the generated summary to a reference summary (often human-generated) by measuring the overlap of words, phrases, or n-grams between them.

There are several variants of ROUGE, including:

**ROUGE-1:**

Measures the overlap of unigrams (individual words) between the reference and generated summaries.

**ROUGE-2:**

Measures the overlap of bigrams (two consecutive words) between the reference and generated summaries.

**ROUGE-L:**

Measures the longest common subsequence (LCS) between the reference and generated summaries, which reflects sentence-level structure similarity.

### 3.2.6 Model Deployment

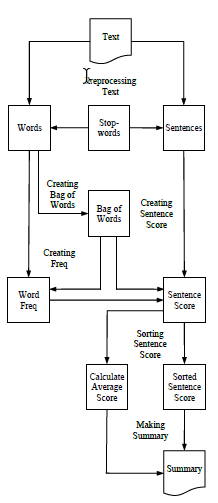
For the sake of model deployment, streamlit is used to create a simple web application for text summarization.

### 3.2.7 Deep Learning and NLP Techniques

Following deep learning techniques have been used in this research work.

#### 3.2.7.1 Word Frequency Algorithm

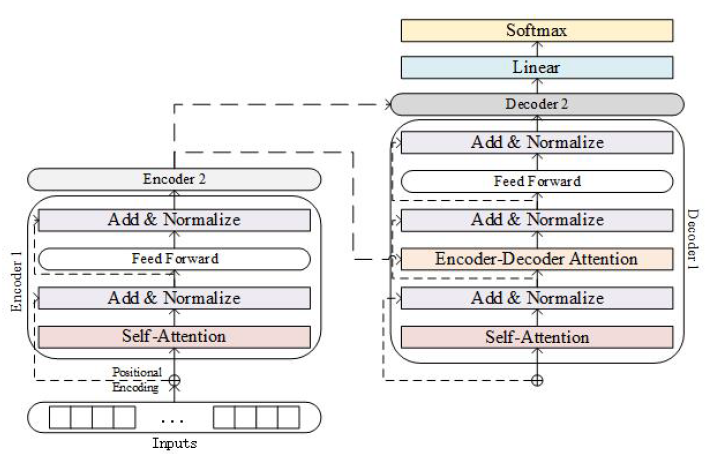
In this research, a Python-based text summarization tool that utilizes word frequency is designed. The program executes essential text summarization processes: text preprocessing, word listing, bag of words formation, word frequency computation, sentence listing, sentence scoring, score sorting, average score calculation, and summary generation. The flowchart of this technique is given in the following flowchart as mentioned by AlShammari (2023), to have an insight into the generation of summary through an extractive approach.



**Figure 3.2: Flowchart of Extractive Text Summarization**

#### 3.2.7.2 T5 model

The T5 model is a top-tier pre-trained language model built on transformer technology. It uses a simple text-to-text format, which means it can manage any task in natural language processing (NLP) by turning both the input and output into plain text. Architecture of T5 is also represented by Wang (2023) in the following diagram:



**Figure 3.3: Architecture of T5 model**

#### 3.2.7.3 BART model

The BART (Bidirectional and Auto-Regressive Transformers) model is a poweful tool for processing natural language tasks. According to Goodwin et al., (2020), BART uses both bidirectional and autoregressive transformers to achieve this. BART is trained with tasks like arranging sentences and filling in missing words, helping it understand context and produce clear text. The model has a bidirectional encoder that reads text in both directions and an autoregressive decoder that generates text one word at a time. BART's decoder also uses cross-attention to improve the accuracy of the generated text. BART-Large, a version of this model, has 12 layers in both the encoder and decoder, making it very effective for tasks like summarization, translation, and text generation. BART-IT is one of the extensions of BART that is used for text summarization in multiple languages as used by Quatra et al., (2022) and the results shown in their paper show that it is wonderful for text summarization of Italian.

# 

# CHAPTER 4

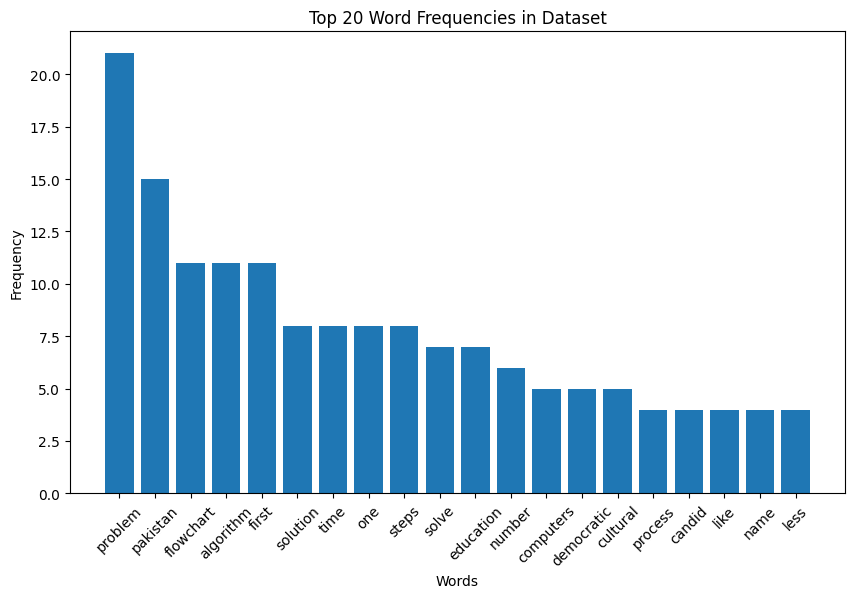
## RESULTS AND DISCUSSION

This chapter contains the details about the results that were predicted by the models after fine-tuning. The results are given in tabular form and the performances of two transformer models that are T5 and BART are also compared. Sequentially, on the basis of this experiment the models that showed high performance and accuracy is deployed using streamlit app.

## 4.1 Visualization

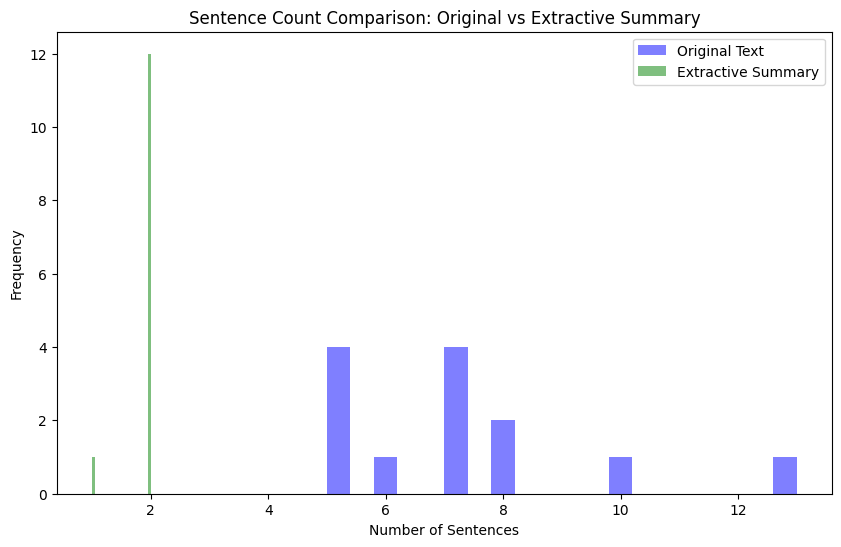
Visualization plays a critical role in this research by enhancing the understanding, analysis, and communication of our findings. Specifically, in the context of text summarization research using models like T5 and BART, visualization helps to interpret complex data, communicate results effectively, identify trends and patterns, and support decision-making. Here in this research visualization serves as a guide for understanding the dataset, identifying patterns and trends, evaluating the impact on summarization, and simplifying complex data for visualization. Through this visual exploration, the study gives a simple yet comprehensive understanding of the characteristics of the dataset as described below:

Firstly, the study has analyzed the educational dataset, the words have been analyzed against different words frequencies. Fig 4.1 shows the top 20 word frequencies in the dataset.



**Figure 4.1: Graph of Word’s Frequencies in Dataset**

Upon delving deeper into the dataset, a distinct number of sentences are to be extracted using the Word frequency Algorithm to predict the summary of the unseen content given by the user. A less number of sentences through WFA(Word frequency Algorithm) can lead to difficulty in training phase. Fig 4.2 represents this consideration and underscores the significance of the length of sentences in influencing the performance of the underlined models.



**Figure 4.2: Graph of Sentence Count Comparison of Original text vs Extractive Summary**

Similarly, one of the important thing that is needed to consider is the number of sentences that are taken in the dataset. As the dataset belongs from educational books , that's why it can vary in semantics but alike in length. Fig 4.3 represents this consideration and exhibits about the variation in number of sentences across multiple data items. This graph emphasizes the relationship between sentence lengths and their impact on the summarization process.

## IMG_256

**Figure 4.3: Graph of Sentence’s Lengths in dataset**

## 4.2 Experimental setup

For this research, a detailed dataset from educational books with different types of data was carefully put together to fine-tune our T5 and BART models for text summarization. This dataset included a collection of topics covered from educational books containing several rows. The variety in the dataset was designed to provide rich information for the summarization task.

Before training, the raw data went through thorough preprocessing, which included text cleaning, tokenization, calculation of word frequency, and removal of special words. After this step, the same dataset is passed under the Word Frequency Algorithm to get summaries of the text that was taken earlier. This step refined the dataset by removing irrelevant or redundant information and emphasizing important features needed for effective model training. The final dataset offered a wide range of topics and writing styles, as well as different sentiment expressions, making it perfect for training our models to handle complex and lengthy texts with high accuracy and speed. After the data collection and preprocessing, the training of this dataset is performed. It involves the fine-tuning of the models, which are T5 and Bart, sequentially. The fine-tuning of both the models is performed on the GPU of the HP Laptop 14-fq0xxxCPU with a processor speed of 2.4 GHZ, 6024 MB of RAM, and 256 GB of Hard Disk space. All this process involves the installation of numerous libraries, e.g. rouge\_score, numpy, panda, scikit-learn, BartForConditionalGeneration, and Bart Tokenizer from transformers, which are used for text generation tasks. The primary metric used to evaluate the performance of the summarization model in this research is ROUGE. ROUGE-1, ROUGE-2, and ROUGE-L scores are computed to measure the overlap between the generated summaries and the reference summaries at different levels. It measures the overlap of n-grams (contiguous sequences of words), longest common subsequences, and word sequences between the generated and reference summaries. In this research, ROUGE is used as the primary evaluation metric to assess the performance of the summarization model.

## 4.2 Results

The hybrid method explored in this research combines the Word Frequency Algorithm (WFA) and transformer models like T5 and BART. The WFA provided extractive summaries, which were then used as input for fine-tuning the transformer models for abstractive summarization.The results answer the research questions efficiently as discussed below.

**Q#1: How to implement a hybrid method for text summarization that combines the Word Frequency Algorithm with transformer models like T5 and Bart?**

To address this research question, we implemented a hybrid method for text summarization. The method integrates the Word Frequency Algorithm (WFA) for extractive summarization with transformer-based models like T5 and BART for abstractive summarization.

In this experiment, the source text was preprocessed and prefixed with the keyword "summarize," and the summaries generated by WFA were used as target text for training the T5 and BART models. In this process, the preprocessing involved tokenizing the text, removing stop words, and normalizing the data by converting all text to lowercase to maintain consistency. Additionally, unnecessary punctuation and special characters were removed to clean the text and enhance the model's understanding. This process effectively merged extractive and abstractive summarization techniques to generate high-quality summaries.

The fine-tuning of the models was conducted over three epochs, and the performance was evaluated using ROUGE metrics. Tools like Python, TensorFlow, and PyTorch were utilized for this process, along with pre-trained T5 and BART models. Our hybrid approach successfully generated summaries by first identifying key sentences (using WFA) and then abstracting them into coherent summaries using T5 and BART.

By comparing ROUGE scores across epochs in table 4.1 , the results indicate that the hybrid method not only works but also enhances the performance of the models in summarization tasks. This directly answers the research question by proving the effectiveness of a hybrid approach combining extractive and abstractive techniques.

**Table 4.1: Results of T5 and BART after fine-tuning for text summarization**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Summarization Approach** | **ROUGE type** | **ROUGE1** | **ROUGE2** | **ROUGEL** |
| T5 base | Epoch 1 | 0.497 | 0.402 | 0.406 |
| Epoch 2 | 0.715 | 0.646 | 0.532 |
| Epoch 3 | 0.694 | 0.611 | 0.606 |
| BART base | Epoch 1 | 0.697 | 0.611 | 0.444 |
| Epoch 2 | 0.732 | 0.637 | 0.425 |
| Epoch 3 | 0.715 | 0.621 | 0.456 |

FIGURE 4.4 represents the comparison between the BART and T5 models. The experiment results in TABLE 4.1 and FIGURE 4.4 show that the system achieves better ROUGE scores for ROUGE-1 is 0.715 % against T5 and for ROUGE-1 is 0.732 % against BART. It has also seen that the proposed model outperforms some previous similar hybrid methods such as mentioned by Muia et al.,(2024) T5 (ROUGE-1 35.12 %, ROUGE-L 32.82 %) and BART (ROUGE-1 27.61 %, ROUGE-L 28.52 %). In terms of ROUGE-1 and ROUGE-L metrics, the proposed model obtained the best performance.

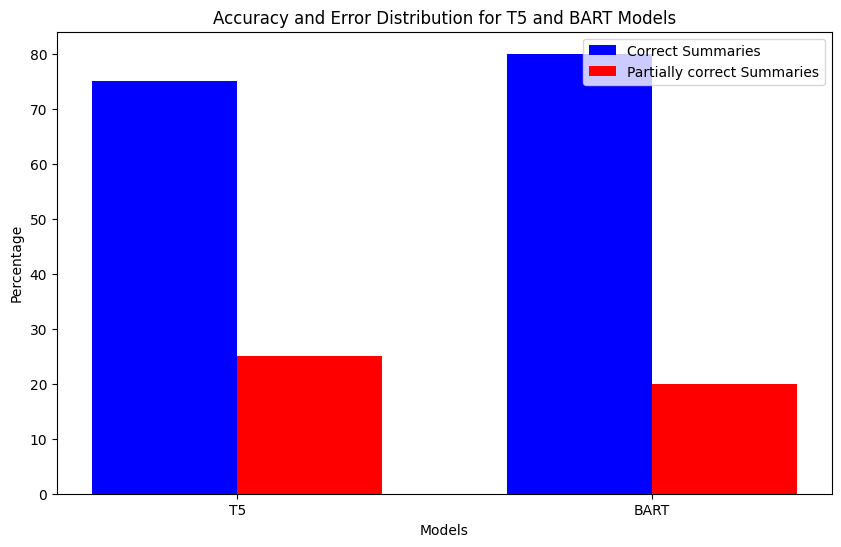
**Figure 4.4: Graphical representation of results of T5 and BART**

**Q#2: How to enhance the accuracy of abstractive models T5 and Bart by fine-tuning on self-created dataset?**

In order to enhance the accuracy of the T5 and BART models, we created and fine-tuned them on a self-created dataset consisting of educational books. The data collection involved preprocessing the text, generating extractive summaries using WFA, and passing this data through T5 and BART models. To fine-tune the T5 model for the hybrid summarization approach, the source text was preprocessed by tokenizing, removing stop words, normalizing the text, and prefixing it with "summarize:" to guide the model. Extractive summaries generated using the Word Frequency Algorithm (WFA) served as the target text, creating a structured dataset that merged extractive and abstractive summarization techniques. The preprocessed data was split into training and testing sets, and a pre-trained T5 model (t5-base) was fine-tuned using the AdamW optimizer with a learning rate of 5e-5. During training, the model learned to map the input text to the extractive summaries, minimizing the loss between predictions and targets. ROUGE-1, ROUGE-2, and ROUGE-L scores were used to evaluate the generated summaries, ensuring alignment with content and structure. This approach enabled the model to produce high-quality, contextually accurate summaries by integrating extractive and abstractive methods.

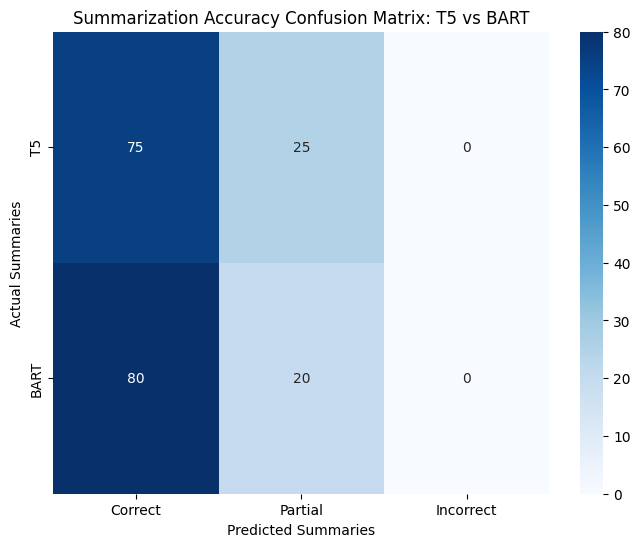
We performed fine-tuning over three epochs, adjusting hyper parameters such as learning rate and batch size to maximize model performance. The ROUGE-1, ROUGE-2, and ROUGE-L metrics were used to evaluate the performance of the models across different epochs.

The results highlight that fine-tuning significantly improved the models' performance. For instance, the T5 model's ROUGE-1 score improved from 0.497 at Epoch 1 to 0.715 at Epoch 2. Similarly, BART’s ROUGE-1 score increased from 0.697 at Epoch 1 to 0.732 at Epoch 2.



**Figure 4.5: Graphical representation of Accuracy Distribution of T5 and BART**

Figure 4.5 presents the error distribution between the two transformer models, T5 and BART, as applied to the summarization task. This visualization provides an in-depth look at the proportion of Incorrect, Partial, and Correct summaries generated by each model, helping address the second research question. This analysis contributes directly to the exploration of Research Question 2, which seeks to improve model accuracy through fine-tuning. By analyzing the error distribution, it becomes clear that T5's fine-tuning resulted in fewer errors, making it a more reliable model for this summarization task.



**Figure 4.6: Graphical representation of results of Accuracy Confusion Matrix of T5 and BART**

The confusion matrix in Figure 4.6 provides a visual comparison of summarization accuracy between the T5 and BART models across evaluated samples. The matrix categorizes the predictions into three types: Correct, Partial, and Incorrect summaries.

The confusion matrix serves as a crucial tool for understanding the comparative strengths and weaknesses of each model.This heat map answers the second research question, "How to enhance the accuracy of abstractive models T5 and BART by fine-tuning on self-created datasets?", by showing the impact of model fine-tuning on different types of summarization outputs. The results clearly indicate that T5 benefits more from fine-tuning on the custom dataset, as evidenced by its higher proportion of correct summaries.

## 4.4 Discussion

The table summarizes the ROUGE scores achieved through different summarization approaches using T5 base and BART base models over multiple epochs. Here's an in-depth discussion of these results:

### 4.4.1 Model Performance Improvement Over Epochs

The importance of extractive and abstractive approaches in automated text summarization can be seen in literature. As Saxena et al., (2023) in their paper demonstrated the potential of using abstractive summarization for podcasts, providing listeners with a quick and accurate summary of episodes. Similarly, Pilault et al., (2020) have also brought light the use of extractive abstractive approach for achieving high efficiency by adopting larger models. These studies provide as a motivation for taking these approaches as a background for modern text summarization.

Both T5 base and BART base models show significant improvement in ROUGE scores from Epoch 1 to Epoch 2. This indicates that both models benefit from additional training, which helps them better capture the relationships and patterns in the text for generating summaries. As compared to the previous work by Wang et al., (2023), it can be clearly observed that the performance of T5 for text summarization can be enhanced by fine-tuning or by adopting any other approach as done in this research. Similarly, by training on a large dataset over multiple epochs can also paves a path towards further improvement, if we have enough GPU.

For the T5 base model, the ROUGE1 score increases from 0.4972 in Epoch 1 to 0.7152 in Epoch 2, demonstrating a considerable enhancement in capturing the overlap of uni-grams between the generated summaries and reference summaries.

Likewise, the performance of BART has also improved through fine-tuning and by adopting hybrid approach. Previous literature shows the BART for text summarization with very low accuracy in terms of ROUGE scores as mentioned by Muia, et al., (2024). They contributed to the field by providing a detailed comparison of different summarization techniques on a dataset of scientific papers and by suggesting future research directions to improve summarization performance for long and complex texts. The results by them have shown ROUGE scores that are not up-to-mark.

**4.4.2 ROUGE2 and ROUGEL Scores:**

ROUGE2 scores (which measure bi-gram overlap) also improve with additional epochs, highlighting that the models are getting better at capturing more contextual and sequential information.

ROUGEL (which measures the longest common sub sequence) shows a similar trend. However, it is interesting to note that for the T5 base model, there is a slight decrease in ROUGE2 and ROUGEL scores from Epoch 2 to Epoch 3, suggesting potential over-fitting or other issues that may need further investigation.One of the reason for this over-fitting is the limited availability of required GPU to handle required dataset.

**Comparison Between T5 base and BART base:**

In Epoch 1, BART base outperforms T5 base across all ROUGE metrics, particularly in ROUGE2 and ROUGEL, indicating that BART base may have better initial performance on this task. By Epoch 2, BART base continues to slightly outperform T5 base in ROUGE1 and ROUGE2, but shows a decrease in ROUGEL. This suggests that while BART base may be better at capturing bi-gram relationships, it might struggle with capturing longer sub sequences compared to T5 base. Similarly, the score of ROUGE that is achieved through this research work is absolutely far high as compared to the work done in previous studies. As Aswani et al., (2024) have mentioned tailored abstractive text summarization approaches in their paper. And according to their results the highest ROUGE score is 50.3. This show s that the techniques in this studies perform well, as our maximum ROUGE score is 0.73.

## 4.5 Deployment

For the sake of model deployment, streamlit is used to create a simple web application for text summarization. The following main tasks are used in this phase:

* Setting the title of the app.
* Creating a text area for user input.
* Adding a button to generate the summary.
* Generating and displaying the summary when the button is pressed.

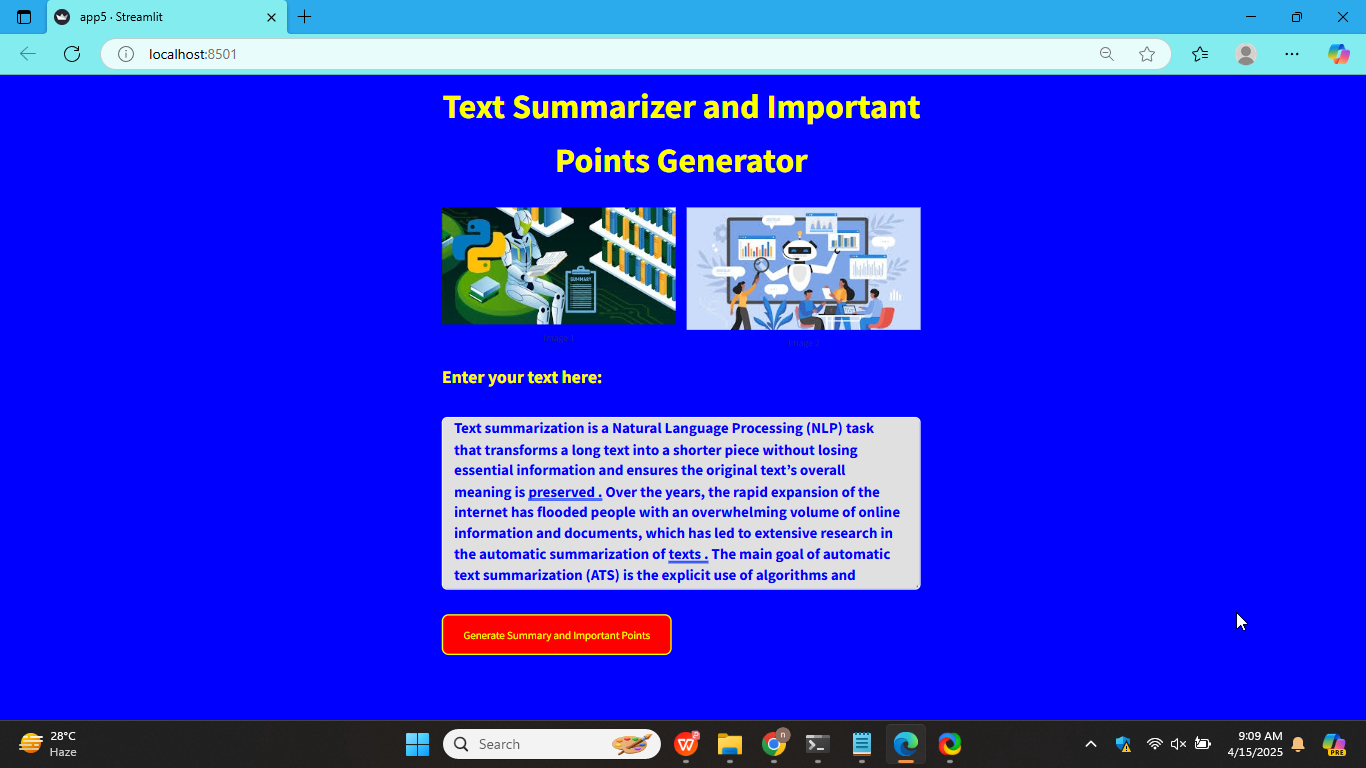
Following are the same figures that show the stepwise performance of this app. In Figure 4.7 the initial interface of the app is shown. There is text box where user enter the text that is needed to summarized.



**Figure 4.7: Text summarizer and important points' generator app**

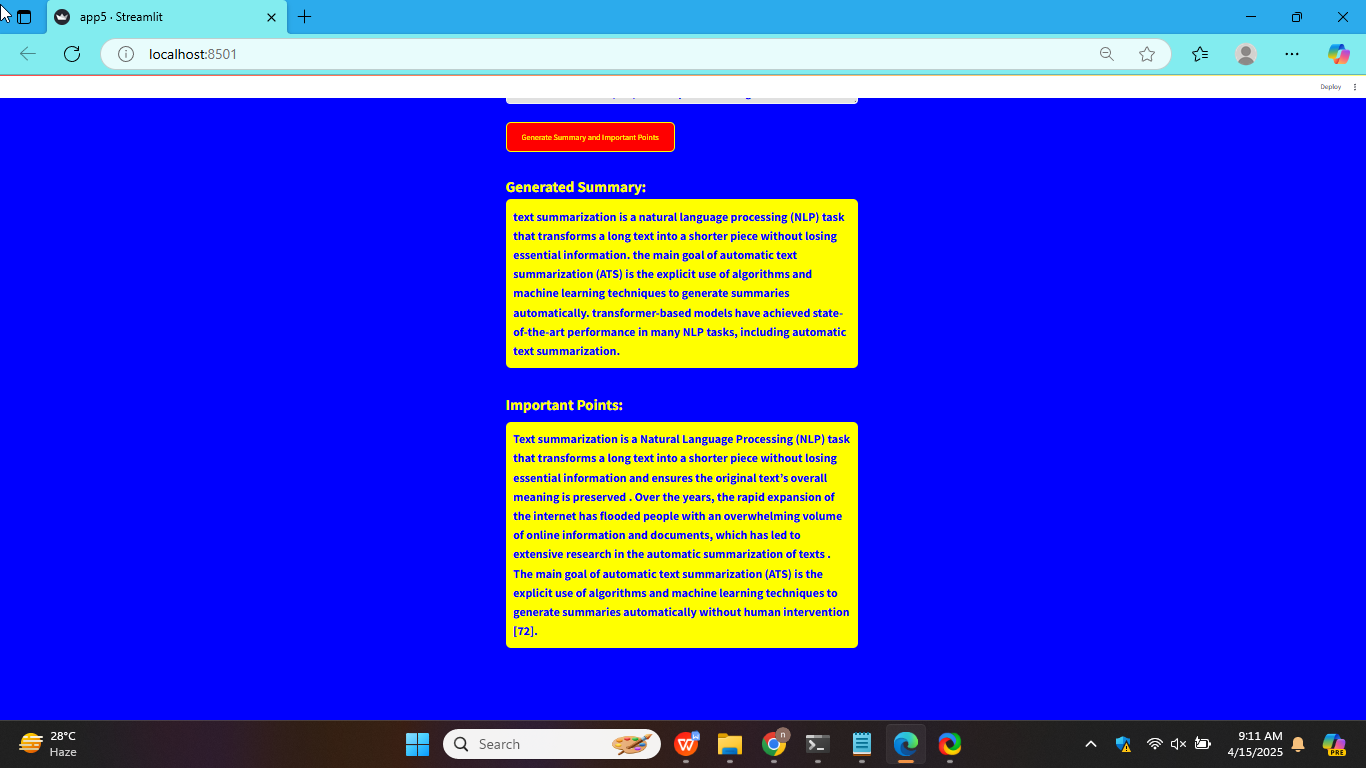
**Figure 4.7:Text summarizer and important points generator app**

In FIGURE 4.8 the user will enter text or simply paste text from any other source.



**Figure 4.8: Input text by the user**

In FIGURE 4.9 when user clicks on the button labeled as Generate Summary and Important points , the app generates the output as mentioned in this figure.



**Figure 4.9: Output summary and highlighted important points by the app**

# CHAPTER 5

# SUMMARY

## 5.1 Summary

The current study has been conducted to enhance the efficiency of text summarization in NLP (Natural Language Processing). Numerous apps and software for text summarization are available on the internet, but all this software works traditionally. Similarly, it is very obvious from the literature review that most of the techniques already developed for text summarization either work on the summarization of newspapers or articles and all of these have very little accuracy.

Furthermore, most of the papers shown the adoption of either extractive or abstractive techniques for text summarization. Some of the literature work shown that the performance of text summarization can be increased by taking suitable dataset, by changing hyper parameters specifically or by merging both extractive abstractive techniques to have hybrid approach.

Likewise, after detailed study of these research papers, the implementation of a hybrid approach for text summarization has adopted in current study. To implement this approach firstly WFA (Word Frequency Algorithm) has implemented which lies under extractive approach and then T5 and BART were implemented after fine-tuning to achieve required results.

## 5.2 Thesis Contributions

The current thesis contributes the following:

1. Development of a summarization model, keeping in view the adoption of hybrid approaches to achieve high accuracy in results
2. Advancements of NLP (Natural Language Processing) techniques e.g. T5 and BART to make its effective use for educational purpose
3. Enhancement of learning outcomes
4. Pedagogical support for educators to bring efficiency in teaching learning process
5. Provision of guidance to the students for consulting multiple resources without being distracted

## 5.2 Limitation of current study

1. **Limited to Educational Content:**  
   The models were trained mainly on educational books, so it might not work well on other types of content like news, stories, or legal documents.
2. **Relies Too Much on Extractive Summarization**  
   The final summary depends on the sentences chosen by the Word Frequency Algorithm. If this step misses important information, the final summary may not be very useful.
3. **Lack of Deep Understanding of Text**  
   The extractive method focuses only on word counts. It doesn’t understand the meaning of the text, so it might leave out important ideas that aren’t repeated often.
4. **Small Dataset for Training**  
   If the training data is not large enough, the models like T5 or BART might not learn well. This can affect the quality of summaries.
5. **Needs High Computer Power**  
   Training transformer models takes a lot of time and needs powerful computers or GPUs. This can be a challenge for researchers with limited resources.
6. **Problems with Repeating or Unclear Sentences**  
   Sometimes, the final summaries may have repeated sentences, unclear wording, or lose the main idea.
7. **Evaluation May Not Reflect True Quality**  
   The ROUGE scores used to measure summary quality only check for word overlap. They don’t show how well the summary reads or if it makes sense.
8. **Model Can Be Biased**  
   The transformer model may learn patterns from the data that are biased or not general. This can affect the fairness or accuracy of the summaries.
9. ·**Difficulty with Technical Terms**  
   The summarizer may not handle technical words, formulas, or diagrams very well, especially in subjects like science or math.

## 5.3 Future Work

Automated Text Summarization provides the fastest way to generate a summary of the input text entered by the user. Text summarization not only paves a path towards modernization in textual data, but it also provides a solution to the biggest challenge faced by the MCQ developers in the form of text summarization. A good and precise summary is used for the development of MCQs it. As text summarization is the biggest challenge in making online MCQs of the text entered by the user.

Furthermore, summarization can be used for question generation by adopting of it as a preprocessing technique. Researchers can further work on question generation by using this research, as it provides them with a background of transformer models.

Moreover, the future work that can be done in near future in the same domain is to enhance its accuracy further for adoption of larger datasets with fully supported and equipped IDE. Researchers can enhance the domain of this text summarization towards images and flowcharts mentioned in educational books.

# LITERATURE CITED

APPENDICES

References

1. Soriano, E. S., Ahuir, V., Hurtado, L. F., & González, J. (2022, July). Dacsa: A large-scale dataset for automatic summarization of catalan and spanish newspaper articles. *IN PROCEEDINGS OF THE 2022 CONFERENCE OF THE NORTH AMERICAN CHAPTER OF THE ASSOCIATION FOR COMPUTATIONAL LINGUISTICS*: Human Language Technologies (pp. 5931-5943).
2. Goloviznina, V., & Kotelnikov, E. (2022). Automatic Summarization of Russian Texts: Comparison of Extractive and Abstractive Methods. arXiv preprint arXiv:2206.09253.
3. Gurusamy, B. M., Rengarajan, P. K., & Srinivasan, P. (2023). A hybrid approach for text summarization using semantic latent Dirichlet allocation and sentence concept mapping with transformer. *INTERNATIONAL JOURNAL OF ELECTRICAL AND COMPUTER ENGINEERING (IJECE),* 13(6), 6663-6672.
4. Zaware, S., Patadiya, D., Gaikwad, A., Gulhane, S., & Thakare, A., (2021). Text Summarization using TF-IDF and Textrank algorithm. *PROCEEDINGS OF THE FIFTH INTERNATIONAL CONFERENCE ON TRENDS IN ELECTRONICS AND INFORMATICS (ICOEI).*
5. Zhao, Z., Hou, Y., Wang, D., Liu, M. Y. & Ma, X. (2022, MARCH 27). Educational question generation of children storybooks via question type distribution learning and event-centric summarization. arXiv:2203.14187v1 [cs.CL]
6. Gunuputi, V. V., & Bharathi, B. R. (2022). Improving hybrid summarization by using abstract and extract model. *INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT*), 10(9), 266-275. <https://www.ijcrt.org/a267>
7. Gopikakrishna, N. G., Sreenivasan, P., Chandran, V., K P, Y. K., Dev, S. S., & V V#, K., (2021). Comparative study on text summarization using nlp and rnn methods. *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT)*, 9(13), [www.ijert.org](http://www.ijert.org)
8. Rani, U., & Bidhan, K. (2021). Comparative assessment of extractive summarization: textrank tf-idf and lda. *JOURNAL OF SCIENTIFIC RESEARCH*, *65*(1), 304-311.
9. Mandal, S., & Singh, G. K. (2020). LSA based text summarization. *INT J RECENT TECHNOL Eng*, *9*, 150-156.
10. Wang, M., Xie, P., Du, Y., & Hu, X. (2023). T5-based model for abstractive summarization: A semi-supervised learning approach with consistency loss functions. *APPLIED SCIENCES*, *13*(12), 7111.
11. Etemad, A. G., Abidi, A. I. & Chhabra, M. (2021, OCTOBER). Fine-tuned t5 for abstractive summarization. *INTERNATIONAL JOURNAL OF PERFORMABILITY ENGINEERING*, 17(10),900-906. DOI: 10.23940/ijpe.21.10.p8.900906
12. Deokar, V. & Shah, K. (2021, SEPTEMBER 6). Automated text summarization of news articles. *INTERNATIONAL RESEARCH JOURNAL OF ENGINEERING AND TECHNOLOGY (IRJET*), 8,1023–1033. https://doi.org/10.26615/978-954-452-092-2\_110
13. Karanja, J. M. & Matheka,A. (2022, NOVEMBER 28). A hybrid model for text summarization using natural language processing. *OPEN JOURNAL FOR INFORMATION TECHNOLOGY*, 5(2), 65-80. <https://doi.org/10.32591/coas.ojit.0502.03065k>
14. Guoy, M. , Joshua, A. , Uthus, D. , Ontañón, S. , Ni, J. , Sung, Y. & Yang, Y. (2022, MAY 3). Longt5: efficient text-to-text transformer for long sequences*.  International Conference on Learning Representations (ICLR),* https://github.com/google/flaxformer/tree/main/flaxformer/architectures/longt52https://github.com/google-research/longt5
15. Mengi, R. , Ghorpade, H., & Kakade, A., (2023, DECEMBER 5). Fine-tuning t5 and Robert a models for enhanced text summarization and sentiment analysis. *THE GREAT LAKES BOTANIST*. <https://www.researchgate.net/publication/376232167>
16. Kumar, S., Chauhan, A., & Kakade, P., (2022, DECEMBER). Augmenting ebooks with recommended questions using contrastive fine-tuned t5. *PROCEEDINGS OF THE 19TH INTERNATIONAL CONFERENCE ON NATURAL LANGUAGE PROCESSING (ICON),* 109 - 115, https://github.com/uci-soe/FairytaleQAData 2 https://ncert.nic.in/textbook.php
17. Kaur, G., & Sharma, A., (2024, APRIL). Automatic customer review summarization using deep learning-based hybrid sentiment analysis. *INTERNATIONAL JOURNAL OF ELECTRICAL AND COMPUTER ENGINEERING (IJECE)* ,14(2), 2110-2125. DOI: 10.11591/ijece.v14i2.pp2110-2125
18. Liu, Y., Dou, Z., & Liu, P., (2021, APRIL 15). RefSum: Refactoring Neural Summarization. arXiv:2104.07210v1 [cs.CL], https://github.com/yixinL7/ Refactoring-Summarization.http://explainaboard.nlpedia. ai/leaderboard/task-summ/index. php.
19. Rehman, T., Das, S., Sanyal, D. K., & Chattopadhyay, S., (2023 FEBRUARY 25). An analysis of abstractive text summarization using pre-trained models. arXiv:2303.12796v1 [cs.CL]
20. Merrouni, Z. A., Frikh, B., & Ouhbi, B., (2023). EXABSUM: a new text summarization approach for generating extractive and abstractive summaries. *JOURNAL OF BIG DATA,* <https://doi.org/10.1186/s40537-023-00836->
21. Widyassaria, A. P., Rustad, S., Shidika, G. F., Noersasongko, E., Syukur, A., Affandy, A., & Setiadi, D. R. I. M., (2022). Review of automatic text summarization techniques & methods. *JOURNAL OF KING SAUD UNIVERSITY – COMPUTER AND INFORMATION SCIENCES* 34, 1029–1046
22. Abidin, A. Z. Z., Murdianingsih, Y., Suryadi U. T., & Setiyadi D., (2020). Text summarizing system of english subjects and text mining subjects. *JOURNAL OF CRITICAL REVIEWS*, Vol 7(5)
23. Christian, H., Agus, M. P., & Suhartono, D., (2016, DECEMBER). Single document automatic text summarization using term frequency-inverse document frequency (tf-idf).*COMTECH COMPUTER MATHEMATICS AND ENGINEERING APPLICATIONS*.DOI: 10.21512/comtech.v7i4.3746
24. Litvak, M., Vanetik, N., & Krimberg, N., (2020).Summarization of financial documents with TF-IDF weighting of multi-word terms
25. Zeng, W. C., (2024, JANUARY 31). Fine-tuning Transformer-based Natural Language Generation Algorithms for USDA Grains Reports for Farmers, Producers, and Small Businesses*. THE UNIVERSITY OF ARIZONA* USA ,http://hdl.handle.net/10150/668771
26. Lubis, A. R., Safitri, H. R., Lubis, M., Hamzah, M. L., Al-Khowarizmi, A. K., & Nugroho, O. (2023). Enhancing Text Summarization with a T5 Model and Bayesian Optimization. *REVUE D'INTELLIGENCE ARTIFICIELLE*, 37(5).
27. Chaurasia, S., Dasgupta, D., & Regunathan, R. (2023). T5LSTM-RNN based Text Summarization Model for Behavioral Biology Literature. *PROCEDIA COMPUTER SCIENCE*, 218, 585-593.
28. Gopikakrishna, N. G., Sreenivasan, P., Chandran, V., KP, Y. K., Dev, S. S., & Krishnaveni, V. V. Comparative Study on Text Summarization using NLP and RNN Methods.
29. Borah, M., Dadure, P., & Pakray, P. (2022). Comparative analysis of T5 model for abstractive text summarization on different datasets.
30. Tsonkov, T., Lazarova, G. A., Zmiycharov, V., & Koychev, I. (2021). A Comparative Study of Extractive and Abstractive Approaches for Automatic Text Summarization on Scientific Texts. In ERIS (pp. 29-34).
31. Aswani, S., Choudhary, K., Shetty, S., & Nur, N. (2024). Automatic text summarization of scientific articles using transformers—A brief review. *JOURNAL OF AUTONOMOUS INTELLIGENCE*, 7(5).
32. Karnik, M. P., & Kodavade, D. (2023). Abstractive Summarization with Efficient Transformer Based Approach. *INTERNATIONAL JOURNAL ON RECENT AND INNOVATION TRENDS IN COMPUTING AND COMMUNICATION*, 11(4), 291-298.
33. Muia, C. M., Oirere, A. M., & Ndungu, R. N. (2024). A Comparative Study of Transformer-based Models for Text Summarization of News Articles.
34. Talukder, M. A. I., Abujar, S., Masum, A. K. M., Akter, S., & Hossain, S. A. (2020, July). Comparative study on abstractive text summarization. In 2020 *11TH INTERNATIONAL CONFERENCE ON COMPUTING, COMMUNICATION AND NETWORKING TECHNOLOGIES (ICCCNT)* (pp. 1-4). IEEE.
35. Odeyemi, G. (2025). Finetuning Encoder-Decoder Transformer Models for Text Summarization using Pause Tokens (MASTER'S THESIS, UNIVERSITY OF WINDSOR (CANADA)).
36. Peddarapu, R. K., Anogna, S., Harshini, M. S., Sireesha, J., & Rathna, K. G. (2025). Document Summarizer: A Machine Learning approach to PDF summarization. *PROCEDIA COMPUTER SCIENCE, 252, 1025-1036.*
37. Fendji, J. L. K., Donatien, D., & Atemkeng, M. (2025). Hybrid Profile based Multi-document Text Summarisation. Procedia Computer Science, 252, 862-872.
38. Bharti, T., Sinha, S. K., Singhal, H., Saini, R., & Parihar, D. (2025). Abstractive Text Summarization Using GAN. International Journal of Innovative Science and Research Technology (IJISRT).
39. Wang, M., Xie, P., Du, Y., & Hu, X. (2023). T5-based model for abstractive summarization: A semi-supervised learning approach with consistency loss functions. *APPLIED SCIENCES*, 13(12), 7111.
40. AlShammar, A. (October, 2023). Flowchart of Text Summarization [photograph]. Implementation of Text Summarization using Word Frequency in Python. *INTERNATIONAL JOURNAL OF COMPUTER APPLICATIONS*, Volume 185 (36)
41. Goodwin, T. R., Savery, M. E., & Demner-Fushman, D. (2020, November). Towards zero-shot conditional summarization with adaptive multi-task fine-tuning. *IN PROCEEDINGS OF THE CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING. CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING*(Vol. 2020, p. 3215). NIH Public Access.
42. La Quatra, M., & Cagliero, L. (2022). Bart-it: An efficient sequence-to-sequence model for italian text summarization*. FUTURE INTERNET*, 15(1), 15.
43. Saxena, P., & El-Haj, M. (2023, September). Exploring Abstractive Text Summarisation for Podcasts: A Comparative Study of BART and T5 Models. *IN PROCEEDINGS OF THE 14TH INTERNATIONAL CONFERENCE ON RECENT ADVANCES IN NATURAL LANGUAGE PROCESSING* (pp. 1023-1033).
44. Pilault, J., Li, R., Subramanian, S., & Pal, C. (2020, November). On extractive and abstractive neural document summarization with transformer language models. *IN PROCEEDINGS OF THE 2020 CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING (EMNLP)* (pp. 9308-9319).
45. Das, B., Majumder, M., Phadikar, S., & Sekh, A. A. (2021). Automatic question generation and answer assessment: a survey. *RESEARCH AND PRACTICE IN TECHNOLOGY ENHANCED LEARNING,* 16(1), 5.
46. Nanda Devi, D. R., Bitla, S., Janagam, S., Kothakapu, A. R., & Atiki, L. (2023). *Question generation system using NLP*. *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT),* 12(04), 375-377. <http://www.ijert.org/IJERTV12IS040277>
47. Habib, M. A., Ema, R. R., Islam, T., Arafat, M. Y., & Hasan, M. (2023). Automatic text summarization based on extractive-abstractive method. Radioelectronic and Computer Systems, (2), 5-17.