

Thesis

by Rishabh Chauhan

Submission date: 22-Jul-2026 12:44PM (UTC+0530)

Submission ID: 3001673097

File name: mtech_thesis_rishabh.pdf (1.39M)

Word count: 5877

Character count: 33549

RESEARCH TRENDS IN TEXT SUMMARIZATION WITH A FOCUS ON FINE-TUNING INSTRUCTION-TUNED LANGUAGE MODELS

1

**Thesis Submitted
in Partial Fulfillment of the Requirements for the
Degree of**

MASTER OF TECHNOLOGY

in

Computer Science And Engineering

by

Rishabh Chauhan

(Roll No. 23/CSE/13)

Under The Supervision Of

Ms. ANUKRITI KAUSHAL

**Assistant Professor, Computer Science and Engineering
Delhi Technological University**



Department of Computer Science and Engineering

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-110042, India

MAY, 2025



DELHI TECHNOLOGICAL UNIVERSITY
 (Formerly Delhi College of Engineering)
 Shahbad Daulatpur, Main Bawana Road, Delhi-42

CANDIDATE'S DECLARATION

I, Rishabh Chauhan, 23/CSE/13 of M.Tech (Computer Science and Engineering), hereby certify that the work which is being presented in the thesis entitled "Research Trends In Text Summarization With A Focus On Fine-Tuning Instruction-Tuned Language Models" in partial fulfillment of the requirements for the award of the degree of Master of Technology, submitted in the Department of Computer Science and Engineering, Delhi Technological University is an authentic record of my own work carried out during the period from to under the supervision of Ms. Anukriti Kaushal.

The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other institute.

Candidate's Signature

This is to certify that the student has incorporated all the corrections suggested by the examiners in the thesis and the statement made by the candidate is correct to the best of our knowledge.

Signature of Supervisor

Signature of External Examiner



DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE BY THE SUPERVISOR

Certified that **Rishabh Chauhan**(23/CSE/13) ¹ has carried out their search work presented in this thesis entitled **"Research Trends In Text Summarization With A Focus On Fine-Tuning Instruction-Tuned Language Models"** ¹ for the award of **Master of Technology** from Department of Computer Science and Engineering, Delhi Technological University, Delhi, under my supervision. The thesis embodies results of original work, and studies are carried out by the student himself and the contents of the thesis do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution.

Signature

(Ms. Anukriti Kaushal)

(Assistant Professor)

¹
(DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Shahbad Daulatpur, Main Bawana Road, Delhi-42)

Date: 30 May 2025

RESEARCH TRENDS IN TEXT SUMMARIZATION WITH A FOCUS ON FINE-TUNING INSTRUCTION-TUNED LANGUAGE MODELS

RISHABH CHAUHAN

ABSTRACT

Text summarisation has emerged as a crucial area within Natural Language Processing, driven by the increasing need to process and condense vast amounts of textual data. This thesis presents a twofold investigation into the domain: (1) a comprehensive bibliometric analysis of global research trends in text summarisation, and (2) a performance-focused study on hyperparameter tuning for domain-specific text generation using transformer-based models.

The first study employs bibliometric methods, visual analytics, and critical review to trace the evolution of text summarisation across decades, highlighting key contributors, publication trends, influential methodologies, and emerging directions such as multilingual and multimodal summarisation. This exploration not only contextualizes the field's growth but also identifies research gaps and strategic opportunities for innovation.

Building upon these insights, the second study investigates the practical challenges of adapting instruction-tuned models for domain-specific summarisation. Using the BillsSum legal dataset, the performance of two models—BART-Large and FLAN-T5 Base—was evaluated. Initial experiments revealed that FLAN-T5 underperformed significantly compared to BART-Large. Traditional strategies such as data augmentation and cross-domain pretraining failed to yield improvements. However, targeted hyperparameter optimization led to substantial gains, emphasizing the critical role of configuration over model architecture alone.

Together, these studies offer both a macro and micro perspective of the field: from understanding its research trajectory to demonstrating how careful tuning can unlock the potential of underperforming models for specialized summarisation tasks. The findings aim to inform both researchers and practitioners seeking to advance summarisation technologies in theory and application.

ACKNOWLEDGEMENTS

I wish to express my sincerest gratitude to Ms. Anukriti Kaushal for her continuous guidance and mentor-ship that she provided during the M.Tech thesis. She showed me the path to achieving targets by explaining all the tasks to be done and explained the importance of this work as well as its industrial relevance. She was always ready to help and clear doubts regarding any hurdles in this project work. Without her constant support and motivation, this project would not have been successful.

(RISHABH CHAUHAN)

CONTENTS

Title	Page No.
Candidate's Declaration	II
Certificate	III
Abstract	IV
Acknowledgements	V
Contents	VI
List of Tables	VIII
List of Figures	IX
List of Abbreviations	XI
CHAPTER 1: INTRODUCTION	1
1.1 Overview	1
1.2 Evolution of Text Summarisation Techniques	1
1.3 Motivation for Bibliometric Exploration	2
1.4 Challenges in Domain-Specific Summarisation	2
1.5 Research Objectives	2
CHAPTER 2: RELATED WORK	3
2.1 Literature review	3
2.2 Instruction Tuned Models	3
2.3 Denoising Language Models	4
2.4 Datasets	6
CHAPTER 3: METHODOLOGY	9

3.1 Model Optimization Methodology	9
3.2 Combined Analysis of Publication Trends, Citations, and Collaboration Networks	10
3.3 Citation and Collaboration Metrics	13
3.4 Co-Authorship and Co-Citation Networks	16
3.5 Publisher Landscape	17
CHAPTER 4: RESULT	18
4.1 Experimental Results on Model Performance	18
CHAPTER 5: CONCLUSION, FUTURE SCOPE, SOCIAL IMPACT	26
5.1 Conclusion	26
5.2 Future Scope	27
5.3 Social Impact	27
REFERENCES	28
LIST OF PUBLICATIONS	30

LIST OF TABLES

Table No.	Table Name	Page No.
Table 3.4.1 ⁴	Top researchers with the highest TLS ratings.	16
Table 4.1.1	Bart-Large On Billsum With 8000 Rows In Train Set And 1000 Rows In Test Set ⁵	18
Table 4.1.2	FLAN-T5-Base on BillSum with 8000 Rows in Train Set and 1000 Rows in Test Set ⁵	20
Table 4.1.3	FLAN-T5-Base on BillSum, With 8000 Rows in Train Set and 1000 Rows in Test Set, After Training on XSum ⁵	21
Table 4.1.4	Flan-T5-Base on BillSum With 18.9K Rows in Training and 3.27K Rows in Test Set.	23
Table 4.1.5	FLAN-T5-Base on BillSum After Hyperparameter Tuning	24

LIST OF FIGURES

Figure No.	Figure Name	Page No.
3 Fig 2.3.1	An overview of Text Summarisation Techniques.	5
Fig 2.4.1	Prisma diagram for search query.	7
Fig 3.2.1.1	An annual publication growth graph illustrating the rise in publications from 2015 to 2024.	11
3 Fig 3.2.1.2	The top 10 categories in the publications.	12
Fig 3.2.2.1	The top ten research directions in the articles indicate the main areas of attention.	13
Fig 3.3.1	The ten most productive locations.	14
Fig 3.3.2	The organizations that produce the greatest number of research publications.	15
Fig 3.3.3	2 The collaborative network of places.	15
Fig 3.4.1	The network of co-citations for articles.	16
Fig 3.5.1	Chord diagram of publisher and publishing year.	17
Fig 4.1.1	Loss graph for Table 4.1.1	19
Fig 4.1.2	Rouge score graph for 22 Table 4.1.1	19
Fig 4.1.3	Loss graph for Table 4.1.2	20
Fig 4.1.4	Rouge score graph for Table 4.1.2	21
Fig 4.1.5	Loss graph for Table 4.1.3	22
Fig 4.1.6	Rouge score graph for Table 4.1.3	22
Fig 4.1.7	Loss graph for Table 4.1.4	23

¹³ Fig 4.1.8	Rouge score graph for ¹⁴ Table 4.1.4	24
Fig 4.1.9	Loss graph for Table 4.1.5	25
Fig 4.1.10	Rouge score graph for Table 4.1.5	25

LIST OF ABBREVIATIONS

Abbreviation	Meaning
TF-IDF	Term Frequency-Inverse Document Frequency
FLAN-T5	Fine-Tuned Language Model - T5
GPT	Generative Pre-training Transformer
BERTSUM	Variant of BERT for extractive summarisation
BART	Bidirectional and Auto-Regressive Transformers
NLP	Natural Language Processing
ROUGE	Recall-Oriented Understudy for Gisting Evaluation
SCI-EXPANDED	Science Citation Index Expanded
SSCI	Social Sciences Citation Index
A&HCI	Arts and Humanities Citation Index
CPCI-S	Conference Proceedings Citation Index – Science

CHAPTER 1

INTRODUCTION

1.1 Overview

15 Within the field of Natural Language Processing (NLP), automatic text summarization has emerged as a crucial field that provides effective means of gleaning pertinent information from vast amounts of data. The skill to write coherent and short summaries has applications across a variety of sectors, including healthcare, education, media, and especially law—where lengthy papers and thick textual material are widespread. This thesis explores two complementary aspects of this field: the evolutionary landscape of text summarisation research and the optimization of modern summarisation models for domain-specific tasks.

1.2 Evolution of Text Summarisation Techniques

Text summarisation approaches have evolved substantially since their inception. Early summarisation methods in the 1990s relied on extractive techniques using statistical and frequency-based scoring methods such as TF-IDF. These models as often as possible delivered outlines that needed consistency and relevant stream since they chose key sentences from the content verbatim. Improved significance estimation and excess decrease were created by machine learning models and graph-based strategies like TextRank to overcome these constraints.

Abstractive summarization, in which models create rundowns by summarizing instead of reproducing the crude content, was made conceivable by the presentation of transformer-based structures, which totally changed the teach. BERTSUM[11], PEGASUS, GPT-2/3/4, and instruction-tuned[4] models like FLAN-T5 are striking turning focuses in this movement. These models significantly moved forward the created summaries' coherence, familiarity, and semantic understanding, which qualified them for down to earth utilization.

1.3 Motivation for Bibliometric Exploration

Given the fast development of summarization investigation, it is basic to chart its advancement all through time. A bibliometric examination offers a deliberate way to discover critical mechanical advancements, creating subjects, and striking donors. This proposition offers a data-driven point of view on the advancement of summarization investigate by applying devices like VOSviewer and RawGraphs to information accumulated from dependable databases like Web of Science. In expansion to recognizing authentic designs, this approach recognizes critical roads to encourage inquiry about.

1.4 Challenges in Domain-Specific Summarisation

Indeed with the progressions in demonstrate models, execution shifts by domain. The complexity of dialect structures, domain-specific lexicon, and the prerequisite for tall truthful precision make summarizing legitimate records especially troublesome. Due to complicated state developments, domain-specific lexicon, and the necessity for tall real precision, lawful record summarization presents special difficulties. Important inquiries concerning demonstrate generalizability and fine-tuning methods are brought up by this execution dissimilarity. Expanding the sum of the dataset or utilizing cross-domain pretraining are cases of routine procedures that do not continuously deliver picks up. Or maybe, there's mounting prove that making strides demonstrate execution for certain assignments requires hyperparameter alteration.

1.5 Research Objectives

Two roads of request are combined in this thesis:

1. To give a basic and bibliometric examination of content summarization investigate, spotting patterns, holes, and modern roads for the discipline.
2. Through deliberate hyperparameter tuning, the FLAN-T5[8] model's execution on the BillSum[10] dataset will be surveyed and made strides in comparison to BART-Large's pattern performance.

This proposal gives an in-depth understanding of the macro-level headway and micro-level challenges in content summarization by combining slant investigation with valuable demonstration optimization.

CHAPTER 2

RELATED WORK

2.1 Literature review

Ordinary studies, such as those conducted by El-Kassas et al. (2021)[16] and Gambhir & Gupta (2017)[17], concentrated for the most part on mechanical improvements in summarization calculations. They did not, be that as it may, give a quantitative appraisal of the state of the field. Later investigate has filled this vacuum by examining distribution patterns, co-authorship systems, and topical clusters in summarization considers utilizing bibliometric instruments like VOSviewer and RawGraphs.

By looking at nearly a thousand distributions from the Net of Science database, this proposal extends on earlier ponders and gives a data-driven and visual picture of the field's advancement. The investigation centers on modern themes counting domain-specific summarization, real-time summarization, and moral concerns like preference and visualization in expansion to more customary subjects like extractive and abstractive models. Master assessment and bibliometric investigation work together to supply a well-organized investigation guide.

2.2 Instruction Tuned Models

A powerful strategy for improving show generalization over NLP assignments is instruction tuning. Instruction tuning trains models on a huge collection of instruction-response pairings, empowering zero-shot and few-shot capabilities, in differentiate to conventional fine-tuning, which customizes a show for a specific work.

One well-known outline of an instruction-tuned transformer that exceeds expectations at summarizing, interpreting, and replying questions is the FLAN-T5 type. Such models perform conflictingly in domain-specific settings, such lawful or therapeutic summarization, however exceeding expectations at non specific errands.

Typically primarily since, especially when domain-specific semantics and dialect are included, instruction tweaking might not offer sufficient task-specific presentation.

2.3 Denoising Language Models

Denoising autoencoders are utilized in pre-training for models such as BART (Bidirectional and Auto-Regressive Transformer). Solid semantic understanding and creation capabilities are made conceivable by their capacity to reproduce the first content from mutilated input. Since BART employs autoregressive decoders and bidirectional encoders to reason over the whole input arrangement, it has beaten various pattern models in abstractive summarization.

Mental trips, on the other hand, can moreover cause denoising models to create content that's liquid but genuinely wrong. In high-stakes areas where truthful precision is significant, like lawful and restorative summarization, this gets to be a genuine concern. Hence, when utilizing such models for domain-specific errands, exhaustive evaluation and alteration are significant.

A careful survey of content summarization strategies is given in Figure 2.3.1, where they are separated into two essential categories: extractive and abstractive. In extractive outline, vital articulations or expressions are chosen straight from the first content, without any rewording. This category incorporates factual methods that gauge sentence pertinence utilizing word recurrence and graph-based strategies like TF-IDF, TextRank[18], and LexRank[19]. To decide which sections of the content to extricate, machine learning procedures for extractive summarization utilize classifiers like Gullible Bayes, Choice Trees, and SVMs in expansion to neural network-based models like RNNs, CNNs, and transformers.

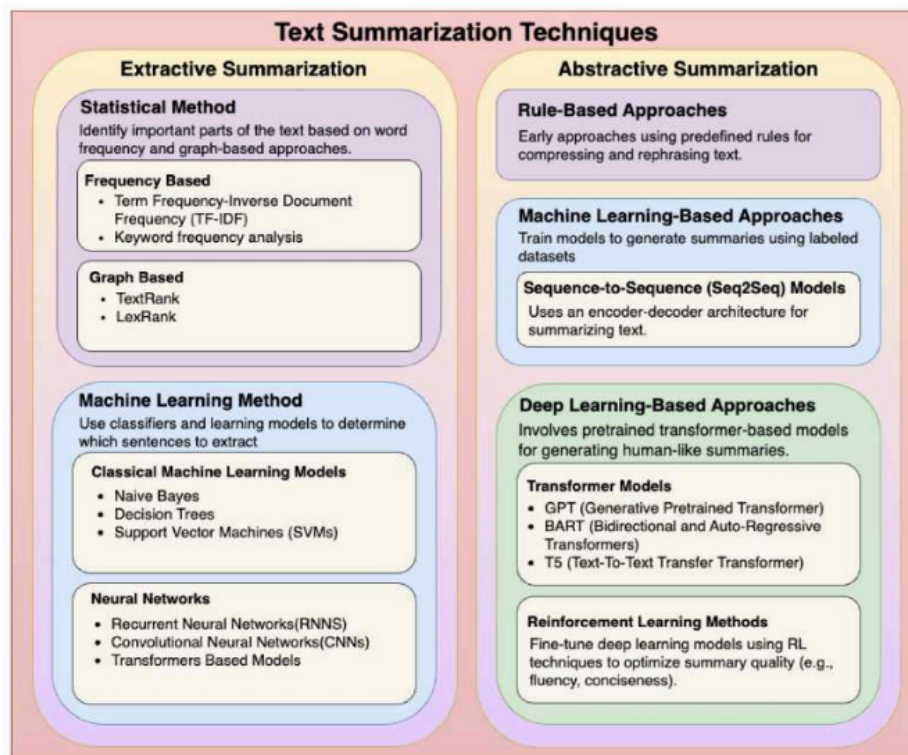


Figure 2.3.1 - An overview of Text Summarisation Techniques.

On the other hand, abstractive summarizing, which regularly requires substance creation and rethinking, produces modern sentences that express the most concepts of the first fabric. Rule-based strategies that utilize pre-established phonetic rules for content modifying and compression drop beneath this category. More advanced strategies depend on machine learning, particularly encoder-decoder models for sequence-to-sequence (Seq2Seq) models. By utilizing powerful pretrained transformer models like GPT, BART, and T5, which are planning to supply consistent and relevantly fitting outlines, profound learning-based methods encourage progress summarization capabilities. These models are also improved through the use of reinforcement learning techniques, which maximize output quality in terms of conciseness and fluency. All things considered, the picture depicts the variety and development of summarization techniques, from basic statistical models to complex neural and transformer-based systems.

2.4 Datasets

Two different kinds of datasets, each representing one of the research components, are used in this thesis: (1) text summarization corpora, particularly XSum and BillSum, for assessing and improving model performance in domain-specific summarization, and (2) a scientific publishing dataset from bibliographic databases for bibliometric study of summarization trends.

The Web of Science (WoS) Core Collection, a well-known bibliographic database renowned for its well curated and superior indexing of scientific articles, provided the data used for the trend analysis of text summarization research. The Core Collection includes the following citation indexes:

- SCI-EXPANDED (Science Citation Index Expanded)
- SSCI (Social Sciences Citation Index)
- A&HCI (Arts and Humanities Citation Index)
- CPCI-S (Conference Proceedings Citation Index – Science)

A carefully formulated topic search query was used:

²
 TS = (“Text summarization” OR “text summarizes” OR “text summarizing”
 OR “Text summarisation” OR “text summarises” OR “text summarising”)

This query was designed to capture all common and variant spellings of the term “text summarisation” across titles, abstracts, author keywords, and keywords plus. Results were filtered using a PRISMA flow (see Figure 2.4.1) to ensure precision and eliminate duplicates. WoS's unique identifier system (UT) helped maintain data integrity and ensured each publication was counted only once.

The final dataset included:

- 983 articles published from 2004 to September 2024
- 18,392 total citations (14,833 excluding self-citations)
- 338 journals, 2952 authors, and a collective H-index of 60

- All documents were in English, filtered by type to emphasize original research and scholarly reviews

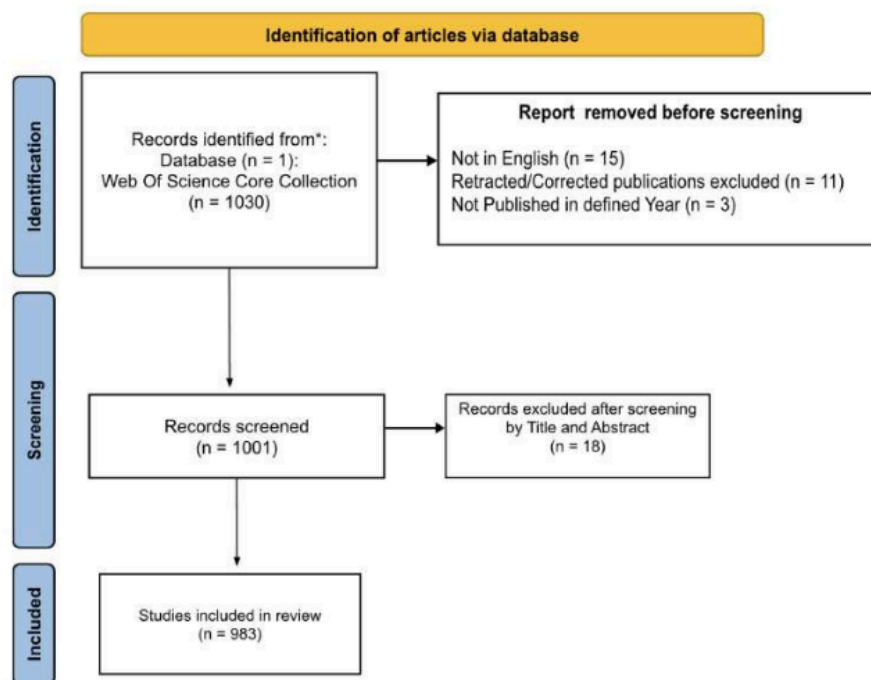


Fig. 2.4.1 - Prisma diagram for search query (the number of records found from each database searched is shown by the symbol "*").

This dataset provided the foundation for identifying thematic evolution, co-authorship networks, and emerging subfields in text summarisation.

To investigate model behavior in domain-specific summarisation, two prominent benchmark datasets were used:

The XSum dataset, created by the BBC, comprises single-sentence news summaries for over 200,000 articles. It was used to explore whether pretraining on highly abstractive summarisation data could enhance generalisation for other tasks. Each summary in XSum is designed to be maximally informative, often rephrasing the content entirely, thus making it well-suited for training generative models like FLAN-T5 and BART.

BillSum is a domain-specific dataset that includes human-written summaries of 22,000 California state and U.S. congressional bills. It illustrates how difficult it is to summarize legal papers due to their language and structural complexity. Due to the formal, technical language and requirement for factual correctness, BillSum serves as a rigorous benchmark for evaluating summarisation model adaptability in legal contexts.

Initial training and evaluation were performed on BillSum, while additional fine-tuning on XSum was explored to test whether cross-domain pretraining could enhance model performance. Despite the high quality of both datasets, performance improvements were realized only after targeted hyperparameter tuning, highlighting the need for tailored configurations in domain-specific applications.

CHAPTER 3

METHODOLOGY

3.1 Model Optimization Methodology

3.1.1 Model Selection

Two transformer-based models from the HuggingFace library were selected:
Model 1: FLAN-T5 Base (Instruction-tuned)
Model 2: BART-Large (Denoising pre-trained)
These models were chosen to evaluate performance disparities on legal-domain summarisation and to explore improvement strategies for underperforming models.

3.1.2 Dataset Description

The following summarisation datasets were used:
BillSum: A legal summarisation corpus of U.S. congressional bills and their summaries.
XSum: A highly abstractive summarisation dataset of BBC news articles.

3.1.3 Experimental Workflow

Step 1: Baseline Evaluation

The BillSum dataset was used to refine both models.
ROUGE-1 scores were computed to establish baseline performance.
BART-Large: 40.09
FLAN-T5 Base: 18.39

Step 2: Dataset Augmentation

Hypothesis: More training data improves performance.
Two attempts:
Pretraining FLAN-T5 on XSum followed by BillSum fine-tuning.
Training on an expanded BillSum dataset.

Outcome: Marginal or no performance gain (ROUGE[1] remained ~18–25).

Step 3: Hyperparameter Optimization

Hypothesis: Adjusting generation parameters would improve summarisation quality.

Parameters tuned:

num_beams = 6

generation_max_length = 300

Result:

FLAN-T5 Base's ROUGE-1 improved to 40.85, nearly matching BART-Large.

3.1.4 Evaluation Metrics

Training Loss and Validation Loss monitored per epoch. ROUGE-1 used for quantifying summary overlap between predictions and ground truth.

Seaborn[9] visualizations were created to illustrate performance changes.

3.2 Combined Analysis of Publication Trends, Citations, and Collaboration Networks

This section presents a consolidated view of publication structures, citation patterns, institutional impact, and publisher contributions in text summarisation research.

3.2.1 Annual Publication Growth and Document Types

Between 2004 and 2024, there has been steady growth in text summarisation research, culminating in a total of 983 publications. The majority (95.7%) are research articles, followed by reviews and conference papers. This indicates a strong foundation of peer-reviewed contributions to the field(Figure 3.2.1.1 and 3.2.1.2).

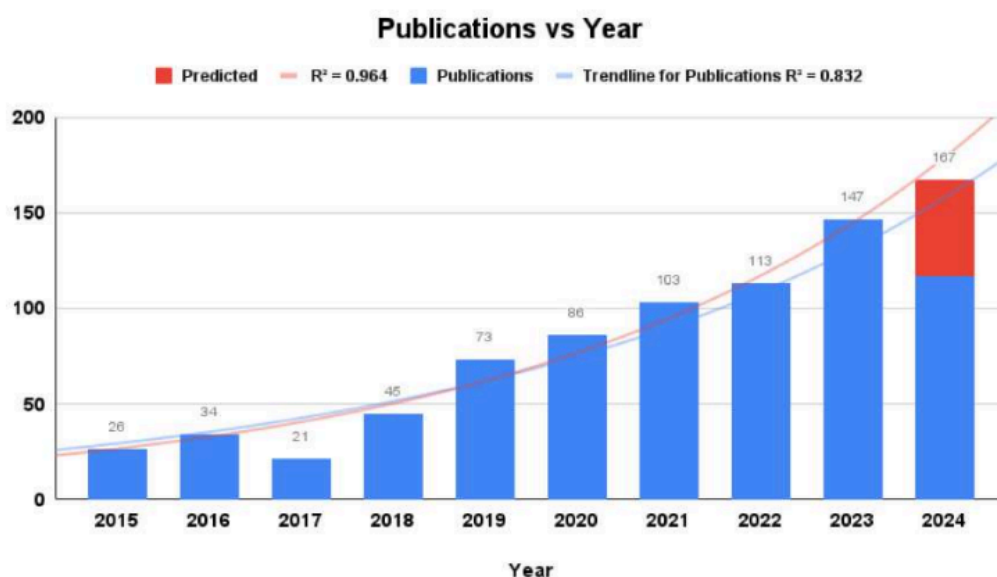


Fig. 3.2.1.1 - An annual publication growth graph illustrating the rise in publications from 2015 to 2024.

A thorough examination of the publishing categories of the 983 documents that were obtained from the Web of Science (WoS) database is shown in Figure 3.2.1.2. The data reveals that research articles constitute the vast majority, with 941 entries, accounting for approximately 95.72% of the total publications. This dominance highlights the field's strong focus on original research contributions. In addition to research articles, the dataset includes 41 review articles (4.17%), which provide comprehensive overviews of existing literature. There are also 25 early access papers (2.54%) and 40 conference proceedings (4.07%), reflecting the dissemination of ongoing and preliminary findings. A smaller proportion is made up of letters and other miscellaneous types, indicating a diverse but research-centric publication landscape.

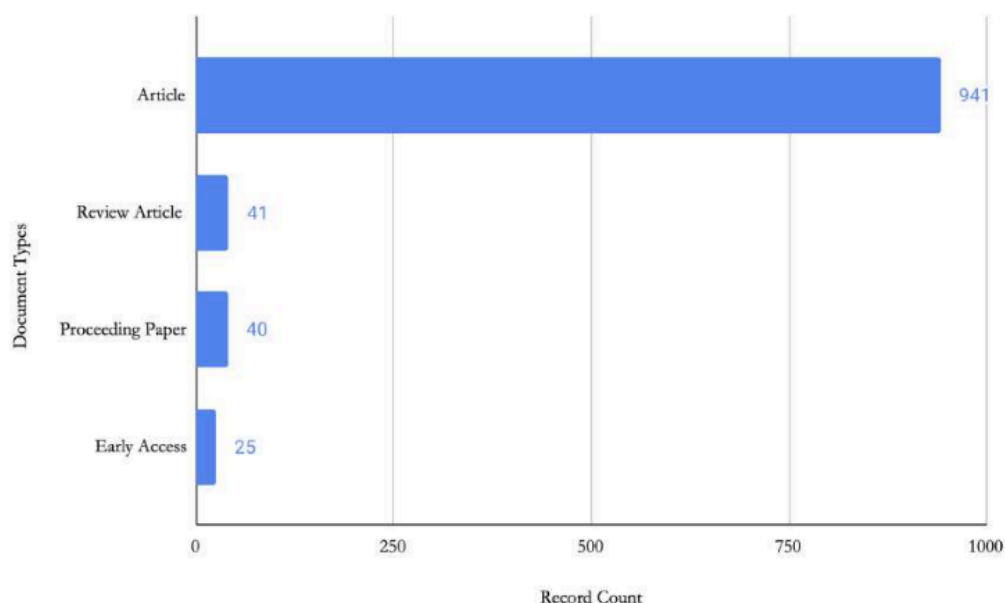


Fig. 3.2.1.2 - The top 10 categories in the publications.

3.2.2 Disciplinary and Geographic Trends

Electrical engineering, information systems, and artificial intelligence are the most active WoS categories. Geographically, the leading contributors are the USA (129), India (172), and China (215 publications). Among the most prominent are the Chinese Academy of Sciences and the Indian Institute of Technology (IIT) system (Figure 3.2.2.1).

Fig.-3.2.2.1 highlights the significant impact of various academic domains on the field of text summarisation. Leading the contributions is the field of Computer Science, particularly the subdomain of Artificial Intelligence, which accounts for 232 publications—approximately 38.86% of the total. A more thorough analysis reveals that the number of publications in computer science not only surpasses but also might overlap by about 955 articles with those in electrical and electronic engineering. Artificial Intelligence, Theory and Methods, Information Systems, Software Engineering, and Multidisciplinary Applications are some of the subfields in computer science that contribute to the literature. The interdisciplinary character and broad applicability of summarization research are further demonstrated by the 80 papers in telecommunications, 46 in multidisciplinary engineering, 34 in medical informatics, and 27 in linguistics.

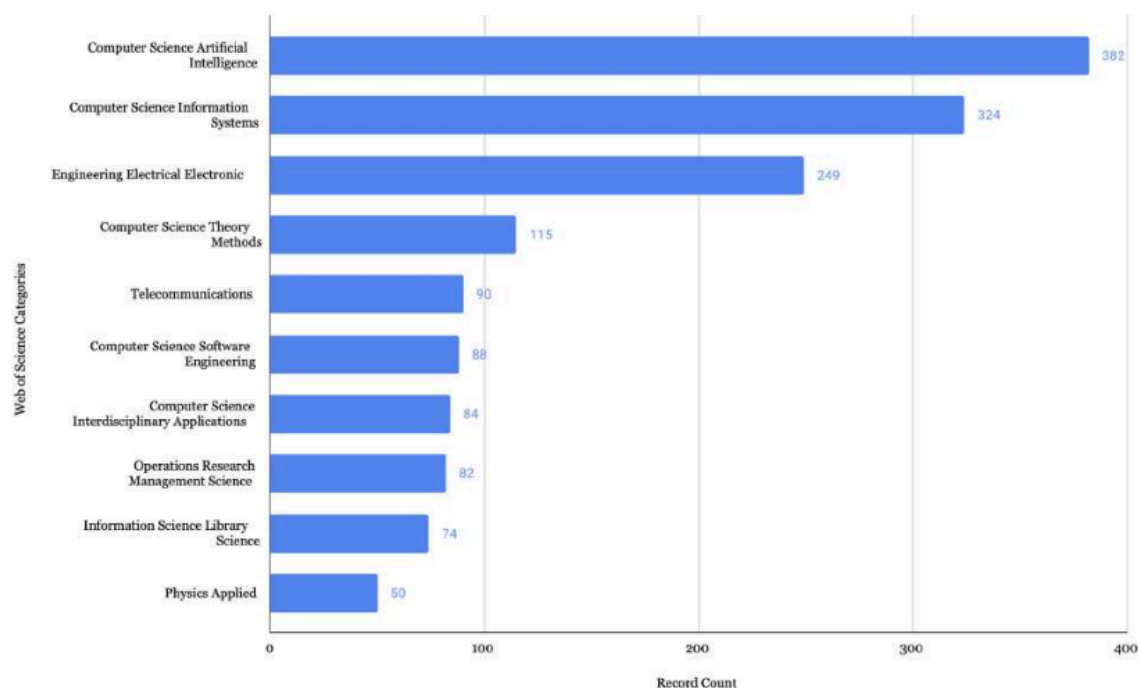


Fig. 3.2.2.1 - The top ten research directions in the articles indicate the main areas of attention.

3.3 Citation and Collaboration Metrics

According to citation analysis, the countries with the highest average number of citations per manuscript are the USA, China, and Italy. Institutions with a high citation strength and collaborative participation include Jawaharlal Nehru University and the University of Chinese Academy of Sciences. Prominent authors include Lloret E., Palomar M., and Devlin J(Figure 3.3.1, 3.3.2 and 3.3.3).

Fig.-3.3.2 illustrates the leading institutional contributors to text summarisation research based on publication count. With 31 publications, the Indian Institute of Technology (IIT) System comes in first, demonstrating its substantial contribution to the subject. Following closely behind are the National Institute of Technology (NIT) System with 28 publications and the Chinese Academy of Sciences with 29, demonstrating China and India as significant centers of research. The Egyptian Knowledge Bank (EKB) also demonstrates notable engagement, contributing 22 publications. Other institutions with substantial outputs include the University of Chinese Academy of Sciences (16), King Saud University and Universitat d'Alacant (15 each), the Institute of Computing Technology CAS (11), and both Islamic Azad

University and Universidad de Extremadura, each with 10 publications. Overall, the chart reflects a strong international presence in text summarisation research, with particularly high contributions from institutions in India, China, the Middle East, and parts of Europe.

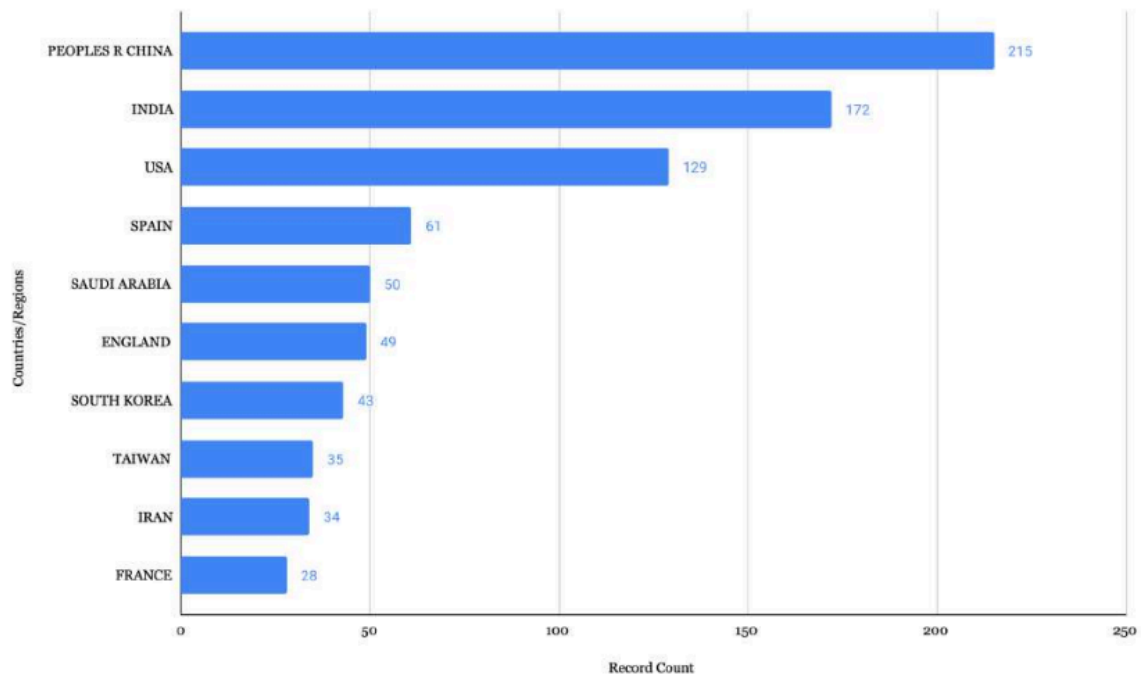


Figure 3.3.1 - The ten most productive locations.

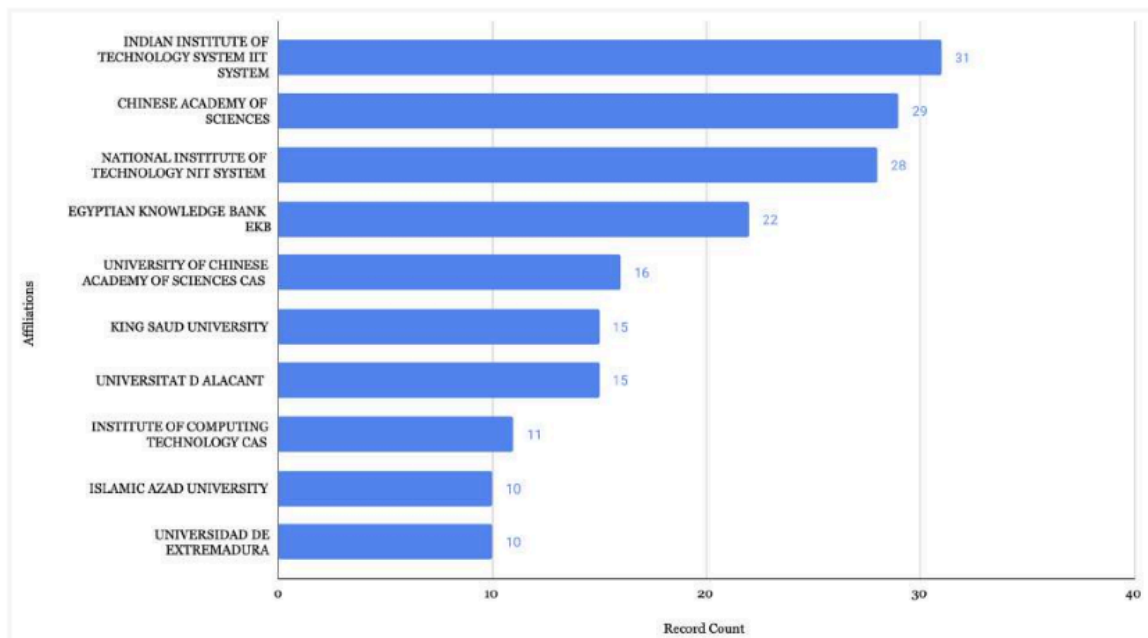


Figure 3.3.2 - The organizations that produce the greatest number of research publications.

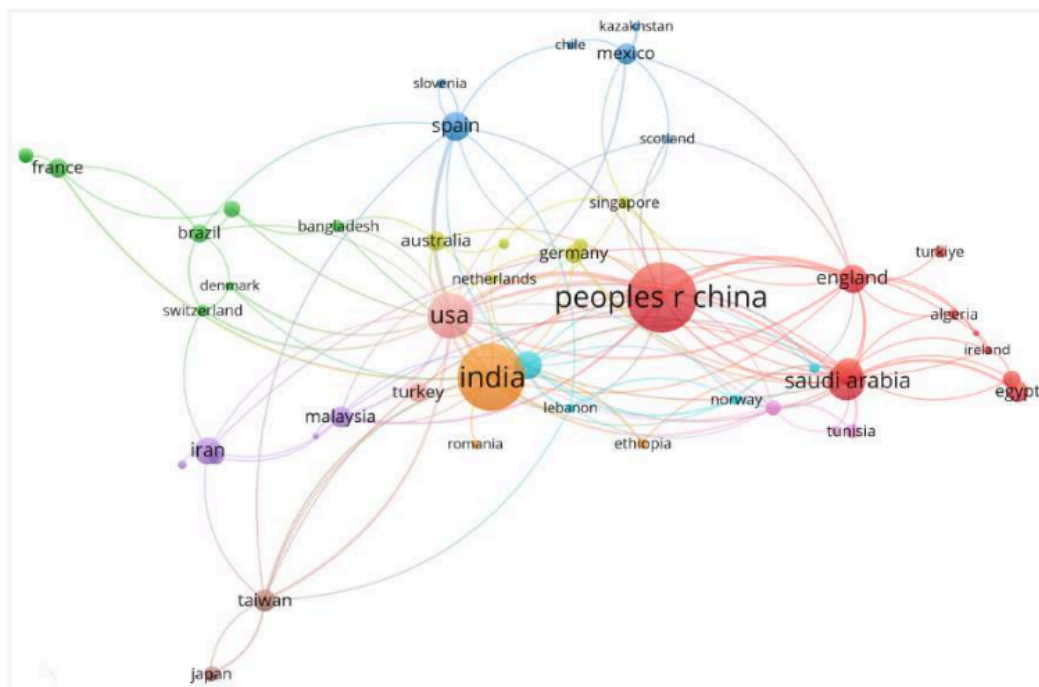


Figure 3.3.3 - The collaborative network of places.

3.4 Co-Authorship and Co-Citation Networks

Table 3.4.1 - Top researchers with the highest TLS ratings.

Rank	Authors	TLS	C
1	Nallapati, R	1565	252
2	Lin, CY	1513	284
3	See, A	1144	187
4	Mihalcea, R	953	154

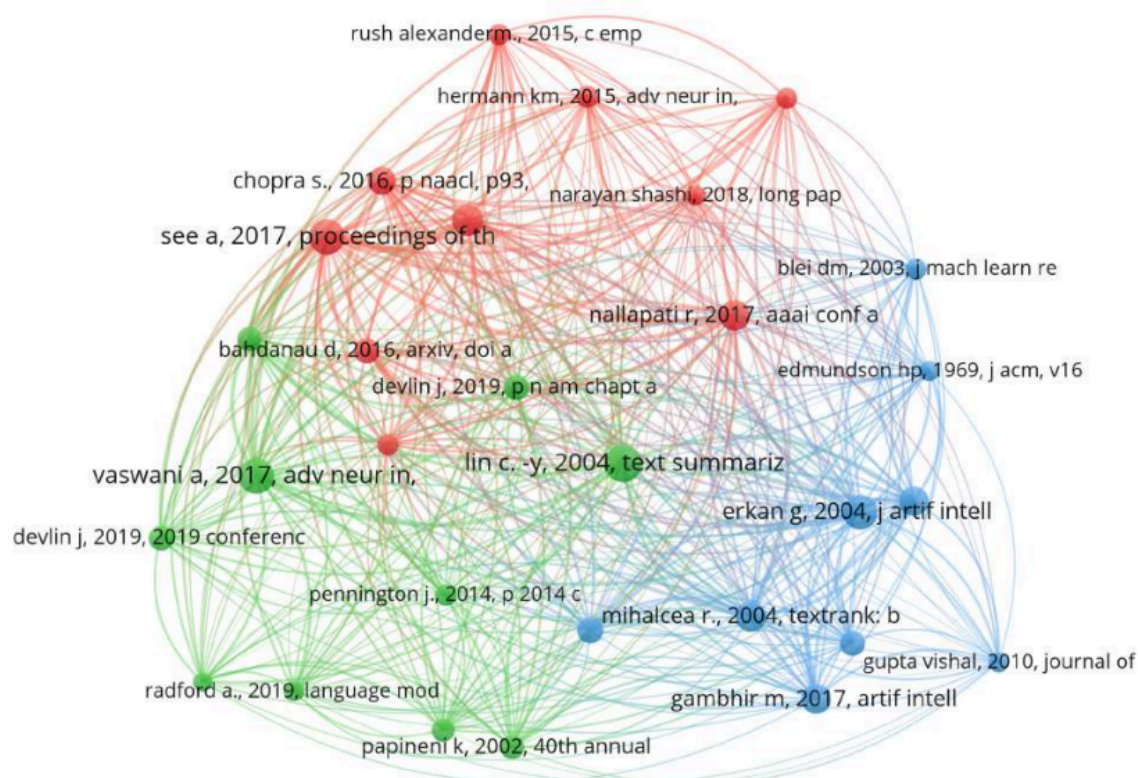


Fig. 3.4.1 - The network of co-citations for articles.

Using VOSviewer, co-authorship networks highlight clusters of collaboration between researchers and institutions. Scholars like Nallapati R. and Lin C.Y. exhibit

strong co-citation metrics. Influential papers include “Attention Is All You Need” (Vaswani et al.) and “BERT: Pre-training of Deep Bidirectional Transformers” (Devlin et al.).

3.5 Publisher Landscape

Chord diagrams generated using RawGraphs show that Springer, Elsevier, and IEEE have been the most active publishers from 2010 to 2024. Emerging contributors include MDPI and PeerJ Inc., indicating a diversification in publication venues.

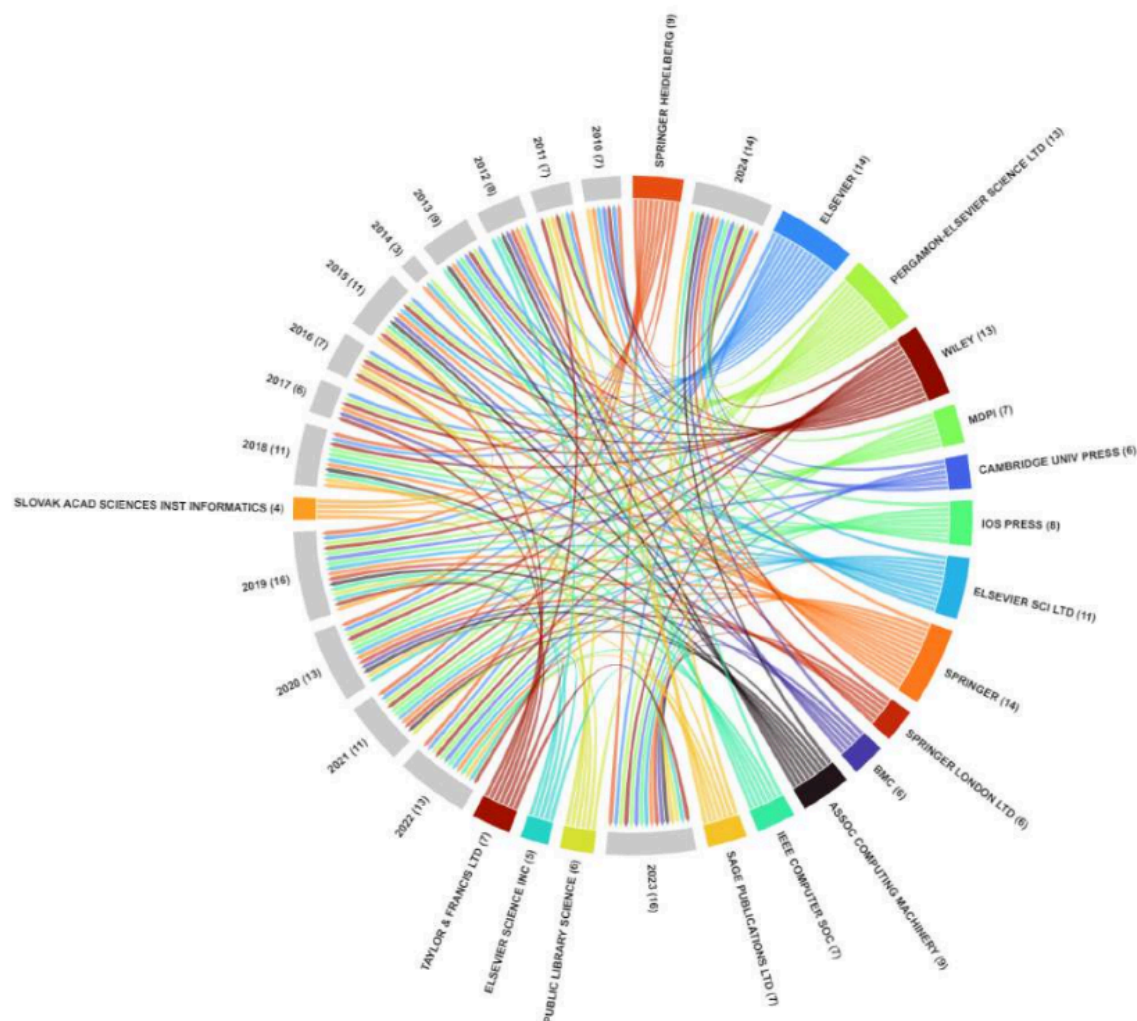


Fig. 3.5.1 - Chord diagram of publisher and publishing year.

CHAPTER 4

RESULTS

4.1 Experimental Results on Model Performance

This chapter presents tables and graphs illustrating ROUGE scores, training loss, and validation loss metrics based on the improvement strategies outlined in the methodology section. All visualizations were generated using the Seaborn Python library. Below is a brief explanation of the columns included in the tables.

1. Epoch: Shows the number of training cycles, with each epoch representing a full run of the training dataset.
2. Training Loss: Shows the error of the model on the training set; smaller values indicate a better match.
3. Validation Loss: Evaluates the model's capacity for generalization by reflecting its inaccuracy on unseen validation data.
4. ROUGE-1: Determines the content similarity between generated and reference summaries by measuring the unigram (single-word) overlap.

In Table 4.1.1, With training loss dropping from 1.4 to 1.08 and validation loss down from 1.4 to 1.3, BART-Large exhibits good performance. Additionally, the ROUGE-1 score somewhat improves, going from 39.8 to 40.0.

Table 4.1.1 - Bart-Large On Billsum With 8000 Rows In Train Set And 1000 Rows In Test Set

Epoch	Training Loss	Validation Loss	Rouge1
1	1.4711	1.430512	39.8935
2	1.2866	1.375137	39.9252
3	1.1315	1.372435	39.9512
4	1.0851	1.367797	40.0978

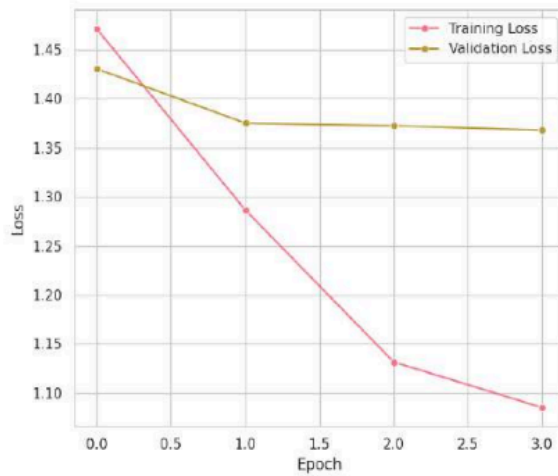


Fig. 4.1.1 - Loss graph for Table 4.1.1

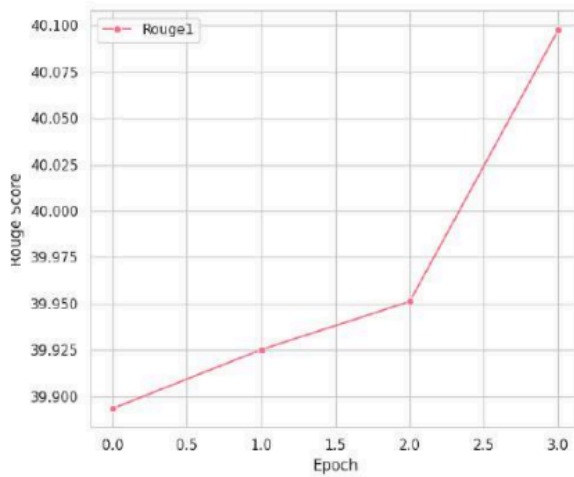


Fig. 4.1.2 - Rouge Score for Table 4.1.1

Table 4.1.2 demonstrates that FLAN-T5 performs worse than BART-Large¹⁶. The validation loss falls from 1.7 to 1.6, and the training loss falls from 1.8 to 1.6. A slight improvement is shown in the ROUGE-1 score, which rises from 18.2 to 18.3.

Table 4.1.2 - FLAN-T5-Base on BillSum with 8000 Rows in Train Set and 1000 Rows in Test Set

<i>Epoch</i>	<i>Training Loss</i>	<i>Validation Loss</i>	<i>Rouge1</i>
1	1.8348	1.740196	18.2941
2	1.8827	1.691753	18.2926
3	1.6701	1.674098	18.4369
4	1.65	1.670359	18.3968

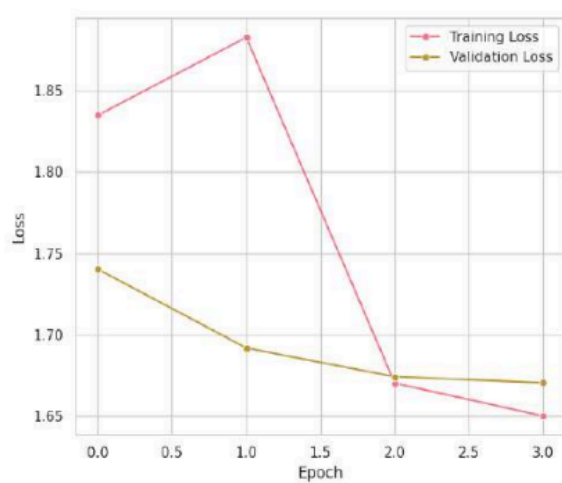


Fig. 4.1.3 - Loss graph for Table 4.1.2

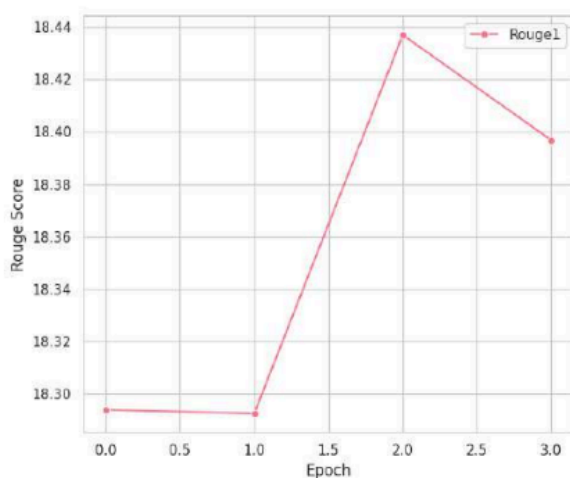


Fig. 4.1.4 - Rouge score graph for Table 4.1.2

In Table 4.1.3, FLAN-T5 shows a slight improvement in performance. The validation loss falls from 1.19 to 1.17, and the training loss somewhat decreases from 1.28 to 1.27. The ROUGE-1 score shows a minimal increase, rising from 24.9 to 25.0.

Table 4.1.3 - FLAN-T5-Base on BillSum, With 8000 Rows in Train Set and 1000 Rows in Test Set, After Training on XSum

<i>Epoch</i>	<i>Training Loss</i>	<i>Validation Loss</i>	<i>Rouge1</i>
1	1.2891	1.195633	24.9941
2	1.2783	1.184432	25.0385
3	1.2178	1.177504	25.0069
4	1.2701	1.174383	25.009

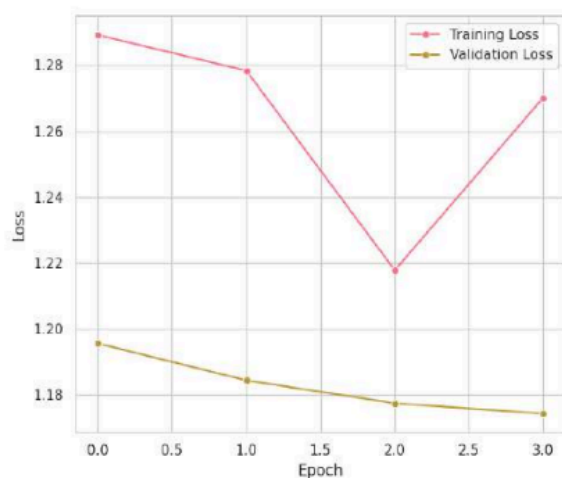


Fig. 4.1.5 - Loss graph for Table 4.1.3

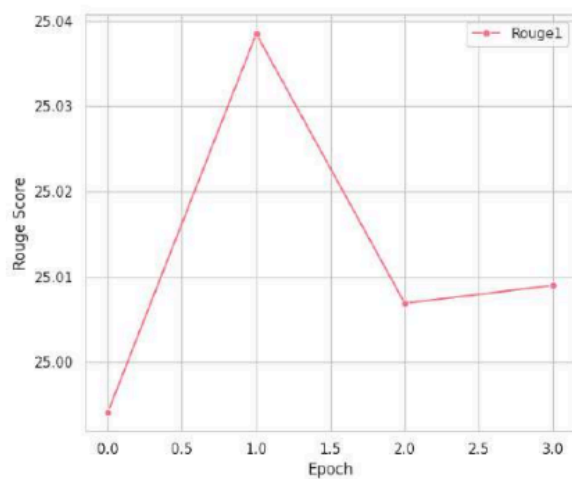


Fig. 4.1.6 - Rouge Score graph for Table 4.1.3

In Table 4.1.4, FLAN-T5 exhibits performance similar to that observed in Table II. While the validation loss falls from 1.8 to 1.7, the training loss decreases somewhat from 2.0 to 1.9. The ROUGE-1 score shows a modest increase from 17.9 to 18.1.

Table 4.1.4 - Flan-T5-Base on BillSum With 18.9K Rows in Training and 3.27K Rows in Test Set.

Epoch	Training Loss	Validation Loss	Rouge1
1	2.0935	1.870951	17.9862
2	1.9908	1.823429	18.1394
3	2.0161	1.802811	18.1797
4	1.9716	1.797551	18.1915

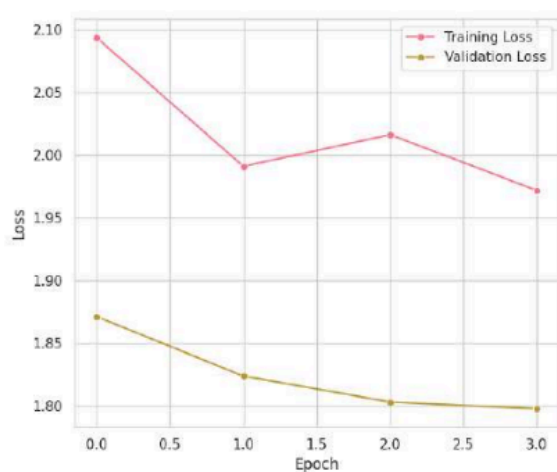


Fig. 4.1.7 - Loss graph for Table 4.1.4

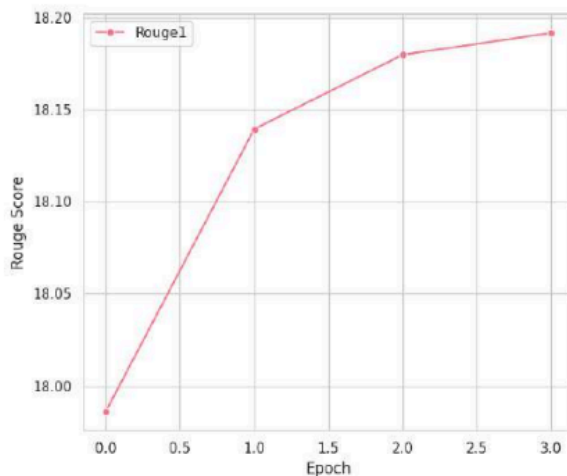


Fig. 4.1.8 - Rouge Score for Table 4.1.4

In Table 4.1.5, FLAN-T5 performs comparably to BART-Large as shown in Table I. The validation loss falls from 1.9 to 1.8, while the training loss falls from 2.1 to 2.0. The ROUGE-1 score significantly improves, rising from 39.1 to 40.8.

Table 4.1.5 - FLAN-T5-Base on BillSum After Hyperparameter Tuning

<i>Epoch</i>	<i>Training Loss</i>	<i>Validation Loss</i>	<i>Rouge1</i>
1	2.1837	1.926986	39.1855
2	2.126	1.88247	39.6472
3	2.0353	1.863689	41.1076
4	2.0528	1.857668	40.8503

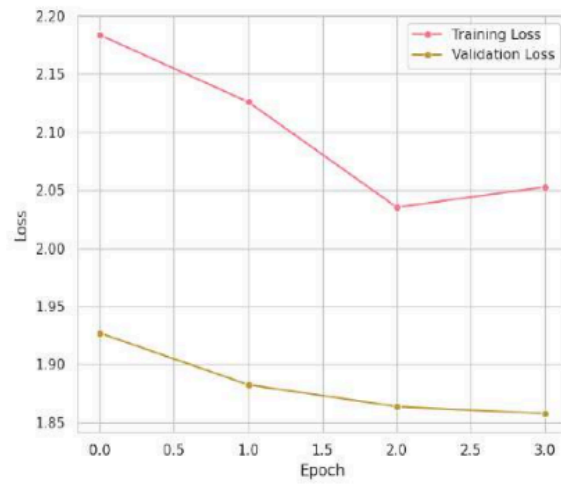


Fig. 4.1.9 - Loss graph for Table 4.1.5

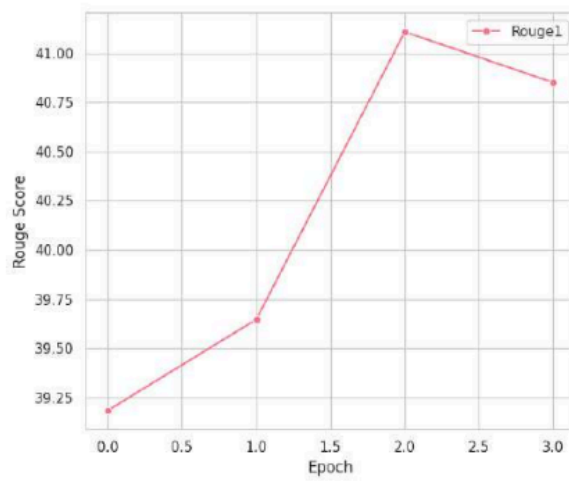


Fig. 4.1.10 - Rouge Score for Table 4.1.5

CHAPTER 5

CONCLUSION, FUTURE SCOPE, SOCIAL IMPACT

5.1 Conclusion

Through the integration of bibliometric analysis and practical testing on transformer-based models, this thesis offers a dual-perspective investigation of the developing subject of text summarization. The two-pronged approach has provided valuable insights into both the macro-level research landscape and the micro-level technical performance improvements in domain-specific summarisation tasks.

From a bibliometric standpoint, the field of text summarisation ²³ has witnessed significant growth over the past two decades, with a surge in research output particularly after 2015. High-impact papers, eminent institutions, and significant contributors were identified through the examination of 983 publications from the Web of Science database. Journals like Expert Systems with Applications and IEEE Access have been important publication sites, and nations including the US, China, and India have been significant players in this field. There has been a noticeable shift from extractive techniques like TextRank to sophisticated abstractive strategies that make use of transformer structures like BERT and GPT. Also, modern patterns that demonstrate a develop and differentiated consider space incorporate topic moves toward multimodal, multilingual, and real-time summarization.

In terms of experimentation, this thesis accesses how well two transformer models—FLAN-T5 Base and BART-Large—summarize data from the BillSum dataset. Agreeing to preparatory comes about, FLAN-T5 performed discernibly more awful than BART-Large. Negligible picks up were gotten when FLAN-T5's execution was endeavored to be moved forward by extending the dataset and doing more pretraining on the XSum dataset. Targeted hyperparameter optimization, namely beam width and generation length modifications, resulted in a significant improvement in performance, with ROUGE-1 scores increasing from 18.3 to 40.85, which essentially matched BART-Large. These discoveries cast question on the idea that more current models naturally perform way better than more seasoned ones without satisfactory tuning and highlight the importance of model-specific arrangement.

This proposition basically emphasizes how academic progression and real-world optimization are associated within the field of content summarization. It

affirms that both comprehending inquire about patterns and moving forward the fundamental advances can lead to noteworthy breakthroughs. For up and coming researchers and experts trusting to improve in this dynamic and critical field, the combined discoveries give a street outline.

5.2 Future Scope

There are several directions ² for future research in the developing field of text summarization, especially in improving model flexibility across various domains with little supervision. There are several directions for future research in the developing field of text summarization, especially in improving model flexibility across various domains with little supervision. Furthermore, utilizing multimodal data sources and cross-lingual transfer learning offers the possibility of developing more resilient and broadly applicable summarization systems, opening the door for context-aware and adaptive language models.

5.3 Social Impact

Better text summarizing models could have a significant impact on how people and organizations use information, encouraging accessibility and well-informed choices. Reliable summary can improve transparency, lessen information overload, and facilitate quicker understanding of important papers in the legal, medical, and governmental sectors. Such innovations promote inclusive technology growth and more equitable access to information by reducing the digital divide and enabling the spread of knowledge in under-resourced languages or geographical areas.

REFERENCES

- [1] ROUGE: A Package for Automatic Evaluation of Summaries. (Lin, 2004)
- [2] Chung, Hyung Won, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Yunxuan Li et al. "Scaling instruction-finetuned language models." *Journal of Machine Learning Research*, 25(70), 2024, pp. 1–53.
- [3] Raffel, Colin, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. "Exploring the limits of transfer learning with a unified text-to-text transformer." *Journal of Machine Learning Research*, 21(140), 2020, pp. 1–67.
- [4] Zhang, Shengyu, Linfeng Dong, Xiaoya Li, Sen Zhang, Xiaofei Sun, Shuhe Wang, Jiwei Li et al. "Instruction tuning for large language models: A survey." *arXiv preprint arXiv:2308.10792* (2023).
- [5] Lewis, Mike, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension." *Annual Meeting of the Association for Computational Linguistics*, 2019.
- [6] Liu, Xueqing, and Chi Wang. "An empirical study on hyperparameter optimization for fine-tuning pre-trained language models." *arXiv preprint arXiv:2106.09204* (2021).
- [7] Ülker, M., and Özer, A. B. "The Bart-based Model for Scientific Articles Summarization." *Journal of Universal Computer Science*, 30(13), 2024, pp. 1807–1828. <https://doi.org/10.3897/jucs.115121>
- [8] Abdul Ghafoor Etemad, Ali Imam Abidi, and Megha Chhabra. "Fine-Tuned T5 for Abstractive Summarization." *International Journal of Performability Engineering*, 17(10), 2021, pp. 900–906.
- [9] Waskom, M. L. "Seaborn: Statistical data visualization." *Journal of Open Source Software*, 6(60), 2021, Article 3021. <https://doi.org/10.21105/joss.03021>
- [10] Kornilova, Anastassia, and Vladimir Eidelman. "BillSum: A Corpus for Automatic Summarization of US Legislation." In *Proceedings of the 2nd Workshop on New Frontiers in Summarization*, 2019, pp. 48–56. Hong Kong, China: Association for Computational Linguistics.

- [11] Zhang, Tianyi, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. "BERTScore: Evaluating text generation with BERT." *arXiv preprint* arXiv:1904.09675 (2019).
- [12] Hu, Edward J., Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. "LoRA: Low-Rank Adaptation of Large Language Models." *International Conference on Learning Representations (ICLR)*, 1(2), 2022, Article 3.
- [13] Wolf, Thomas, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac et al. "HuggingFace's Transformers: State-of-the-Art Natural Language Processing." *arXiv preprint* arXiv:1910.03771 (2019).
- [14] Narayan, Shashi, Shay B. Cohen, and Mirella Lapata. "Don't Give Me the Details, Just the Summary! Topic-Aware Convolutional Neural Networks for Extreme Summarization." *arXiv preprint* arXiv:1808.08745 (2018).
- [15] Kaushal, Anukriti, Chia-Chen Lin, Rishabh Chauhan, and Rajeev Kumar. "Charting the Growth of Text Summarisation: A Data-Driven Exploration of Research Trends and Technological Advancements." *Applied Sciences*, 14(23), 2024, Article 11462. <https://doi.org/10.3390/app142311462>
- [16] El-Kassas, W.S.; Salama, C.R.; Rafea, A.A.; Mohamed, H.K. Automatic text summarization: A comprehensive survey. *Expert Syst. Appl.* **2021**, *165*, 113679.
- [17] Gambhir, M.; Gupta, V. Recent automatic text summarization techniques: A survey. *Artif. Intell. Rev.* 2017, *47*, 1–66.
- [18] Mihalcea, R.; Tarau, P. TextRank: Bringing order into text. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, Barcelona, Spain, 25–26 July 2004; pp. 404–411.
- [19] Erkan, G.; Radev, D. LexRank: Graph-based lexical centrality as salience in text summarization. *J. Artif. Intell. Res.* 2004, *22*, 457–479.

Thesis

ORIGINALITY REPORT

16%
SIMILARITY INDEX

15%
INTERNET SOURCES

%
PUBLICATIONS

11%
STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Delhi Technological University Student Paper	7%
2	www.mdpi.com Internet Source	2%
3	123dok.net Internet Source	1%
4	www.collectionscanada.ca Internet Source	1%
5	Submitted to Gisma University of Applied Sciences GmbH Student Paper	<1%
6	Submitted to King's College Student Paper	<1%
7	f1000research.com Internet Source	<1%
8	Submitted to Jawaharlal Nehru University (JNU) Student Paper	<1%
9	Submitted to dtusimilarity Student Paper	<1%
10	mdpi-res.com Internet Source	<1%
11	text-id.123dok.com Internet Source	<1%
12	www.spiedigitallibrary.org Internet Source	<1%
13	kipdf.com Internet Source	<1%
14	opus4.kobv.de Internet Source	<1%
15	export.arxiv.org Internet Source	<1%

16	aclanthology.org Internet Source	<1 %
17	arxiv.org Internet Source	<1 %
18	datafield.dev Internet Source	<1 %
19	johal.in Internet Source	<1 %
20	open.library.ubc.ca Internet Source	<1 %
21	reposit.haw-hamburg.de Internet Source	<1 %
22	www.docstoc.com Internet Source	<1 %
23	www.newbank.sit.internal.pwc.in Internet Source	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		