Automated Research Paper Summarization with Multiple Model and Accessibility Enhancements

A.I.E Wijesooriya
(MS24022774)

A THESIS
SUBMITTED TO
SRI LANKA INSTITUTE OF INFORMATION TECHNOLOGY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE IN INFORMATION TECHNOLOGY

December 2025

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Mrs. Anjalie Gamage

Approved for MSc. Research Project:

MSc in IT Programme Co-ordinator, SLIIT

Approved for MSc:

Head of Graduate Studies, FoC, SLIIT

DECLARATION

This is to certify that the work is entirely my own and not of any other person, unless explicitly acknowledged (including citation of published and unpublished sources). The work has not previously been submitted in any form to the Sri Lanka Institute of Information Technology or to any other institution for assessment for any other purpose.

Sign: 

(A.I.E Wijesooriya)

Date: ……03.01.2026……

ABSTRACT

Automated Research Paper Summarization with Multi-Model and Accessibility Enhancements

Achini Wijesooriya
MSc. in Information Technology
Supervisor: Mrs. Anjalie Gamage
December 2025

The number of research papers published each year is growing at an overwhelming pace, making it difficult for students, researchers, and professionals to keep up with new knowledge. Existing summarization tools can help, but most of them rely on large models like GPT, Pegasus, or BERT, which need powerful hardware and constant internet access. This limits their use, especially in low-resource or offline environments. This work introduces a novel framework for Automated Research Paper Summarization that employs a multi-model hybrid pipeline, integrating both extractive and abstractive strategies. Unlike resource intensive models, this approach emphasizes lightweight architectures, enabling efficient performance even in low-resource settings while preserving summary quality. To further enhance usability, the system includes keyword extraction modules that highlight central concepts and accessibility features such as text-to-speech, supporting users with visual or cognitive challenges. A distinctive feature of this framework is its section-wise summarization output, which mirrors the logical flow of research papers allowing users to quickly access context, methodology, findings, or conclusions as needed. System performance is assessed through standard metrics like ROUGE and BLEU, complemented by qualitative evaluations of readability, informativeness, and coherence. By avoiding full dependence on large, pre-built models such as GPT or Pegasus, this work prioritizes component level innovation, offline functionality, and greater privacy, making it adaptable across diverse use cases. The study advances the field of scientific summarization by offering a practical, modular, and accessible tool that supports knowledge discovery and management in research intensive domains.

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to my supervisor, Ms. Anjalie Gamage, for their continuous guidance, encouragement, and support throughout the course of this research. Their insightful feedback and patience have been invaluable in shaping both the direction and the quality of this thesis.

I am also sincerely thankful to the lecturers of Faculty of Graduate Studies and Research department, whose teaching and advice have broadened my academic perspective and helped me strengthen this study. I also wish to thank my friends, both in the university and at my workplace, for their constant encouragement, motivation, and support during this journey.

A very special thanks goes to my family, whose love, understanding, and unwavering belief in me have been my greatest source of strength.

Finally, I would like to acknowledge all the authors and researchers whose works I have referred to in this thesis. Their valuable contributions laid the foundation for this research and provided me with the knowledge and direction needed to move forward.

TABLE OF CONTENTS

DECLARATION ii
ABSTRACT iii
ACKNOWLEDGEMENT iv
TABLE OF CONTENTS v
List of Figures vii
List of Tables viii
Chapter 1 – Introduction 1
1.1 Background of Study 1
1.2 Problem Statement 2
1.3 Research Objectives 3
1.3.1 Main Objective 3
1.3.2 Sub Objectives 3
1.4 Research Questions 4
1.5 Significance of Study 5
1.6 Scope and Limitations 7
1.7 Thesis Organization 10
Chapter 2 - Literature Review 11
2.1 Overview of Text Summarization 11
2.2 Extractive Summarization Approaches 11
2.3 Abstractive Summarization Approaches 12
2.4 Hybrid Summarization Approaches 13
2.5 Transformers and LLM-Based Summarization 13
2.6 Accessibility in Summarization Tools 14
2.7 Evaluation Metrics (ROUGE, BLEU, etc.) 15
2.8 Research Gaps 15
Chapter 3 – Methodology 17
3.1 Research Design 17
3.2 Datasets and Preprocessing 17
3.3 Model Architectures Considered 18
3.4 Proposed Hybrid Summarization Framework 19
3.5 Implementation of Accessibility Features 19
3.6 Section-Wise Summarization Workflow 20
3.7 Tools and Environment 20
3.8 Evaluation Strategy 21

3.9 Offline Deployment Strategy 22
Chapter 4 - System Design and Implementation 23
4.1 System Architecture 23
4.2 Extractive Summarization 23
4.3 Abstractive Summarization 24
4.4 Hybrid Pipeline and Refinements 24
4.5 Keyword Extraction and Visualization 25
4.6 Accessibility Features (TTS, Section Summaries) 26
4.7 Offline Tool Implementation 27
Chapter 5 - Results and Evaluation 28
5.1 Quantitative Evaluation (ROUGE, BLEU Scores) 28
5.2 Qualitative Analysis of Summaries 29
5.3 Comparison with Existing Models 29
5.4 Ablation Studies (Impact of Individual Components) 31
5.5 Key word Extraction 32
5.6 Section-Wise Summarization Outputs 34
5.7 Offline Application Development 35
Chapter 6 – Discussion 39
6.1 Key Findings 39
6.2 Implications for Academic Research 40
6.3 Strengths and Limitations of the Proposed Framework 41
6.4 Comparison with Literature 44
Chapter 7 - Conclusion and Future Work 46
7.1 Future Enhancements 46
References 48
Appendix 51
Appendix A 51
Appendix B 67

List of Figures

Figure 3.1 System Architecture..... 17
Figure 3.2 Data Set..... 19
Figure 4.1 Evaluation of different models.....28
Figure 5.2 Example side-by-side comparison of system-generated vs. gold-standard abstract.....29
Figure 5.3 Comparison with other models. 30
Figure 5.4 Ablation study results.....31
Figure 5.5 Keyword extraction(test1).....32
Figure 5.6 Key Word Extraction (Test 2).....33
Figure 5.7 COSINE Methodology Key sentence extraction.....33
Figure 5.8 NORM Methodology Key sentence extraction 33
Figure 5.9 Section Extraction (test1).....34
Figure 5.10 Example of section-wise summaries generated for a paper 34
Figure 5.11 Web Application Interface 35
Figure 5.12 PDF Upload Section.....37
Figure 5.13 Research paper replace and remove options 376
Figure 5.14 Set parameters.....376
Figure 5.15 Select full paper summery.....376
Figure 5.16 Section wise summary 387
Figure 5.17 Display keywords.....388

List of Tables

Table 1 Uniqueness with other tools 6
Table 2 Comparative evaluation of summarization models 30
Table 3 Comparing Strengths and Limitations of the Proposed Framework 43