AUTOMATIC GENERATION OF RESEARCH PAPER ABSTRACTS USING DEEP-HYBRID MODELS

R.P.D. Kumarasinghe

219354V

Master of Science in Computer Science

Department of Computer Science & Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

February 2025

AUTOMATIC GENERATION OF RESEARCH PAPER ABSTRACTS USING DEEP-HYBRID MODELS

R.P.D. Kumarasinghe

219354V

Thesis submitted in partial fulfillment of the requirements for the degree
Master of Science in Computer Science

Department of Computer Science & Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

February 2025

DECLARATION

I declare that this is my own work and this Thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date: 28/02/2025

The supervisor should certify the Thesis with the following declaration.

The above candidate has carried out research for the Master of Science in Computer Science Thesis under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Nisansa de Silva

Signature of the Supervisor: Digitally signed by Nisansa de Silva Date: 2025.02.28 13:09:33 +05'30' Date: 28/02/2025

DEDICATION

I dedicate this to all who strive to seek knowledge, believing that every step forward is a step towards a better world.

ACKNOWLEDGEMENT

I would like to express my deepest gratitude to my research supervisor, Dr. Nisansa de Silva, for their unwavering guidance, support, and encouragement throughout the course of this thesis. Their insightful advice, constructive feedback, and immense patience have been invaluable in shaping the direction and quality of this research. I am deeply thankful for the time and effort they dedicated to helping me grow academically and professionally, and for always challenging me to think critically and aim for excellence.

Lastly, I extend my appreciation to my family and friends for their constant encouragement and understanding during this challenging process. Their support made this accomplishment possible.

ABSTRACT

Condensing important information into a summary is crucial for readers navigating lengthy documents. In the context of research papers, the **abstract** serves as a concise overview of the study. This thesis focuses on enhancing research paper summarization by introducing a novel section-wise relevance matrix. To address the token size limitations of Large Language Models (LLMs) , such as GPT-Neo, we developed a two-fold approach. First, we employed extractive summarization to condense lengthy texts into key sentences, followed by the application of abstractive summarization to generate coherent and concise summaries from these extracts.

Our approach, combining both extractive and abstractive techniques, leverages section-wise involvement ratios, with particular attention to the abstract section, improving the accuracy and quality of generated summaries. We introduced a pioneering dataset of research papers organized into sections, which plays a crucial role in this summarization process. Experimental results demonstrated that our method produces high-quality summaries while effectively overcoming token limitations, offering significant potential for summarizing long documents in low-resource and cost-effective environments.

However, challenges arise when section-wise segmentation is unclear, impacting the accuracy of summaries. This research underscores the need for further refinements and offers a promising framework for enhancing summarization techniques, benefiting researchers, educators, and information seekers alike.

Keywords: NLP, ATS, Summarization, Text Generation, LLM

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
BLEU	Bilingual Evaluation Understudy
CPUs	Central Processing Units
GPUs	Graphics Processing Units
I/O	Input/Output
LCS	Longest Common Subsequence
LLMs	Large Language Models
LSA	Latent Semantic Analysis
NLP	Natural Language Processing
ROUGE	Recall-Oriented Understudy for Gisting Evaluation
TPUs	Tensor Processing Units