

INERA – INTELLIGENT EMAIL REPLY ASSISTANT TO REDUCE EMAIL OVERLOAD BASED ON DEEP LEARNING

Dinuka Malinda Samarasinghe

208548D

Thesis/Dissertation submitted in partial fulfillment of the requirements for the degree Master of
Science in Artificial Intelligence

Department of Computational Mathematics
Faculty of Information Technology

University of Moratuwa
Sri Lanka

May 2023

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.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis/dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature:

Date: 28/06/2023

The above candidate has carried out research for the Masters thesis/dissertation under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of the Supervisor: Dr. Thushari Silva

ABSTRACT

Email overload is a pressing concern in modern workplaces and businesses, as it hampers productivity and efficiency. Repetitive tasks, such as responding to customer inquiries, contribute to this escalating issue. Previous literature has investigated various methods to mitigate email overload, including email prioritization, structuring, missing attachment prediction, reply suggestion, keyword summarization, and task scheduling. Despite these efforts, existing solutions exhibit notable limitations, such as reliance on manual user validation and generating responses that lack comprehensiveness and context. In this research, we developed InERA – Intelligent Email Reply Assistant, a deep learning-based solution designed to alleviate email overload. InERA generates comprehensive, context-aware responses to general inquiries by leveraging insights from previous similar emails handled by the user. The system employs a classification module to identify responses that match the current email intent, utilizing natural language processing techniques and machine learning algorithms, such as k-means clustering, Multi-Layer Perceptron classifier, and LSTM model. InERA generates meaningful replies by considering previous user responses and email thread information, ultimately reducing the need for manual intervention. The thesis is structured with chapters on Introduction, Literature Review, Technology Adoption, Approach, Design, Implementation, Evaluation, and Conclusion. These chapters detail the development, implementation, and evaluation of InERA, including its user interface design, integration with email clients, and evaluation through user testing and performance metrics. By implementing InERA, organizations can alleviate the burden of email overload, significantly improving productivity and fostering more efficient communication. Furthermore, this research highlights the broader implications and potential future enhancements of the automatic email reply assistance system in various industries and academic fields, emphasizing ethical considerations and privacy concerns.

Keywords: Email Overload, Deep learning, Feature Extraction, Unsupervised Learning

DEDICATION

To my loving mother, Renuka Wanigasinghe, who have always believed in me and supported my dreams, and to my wife, Eshani Hansamali, who have been a constant source of encouragement and inspiration. This thesis is a testament to your unwavering love and faith in me.

To my dearest friends, who have been by my side throughout this journey, providing laughter, comfort, and companionship. Your friendship has made this experience all the more rewarding. And to all those who have inspired and motivated me to pursue my passion for research and knowledge, this work is dedicated to you.

ACKNOWLEDGEMENTS

My supervisor, Dr. Thushari Silva, has provided me with tremendous advice, inspiration, and support during the course of my project, and I would like to extend my sincere gratitude to her. This thesis' completion would not have been possible without their knowledge, understanding, and mentorship. I am also grateful to the members of the Department of Computational Mathematics at University of Moratuwa for providing a stimulating and supportive environment during my studies. Special thanks go to my fellow researchers and colleagues, whose camaraderie and insightful discussions have been a source of inspiration and motivation.

I would like to extend my appreciation to the external collaborators and industry partners who contributed their time, resources, and knowledge to this research. Their input has significantly enriched the quality and practicality of the study. I owe a debt of gratitude to my family for their consistent love, patience, and support throughout my academic career. I dedicate this work to them since their confidence in me has been an ever-present source of strength for me. Also, I would want to express my gratitude to my friends for their assistance and for being present throughout both the difficult and gratifying parts of our research. Your friendship has made this experience truly enjoyable and memorable.

TABLE OF CONTENT

CHAPTER 1.....	1
INTRODUCTION	1
1.1 Prolegomena	1
1.2 Objectives	2
1.3 Background and Motivation	2
1.4 Problem definition	3
1.5 InERA to reduce email overload based on Deep Learning	3
1.6 Structure of the thesis	4
1.7 Summary	4
CHAPTER 2.....	6
LITERATURE REVIEW ON EMAIL REPLY ASSISTANCE SYSTEMS	6
2.1 Introduction	6
2.2 Early developments of email reply assistance systems	6
2.3 Recent development of email reply assistance systems.....	7
2.4 Analysis of literature	8
2.5 Problem definition	9
2.6 Summary	9
CHAPTER 3.....	10
TECHNOLOGY ADOPTED.....	10
3.1 Introduction	10
3.2 Feature Extraction Techniques	11
3.3 Intent Classification Algorithms	12
3.4 Unsupervised Learning Techniques	17
3.5 Deep Learning Algorithms.....	19
3.6 Programming Languages and Tools	21
3.7 Deep Learning Libraries and Frameworks.....	22
3.8 Natural Language Processing Libraries	24
3.9 Data Collection and Preprocessing Tools.....	25
3.10 Machine Learning Libraries and Algorithms	27
3.11 Evaluation and Visualization Tools.....	29
3.12 Email Client Integration and User Interface Development	31
3.13 Hardware and Computational Resources	33
3.14 Conclusion.....	35
CHAPTER 4.....	37
APPROACH TO INTELLIGENT EMAIL REPLY ASSISTANT SYSTEM.....	37
4.1 Introduction	37

4.2	Hypothesis of Intelligent Email Reply Assistant	37
4.3	Input.....	37
4.4	Process	39
4.5	Output.....	42
4.6	Features	43
4.7	Users	45
4.8	Summary	46
CHAPTER 5.....		47
DESIGN		47
5.1	Introduction	47
5.2	High-level system overview	47
5.3	Description of the dataset used for training and validation	49
5.4	Choice of Machine Learning Models.....	52
5.5	Description of the features extracted from the input data	59
5.6	Training and Optimization	61
5.7	Loss function and optimization algorithm selection	62
5.8	Selection of evaluation metrics.....	69
5.9	Description of the user interface (UI) for InERA	71
5.10	Overview of the process for integrating the system with popular email clients	74
5.11	Summary of the design chapter.....	76
CHAPTER 6.....		78
IMPLEMENTATION		78
6.1	Introduction	78
6.2	Software and Tools	79
6.3	Data Collection and Preprocessing Implementation	80
6.4	Feature Engineering Implementation	82
6.5	K-Means Clustering Implementation	84
6.6	Multi-Layer Perceptron Model Implementation.....	86
6.7	Deep Learning Model Implementation.....	88
6.8	Training, Optimization, and Evaluation Implementation.....	91
6.9	User Interface Implementation	93
6.10	Integration with Email Clients Implementation.....	97
6.11	Testing and Debugging.....	99
6.12	Conclusion.....	102
CHAPTER 7.....		103
EVALUATION		103
7.1	Introduction	103
7.2	Overview of the evaluation process and its significance	104
7.3	Details of the dataset used for evaluation	104

7.4	Evaluation Machine Learning Models.....	105
7.5	Comparison of the system's performance with existing state-of-the-art solutions	112
7.6	Evaluation of user opinion and usability testing.....	114
7.7	Identification of the system's limitations and challenges faced during evaluation	117
7.8	Conclusion.....	119
CHAPTER 8.....		121
CONCLUTION AND FURTHER WORK		121
8.1	Introduction	121
8.2	Summary of the Research	121
8.3	Key Contributions	123
8.4	Limitations of the Study	125
8.5	Future Work.....	126
8.6	Broader Implications.....	127
8.7	Conclusion.....	128
REFERENCE		129
APPENDICES		131

LIST OF FIGURES

Figure 5. 1: High level architecture diagram	48
Figure 5. 2 : Dataset distribution.....	50
Figure 6. 1 : Pseudocode for LSTM model implementation	90
Figure 6. 2 : InERA system UI for email inbox.....	95
Figure 6. 3 : InERA system UI for email cluster distribution.....	96
Figure 6. 4 : InERA system configuration UI.....	96
Figure 7. 1 : Silhouette Scores	106
Figure 7. 2 : WCSS for Elbow Method.....	106
Figure 7. 3 : Confusion Matrix for MLP classifier	108
Figure 7. 4 : Convergence of MLP classifier.....	109
Figure 7. 5 : ROC Curve for MLP classifier.....	109
Figure 7. 6 : Precision-Recall Curve for MLP classifier	110
Figure 7. 7 : Precision, Recall and F1-score for MLP classifier.....	111
Figure 7. 8 : Questionnaire results – Pie chart 1	114
Figure 7. 9 : Questionnaire results – Pie chart 2	115
Figure 7. 10 : Questionnaire results – Pie chart 3	115
Figure 7. 11 : InERA effectiveness – Questionnaire results bar chart.....	115
Figure 7. 12 : Average time taken to reply comparison.....	116
Figure 7. 13 : Percentage of emails handled by InERA.....	117

LIST OF TABLES

Table 7. 1 : Comparison of average response time of InERA with GPT 3.5.....	113
---	-----

LIST OF ABBREVIATIONS

Abbreviation	Description
InERA	Intelligent Email Reply Assistant
MLP	Multi-Layer Perceptron
ML	Machine Learning
RNN	Recurrent Neural Network