

BRIDGING THE GAP

INTEGRATING LINEAR ALGEBRA IN THE
DEVELOPMENT AND UNDERSTANDING OF
LARGE LANGUAGE MODELS FOR SOFTWARE
ENGINEERING APPLICATIONS.

BY

ASGER POULSEN

202106630

BACHELOR'S THESIS

IN

COMPUTER ENGINEERING

SUPERVISOR: HUGO DANIEL MACEDO

Aarhus University, Department of Electrical and Computer Engineering

8 June 2024

Preface

This bachelor's thesis was written for the Department of Electrical and Computer Engineering, Aarhus University. It is part of the Computer Engineering study program, and was written in the spring of 2024.

All source files associated with this thesis are found at: <https://github.com/asgersong/BSc>

Asger Poulsen, February 13, 2024

Abstract

Acknowledgements

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 7 |
| 2 | Literature Review | 9 |
| 2.1 | Overview of Large Language Models | 9 |
| 2.2 | Linear Algebra in Machine Learning | 9 |
| 2.3 | Previous Studies on LLM Compression and Optimization | 9 |
| 3 | Theoretical Foundations | 11 |
| 3.1 | Linear Algebra in Deep Learning | 11 |
| 3.2 | Understanding LLMs | 11 |
| 4 | Methodology | 13 |
| 4.1 | Educational Synergy Development | 13 |
| 4.2 | Software Engineering Application | 13 |
| 4.3 | Evaluation Method | 13 |
| 4.3.1 | Compression of LLM (Flan-T5-Base) Using Low Rank De- composition | 13 |
| 4.3.2 | Metrics for Evaluation | 13 |
| 5 | Implementation | 15 |
| 5.1 | Curriculum Component Implementation | 15 |
| 5.2 | Chatbot Development | 15 |
| 5.2.1 | Tools and Libraries Used | 15 |
| 5.2.2 | Integration of LLM | 15 |
| 6 | Evaluation and Results | 17 |
| 6.1 | Curriculum Effectiveness | 17 |
| 6.2 | Chatbot Performance and Usefulness | 17 |
| 6.3 | Compression and Optimization of LLM | 17 |
| 6.3.1 | Methodology Applied | 17 |

| | | |
|----------|--|-----------|
| 6.3.2 | Results and Analysis | 17 |
| 7 | Discussion | 19 |
| 7.1 | Interpretation of Results | 19 |
| 7.2 | Theoretical and Practical Implications | 19 |
| 7.3 | Limitations and Challenges | 19 |
| 8 | Conclusion and Future Work | 21 |
| 8.1 | Summary of Key Findings | 21 |
| 8.2 | Contributions to the Field | 21 |
| 8.3 | Recommendations for Future Research | 21 |

Chapter 1

Introduction

Chapter 2

Literature Review

2.1 Overview of Large Language Models

2.2 Linear Algebra in Machine Learning

2.3 Previous Studies on LLM Compression and Optimization

Chapter 3

Theoretical Foundations

3.1 Linear Algebra in Deep Learning

3.2 Understanding LLMs

Chapter 4

Methodology

4.1 Educational Synergy Development

4.2 Software Engineering Application

4.3 Evaluation Method

4.3.1 Compression of LLM (Flan-T5-Base) Using Low Rank Decomposition

4.3.2 Metrics for Evaluation

Chapter 5

Implementation

5.1 Curriculum Component Implementation

5.2 Chatbot Development

5.2.1 Tools and Libraries Used

5.2.2 Integration of LLM

Chapter 6

Evaluation and Results

6.1 Curriculum Effectiveness

6.2 Chatbot Performance and Usefulness

6.3 Compression and Optimization of LLM

6.3.1 Methodology Applied

6.3.2 Results and Analysis

Chapter 7

Discussion

7.1 Interpretation of Results

7.2 Theoretical and Practical Implications

7.3 Limitations and Challenges

Chapter 8

Conclusion and Future Work

8.1 Summary of Key Findings

8.2 Contributions to the Field

8.3 Recommendations for Future Research