AARHUS
UNIVERSITY
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Rev. 12/12/2023

Specifications of Bachelor of Science project for students at Aarhus University, Department of Electrical and Computer Engineering

Date	12/12/2023		
Project title	**Bridging the Gap:** Integrating Linear Algebra in the Development and Understanding of Large Language Models for Software Engineering Applications.		
The project applies to students with a specialisation in (mark with X)			
Electrical Engineering		Computer Engineering	x
ECE supervisor (co-supervisor) info including name, email, mobile phone. *Note - main supervisor must be ECE VIP, Assistant Prof or more senior. PhD students, Postdocs and other experts (e.g. from industry, or from a different AU department) can co-supervise.*	Hugo Daniel Macedo (ECE supervisor) hdm@ece.au.dk		
Special demands to: - equipment - place - confidentiality			

Project description

- **Background/Motivations**
  Large Language Models (LLMs) like GPT-4 have revolutionized natural language processing with their advanced text generation and problem-solving skills. These models rely heavily on linear algebra to navigate the complex vector spaces they work in.
  Despite the growing use of AI in software engineering, there's still a notable disconnect between the core mathematical principles and their practical use in software development. Linear algebra, while fundamental to LLMs, is often not directly linked to its real-world applications for software engineers. This thesis aims to bridge this gap by delving into the synergy between LLMs, mathematics, and software engineering, with a focus on practical linear algebra.

- **Objectives**

  - **Theoretical Exploration:** Delve into the mathematics behind LLMs, focusing on the linear algebraic operations, such as matrix multiplication. eigenvalues, and eigenvectors, that underpin neural network architectures, weight optimization, and data representation.
  - **Educational Synergy:** Craft a curricular component that correlates linear algebra principles with the practical aspects of LLM structures and algorithms, thus bridging the gap between abstract mathematical concepts and their practical applications in AI.
  - **Software Engineering Application:** Implement a chatbot.
  - **Evaluation:** Assess the effectiveness of compression of LLM (Flan-T5-Base) with low rank decomposition of attention weight matrices.

- **Learning outcomes**

  - **Deep Understanding of Linear Algebra in LLMs:** Develop a thorough understanding of the specific linear algebra operations employed in LLMs.
  - **Practical Application of Theoretical Concepts:** Acquire hands-on experience in working with LLM's and applying them to real-world language processing tasks.
  - **Replication and Extension:** Build on existing study, specifically the compression of LLMs using low rank decomposition of attention weight matrices.

Activity plan

Week:	Task
5	Introduction and background research
6	Introduction and background research
7	Theoretical Exploration
8	Theoretical Exploration
9	Replicate Study
10	Replicate Study
11	Replicate Study
12	SWENG App Development
13	Easter Break
14	SWENG App Development
15	SWENG App Development
16	SWENG App Development
17	Analysis and evaluation
18	Finalizing research components
19	Buffer Week
20	Report writing and revision
21	Report writing and revision
22	Report writing and revision
23	Final review and submission

Supervision Plan

Hugo Daniel Macedo will be the supervisor. The student and supervisor plan to meet weekly in person.