

DIPARTIMENTO DI INGEGNERIA E ARCHITETTURA

Corso di Laurea Magistrale Ingegneria Informatica

Enhancing Fault Isolation in Hardware Systems Using Large Language Models (LLM)

Miglioramento dell'isolamento dei guasti nei sistemi hardware utilizzando modelli linguistici di grandi dimensioni (LLM)

Relatore:

Prof. Gianfranco Lombardo

Tesi di Laurea di:

Matteo Gianvenuti Matricola: 364979

Table of Contents

[Introduction 2](#_Toc199942774)

[1. State Of The Art 3](#_Toc199942775)

[1.1 Large Language Models (LLMs) 3](#_Toc199942776)

[1.2 Self-supervised learning (SSL) 3](#_Toc199942777)

[1.3 Prompt Engineering 3](#_Toc199942778)

[1.3.1 In-Context Learning (ICL) 3](#_Toc199942779)

[1.3.2 Few shot 3](#_Toc199942780)

[1.4 Generative pre-trained transformer (GPT) 4](#_Toc199942781)

[1.5 Llama 4](#_Toc199942782)

[1.6 Diagnostics 4](#_Toc199942783)

[Bibliography 4](#_Toc199942784)

Introduzione

- Ambito dell’attività di tesi

- Obiettivi dell’attività di tesi

Stato dell’arte

- Documentare tutto ciò che esisteva prima della tesi in termini di tecnologie, metodologie e modelli

2/3 capitoli di metodologia in cui spiega cosa ha effettivamente svolto, progettato e realizzato (architettura software, addestramento di modelli, dati utilizzati e loro composizione etc..)

- 1 capitolo su i risultati

- 1 breve capitolo di conclusioni in cui ricapitola il tutto traendo le conclusioni e proponendo degli sviluppi futuri.

***Enhancing Fault Isolation in Hardware Systems Using Large Language Models (LLM): A Real-Time Diagnostic Framework***

*This thesis explores the application of Large Language Models (LLM) to improve fault isolation in hardware systems. Thanks to their ability to analyse large amounts of unstructured data, LLMs can identify patterns and anomalies in system logs and telemetry data collected from hardware devices. The objective is to develop a framework that uses LLMs to diagnose faults in real-time, reducing downtime and improving the overall reliability of the system. Case studies and experimental results will be presented to demonstrate the effectiveness of the proposed approach.*

# Introduction

In this thesis work we want to explore the use of Large Language Models (LLMs) in hardware diagnostics. LLMs have shown good memorization and "understanding" capabilities of natural language, even in reasoning tasks. Especially in the case in which the models know the environment in which they operate.

In our case, we initially want to make the LLMs understand how a hardware architecture is structured so that the model is able to solve the diagnostics at its best. Subsequently, we want to exploit the capabilities of the LLMs to identify any problems in the hardware by analysing system logs, telemetry data and hardware functionality test results. To carry out these activities it was necessary to create a specific dataset and a simulation environment. We also implemented the Open WebUI graphical interface.

So, we have two objectives, the first is to verify that the LLMs are able to understand the hardware architecture. While the second is to verify that the LLMs are able to understand and solve hardware problems. This work was based on In-Context Learning (ICL) and Prompt Engineering. In particular, we compared the performance of different models according to various metrics.

# 1. State Of The Art

## 1.1 Large Language Models (LLMs)

LLMs are machine learning models with many parameters, designed for natural language processing tasks, especially text generation. They are trained with self-supervised learning on a vast amount of text.

LLMs can be fine-tuned for specific tasks or guided by prompt engineering, as done in this work. These models acquire predictive power regarding syntax, semantics, and ontologies inherent in human language corpora, but they also inherit inaccuracies and biases present in the data they are trained in.

The largest and most capable LLMs are generative pre-trained transformers (GPTs).

## 1.2 Self-supervised learning (SSL)

## 1.3 Prompt Engineering

Prompt engineering is the process of structuring or crafting an instruction in order to produce the best possible output from a generative artificial intelligence (AI) model.

In case of text-to-text language models a *prompt* is natural language text describing the task that an AI should perform. It can be a query, a command, or a longer statement including context, instructions, and conversation history. Prompt engineering may involve phrasing a query, specifying a style, choice of words and grammar, providing relevant context, or describing a character for the AI to mimic.

There are many different types of prompt engineering:

* Chain-of-Thought (CoT): It is a reasoning technique that goes through several intermediate steps before giving a definitive answer to a question.
* In-Context Learning (ICL): it is the ability of models to learn information temporarily. The information of interest is specified in the prompt, typically in the initial part. This information is then used by the model in the session to answer questions by generating text, this allows the model to be trained without necessarily having to perform fine-tuning.

The weights of the network are not updated, so the information is temporary. Starting a new instance the model does not retain this knowledge.

* Self-consistency decoding:
* Tree-of-thought:
* Prompting to estimate model sensitivity?:
* Automatic prompt generation?:

In this work we focused on ICL and CoT, building a few-shot technique.

### 1.3.1 In-Context Learning (ICL)

### 1.3.2 Few shot

## 1.4 Generative pre-trained transformer (GPT)

## 1.5 Llama

## 1.6 Diagnostics

# Bibliography

[1] LLMs. (Online) <https://en.wikipedia.org/wiki/Large_language_model>.

[https://web.stanford.edu/~jurafsky/slp3/](https://eur01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fweb.stanford.edu%2F~jurafsky%2Fslp3%2F&data=05%7C02%7Cmatteo.gianvenuti%40studenti.unipr.it%7C3eb6fbcc055f4e8a0aff08dd8bd29bee%7Cbb064bc5b7a841ecbabed7beb3faeb1c%7C0%7C0%7C638820458860462928%7CUnknown%7CTWFpbGZsb3d8eyJFbXB0eU1hcGkiOnRydWUsIlYiOiIwLjAuMDAwMCIsIlAiOiJXaW4zMiIsIkFOIjoiTWFpbCIsIldUIjoyfQ%3D%3D%7C0%7C%7C%7C&sdata=R76CP%2Fvpd1I40a8MReCrxBrrjUE6ColECYnknFGpswY%3D&reserved=0) book various arguments (it should cover almost all)

[2]