

# **[Thesis Title]**

**Rui Pedro Teles Ribeiro**

**A dissertation submitted in partial fulfillment of  
the requirements for the degree of Master of Science,  
Specialisation Area of Software Engineering**

**Advisor: Prof<sup>ª</sup>. Dra. Piedade Carvalho  
Supervisor: José Soares**

Porto, March 23, 2025



# Statement of Integrity

I hereby declare having conducted this academic work with integrity.

I have not plagiarised or applied any form of undue use of information or falsification of results along the process leading to its elaboration.

Therefore the work presented in this document is original and authored by me, having not previously been used for any other end.

I further declare that I have fully acknowledged the Code of Ethical Conduct of P.PORTO.

ISEP, Porto, March 23, 2025



# Dedictory

The dedicatory is optional. Below is an example of a humorous dedication.

"To my wife Marganit and my children Ella Rose and Daniel Adam without whom this book would have been completed two years earlier." in "An Introduction To Algebraic Topology" by Joseph J. Rotman.



# Abstract

This document explains the main formatting rules to apply to a TMDEI Master Dissertation work for the MSc in Computer Engineering of the Computer Engineering Department (DEI) of the School of Engineering (ISEP) of the Polytechnic of Porto (IPP).

The rules here presented are a set of recommended good practices for formatting the dissertation work. Please note that this document does not have definite hard rules, and the discussion of these and other aspects of the development of the work should be discussed with the respective supervisor(s).

This document is based on a previous document prepared by Dr. Fátima Rodrigues (DEI/ISEP).

The abstract should usually not exceed 200 words, or one page. When the work is written in Portuguese, it should have an abstract in English.

Please define up to 6 keywords that better describe your work, in the *THESIS INFORMATION* block of the `main.tex` file.

**Keywords:** Keyword1, ..., Keyword6





# Resumo

Trabalhos escritos em língua Inglesa devem incluir um resumo alargado com cerca de 1000 palavras, ou duas páginas.

Se o trabalho estiver escrito em Português, este resumo deveria ser em língua Inglesa, com cerca de 200 palavras, ou uma página.

Para alterar a língua basta ir às configurações do documento no ficheiro `main.tex` e alterar para a língua desejada ('english' ou 'portuguese')<sup>1</sup>. Isto fará com que os cabeçalhos incluídos no template sejam traduzidos para a respetiva língua.

---

<sup>1</sup>Alterar a língua requer apagar alguns ficheiros temporários; O target **clean** do **Makefile** incluído pode ser utilizado para este propósito.



# Acknowledgement

The optional Acknowledgment goes here... Below is an example of a humorous acknowledgment.

"I'd also like to thank the Van Allen belts for protecting us from the harmful solar wind, and the earth for being just the right distance from the sun for being conducive to life, and for the ability for water atoms to clump so efficiently, for pretty much the same reason. Finally, I'd like to thank every single one of my forebears for surviving long enough in this hostile world to procreate. Without any one of you, this book would not have been possible." in "The Woman Who Died a Lot" by Jasper Fforde.



# Contents

<b>List of Figures</b>	<b>xv</b>
<b>List of Tables</b>	<b>xvii</b>
<b>List of Algorithms</b>	<b>xix</b>
<b>List of Source Code</b>	<b>xxi</b>
<b>List of Symbols</b>	<b>xxiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Context . . . . .	1
1.2 Problem . . . . .	1
1.3 Objectives . . . . .	1
1.4 Methodology . . . . .	1
1.5 Work Plan . . . . .	1
1.6 Document Structure . . . . .	1
<b>2 Key Concepts</b>	<b>3</b>
2.1 Artificial Intelligence . . . . .	3
2.2 Machine Learning & Deep Learning . . . . .	3
2.3 Large Language Models . . . . .	3
2.4 Retrieval-Augmented Generation . . . . .	4
2.5 Bases de Dados Vetoriais . . . . .	4
2.6 Aplicação ao Suporte Técnico . . . . .	4
<b>3 State of the Art</b>	<b>5</b>
3.1 Methodology . . . . .	5
3.1.1 Research Questions . . . . .	5
3.1.2 Research Scope . . . . .	5
3.1.3 Eligibility Criteria . . . . .	5
3.1.4 Selection Process . . . . .	5
3.1.5 Data Collection . . . . .	5
3.2 Results and Analysis . . . . .	5
3.3 Related Work . . . . .	5
<b>Appendix A Appendix Title Here</b>	<b>7</b>



# List of Figures





# List of Tables



# List of Algorithms



## List of Source Code



# List of Symbols

$a$	distance	m
$P$	power	W ( $\text{Js}^{-1}$ )
$\omega$	angular frequency	rad





# **Chapter 1**

## **Introduction**

**1.1 Context**

**1.2 Problem**

**1.3 Objectives**

**1.4 Methodology**

**1.5 Work Plan**

**1.6 Document Structure**



## Chapter 2

# Key Concepts

### 2.1 Artificial Intelligence

A inteligência artificial (IA) é um ramo científico da computação que se dedica ao desenvolvimento de sistemas capazes de executar tarefas que normalmente exigiriam inteligência humana. Estes sistemas têm a capacidade de executar funções avançadas e analisar dados de grande escala a fim de gerar respostas precisas. Baseado num conceito do filósofo do grego Aristoteles, a IA surgiu na década de 1950 por Allan Turing, onde o mesmo escreveu sobre a possibilidade de uma máquina pensar e imitar o comportamento humano inteligente. Atualmente a IA é aplicada em diversos setores, como na saúde através do diagnóstico automatizado de doenças, no setor financeiro para análises de mercado e deteção de fraudes, entre outros. Recentemente a IA sofreu um "boom" tecnológico, com a corrida da IA generativa, sendo o seu componente-chave a fundação da OpenAI em 2015 e surgimento do ChatGPT em 2022, sistema este capaz de processar linguagem natural (NLP) e gerar respostas precisas e corretas sobre variados assuntos (<https://hai.stanford.edu/news/ai-spring-four-takeaways-major-releases-foundation-models>).

Dentro da IA existem diferentes sub-ramos científicos, como:

- Machine Learning (ML): Ensina computadores a aprender padrões a partir de dados através de redes neuronais ou árvores de decisão;
- Deep Learning (DL): Sub-ramo do ML que faz uso de redes neuronais para modelar e interpretar padrões complexos;
- Processamento de linguagem natural (NLP): Interpretação de linguagem natural humana.
- Visão computacional: Interpretação de imagens e vídeos

### 2.2 Machine Learning & Deep Learning

Diferença entre aprendizado de máquina e aprendizado profundo. Como esses conceitos se relacionam com modelos de IA modernos.

### 2.3 Large Language Models

Os Large Language Models (LLMs) representam um avanço significativo na IA. Proposta pela Google em 2017, atualmente, Transformer é a arquitetura de DL mais explorada para

esta componente. Os Transformers foram inicialmente desenvolvidos como melhoria das arquiteturas anteriores para a tradução automática, mas desde então têm encontrado muitas aplicações, como na visão computacional e NLP. Conduziram ao desenvolvimento de sistemas pré-treinados, tais como Generative Pre-trained Transformers (GPTs) and Bidirectional Encoder Representations from Transformers (BERT). Estes modelos são treinados através do paradigma Self-supervised learning (SSL), no qual aprendem representações úteis dos dados sem a necessidade de rótulos manuais. No SSL, o próprio modelo gera os seus rótulos a partir dos dados brutos, criando tarefas preditivas auxiliares chamadas pretext tasks. Masked Language Modeling é um exemplo de tarefa preditiva, utilizada pelo BERT, onde palavras aleatórias são ocultadas em uma frase e o modelo aprende a prever as palavras corretas, isto no contexto de NLP. Em contraste o GPT faz uso do Casual Language Modeling onde o modelo prevê a próxima palavra numa sequência de texto, dado o contexto anterior.

## 2.4 Retrieval-Augmented Generation

Definição e funcionamento do RAG. Diferença entre RAG e abordagens puramente gerativas. Benefícios do RAG para suporte técnico.

## 2.5 Bases de Dados Vetoriais

Conceito de embeddings e busca vetorial. Exemplos de ferramentas (FAISS, Weaviate, Pinecone).< Importância da base vetorial no contexto do RAG.

## 2.6 Aplicação ao Suporte Técnico

Como esses conceitos são aplicáveis ao problema da dissertação. Benefícios esperados da implementação.

## **Chapter 3**

# **State of the Art**

### **3.1 Methodology**

#### **3.1.1 Research Questions**

RQ1 - Quais as melhores práticas para alimentar base de dados vetoriais

#### **3.1.2 Research Scope**

#### **3.1.3 Eligibility Criteria**

#### **3.1.4 Selection Process**

#### **3.1.5 Data Collection**

### **3.2 Results and Analysis**

### **3.3 Related Work**



## **Appendix A**

### **Appendix Title Here**

Write your Appendix content here.