Answering consumer health questions with non-expert language

FINAL PRESENTATION
PREPARAÇÃO PARA DISSERTAÇÃO/ESTÁGIO
2024/2025

Daniel Ferreira, 102885

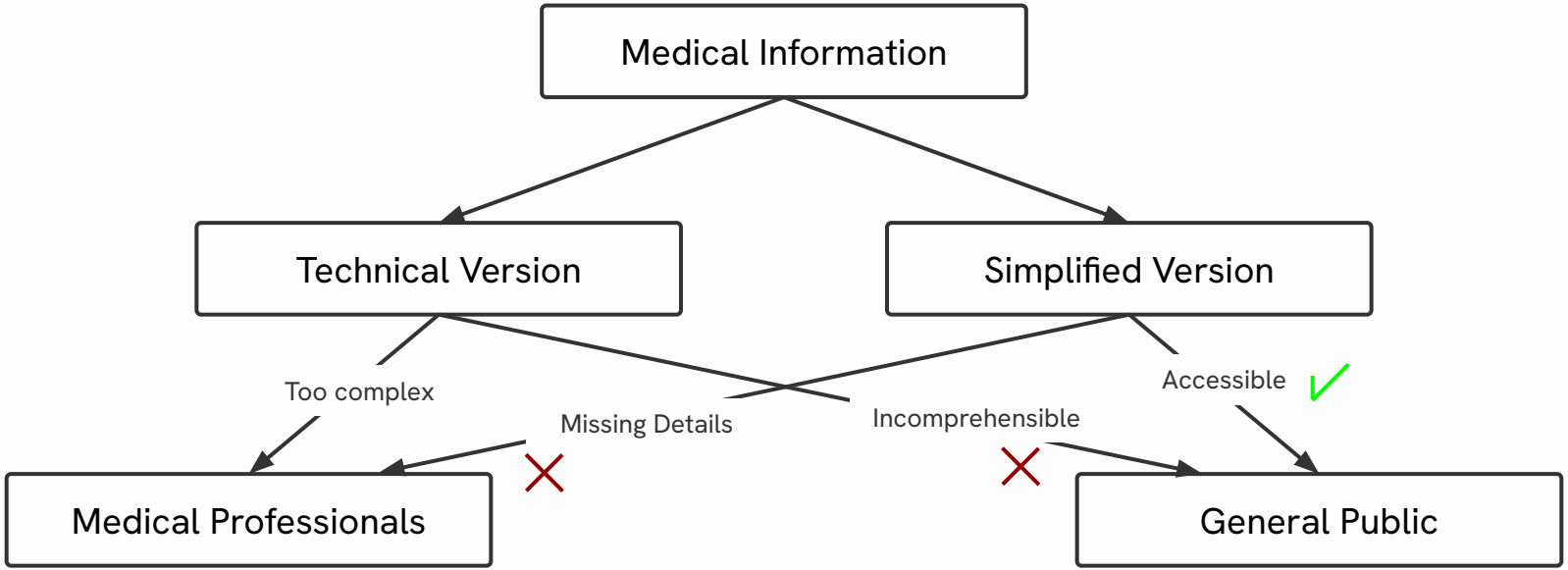
Overview

The following outlines the key participants involved in this project.

Student	Daniel Jorge Bernardo Ferreira
Supervisor	Sérgio Matos
Co-supervisor	Tiago Almeida
Summary	With the appearance of chatbot applications, searches and conversations regarding health issues can be expected to increase, which raises the challenges on how to reply with verified and reliable information. The objective of this work is to develop and evaluate a method for retrieval and summarization of scientific articles that contain answers to consumer health questions.

For more info:
Portfolio

Motivation


Medical Information
Technical Version
Simplified Version
Too complex
Missing Details
Incomprehensible
Accessible
Medical Professionals
General Public

Objectives

Quantify Medical Text Complexity

Develop a set metrics for measuring medical text complexity

Create Complexity-Annotated Dataset

Build a high-quality dataset of health QA pairs with complexity levels

Develop Controllable Language Model

Create a model that generates answers with adjustable complexity

Controlled text generation

Control methods

Training phase

Retraining

Fine-tuning

Reinforcement learning

Inference phase

Prompt engineering

Latent space manipulation

Guided Decoding

Control types

Content control

Structure

Vocabulary

Length

Format

Attribute control

Sentiment

Style

Topic

Safety

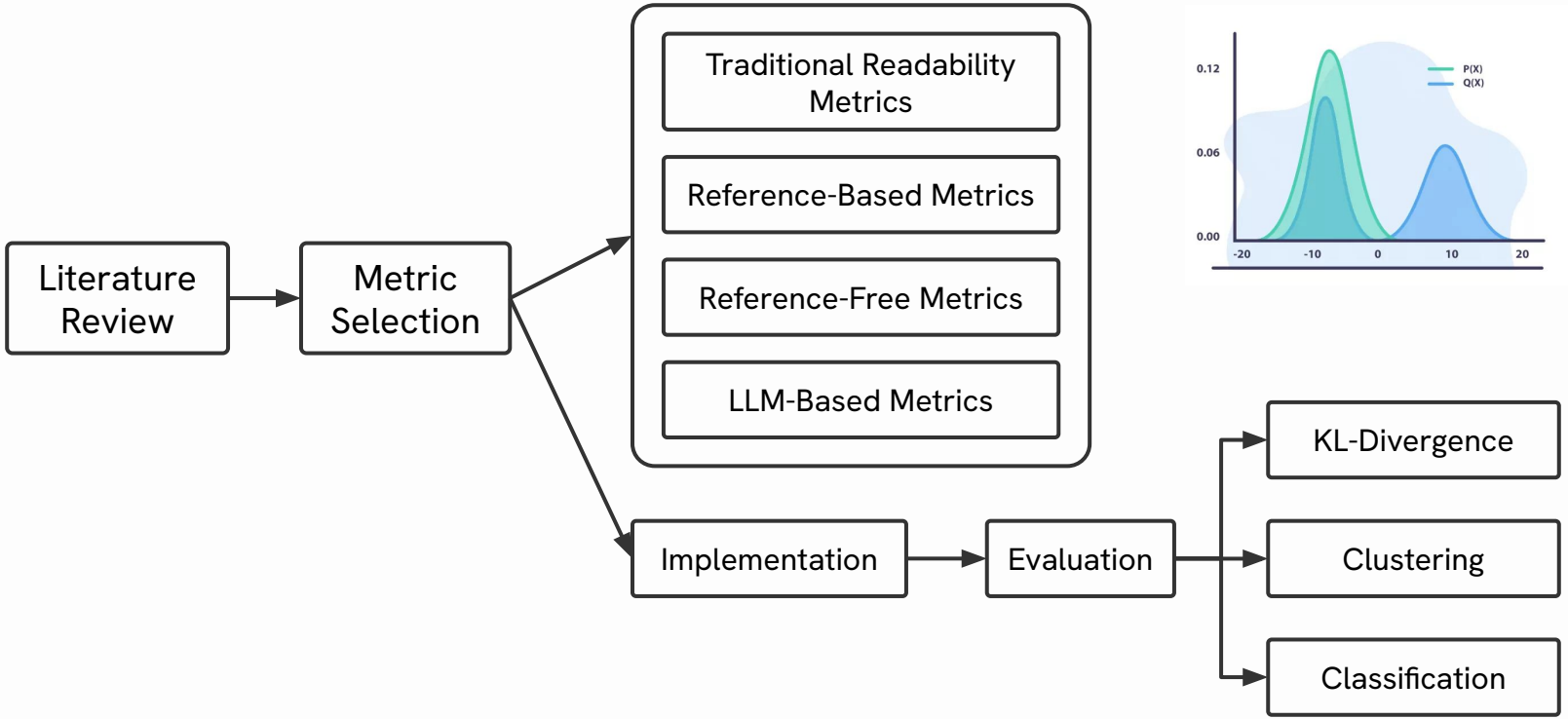
Quality requirements

Fluency

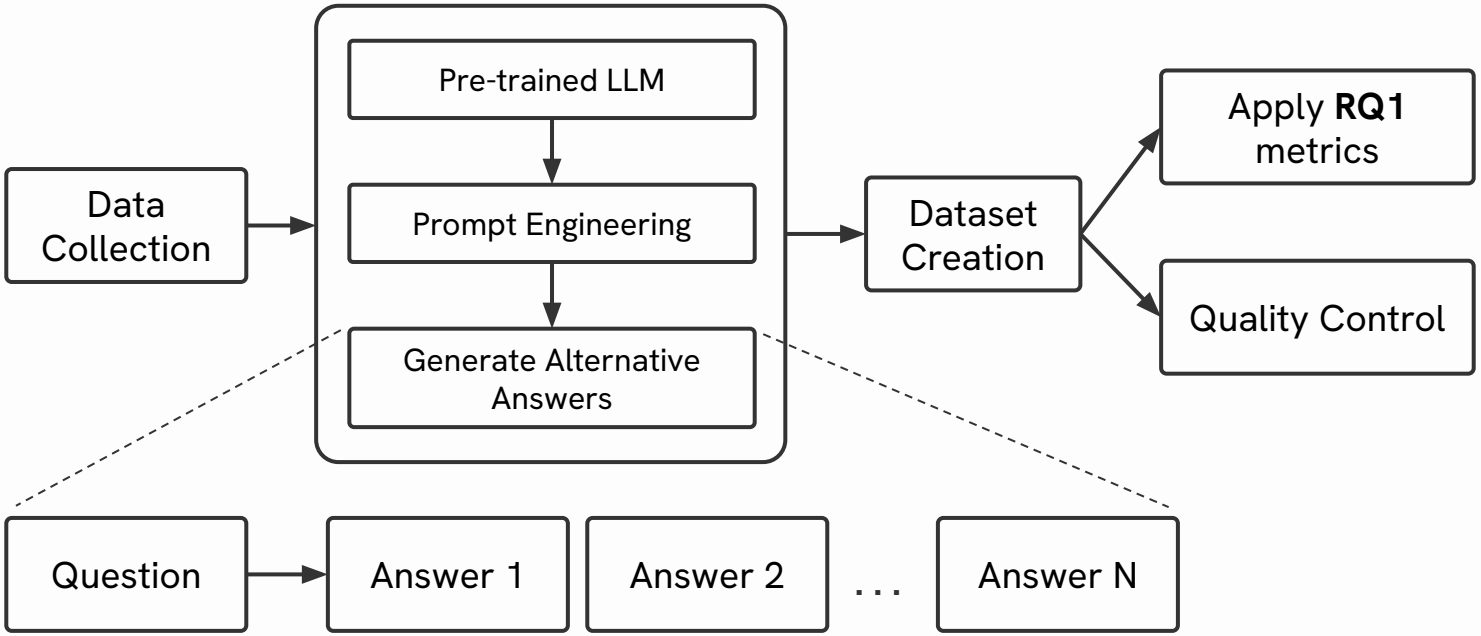
Helpfulness

Diversity

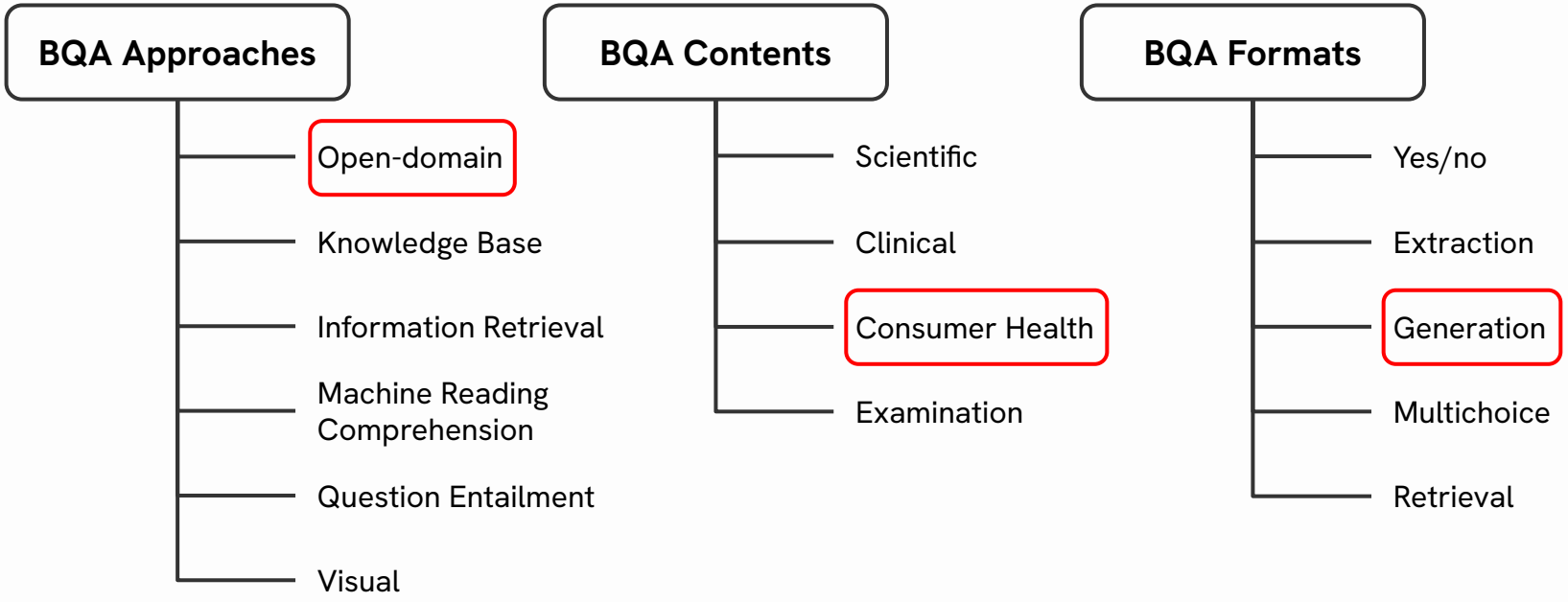
Methodology (RQ1)



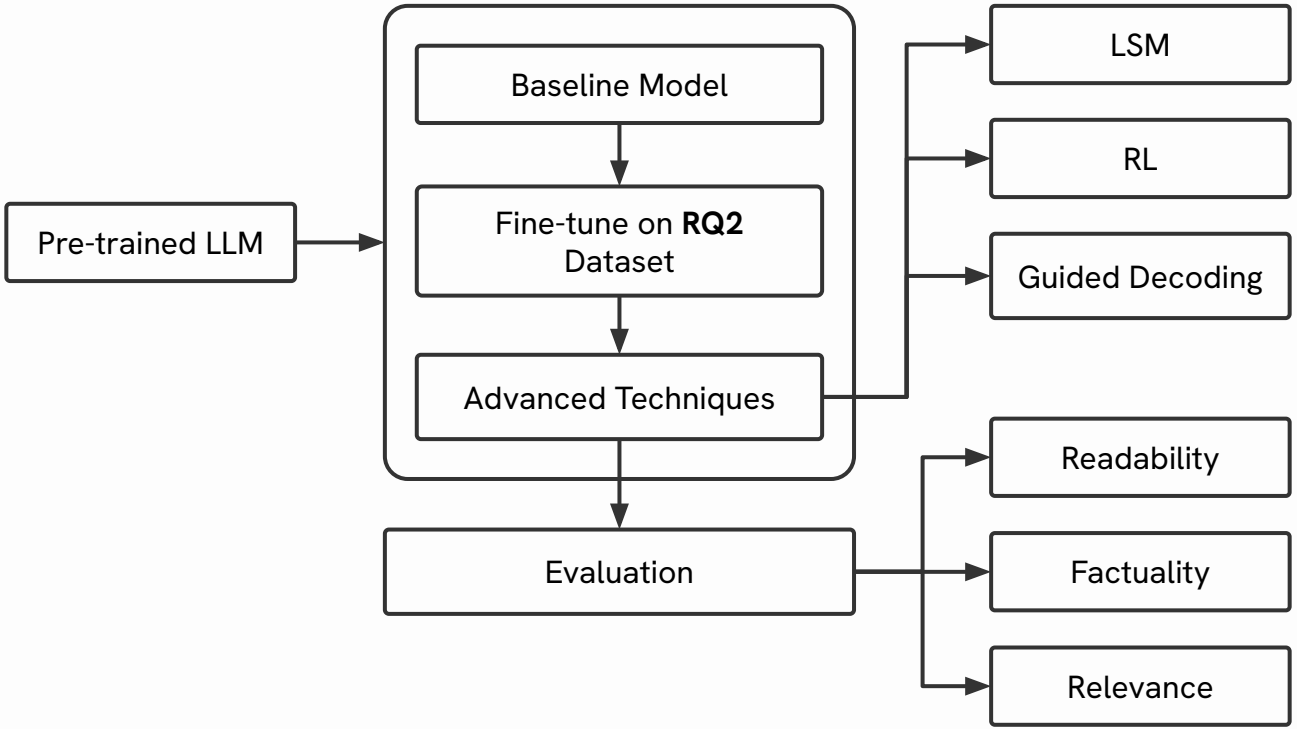
Methodology (RQ2)

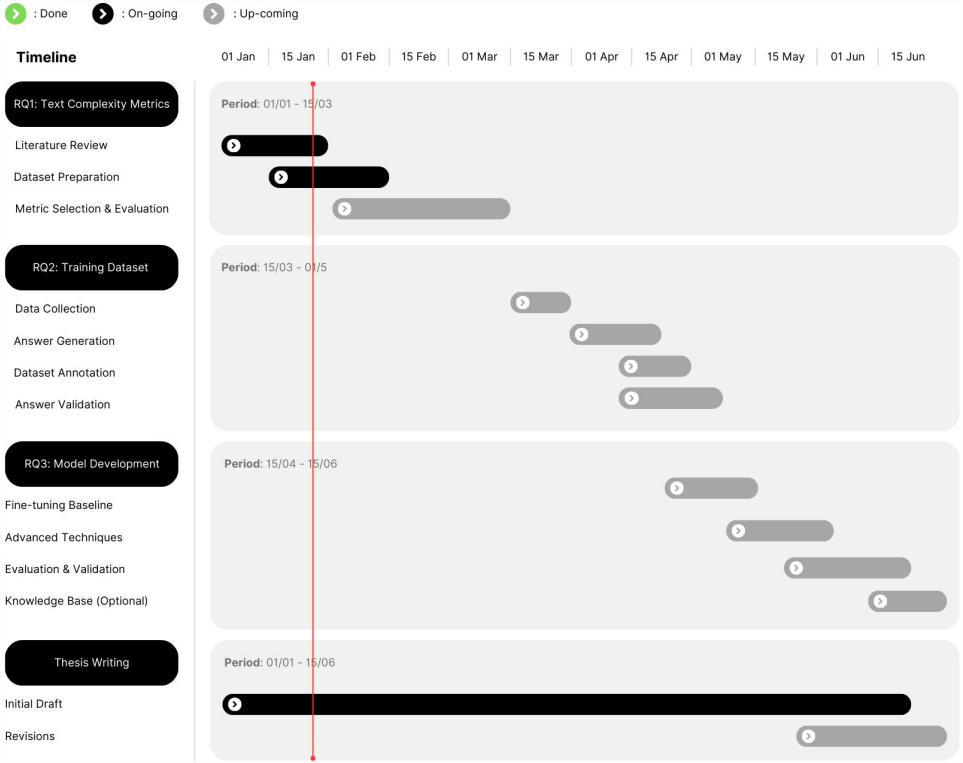


Datasets


BQA Approaches
Open-domain
Knowledge Base
Information Retrieval
Machine Reading Comprehension
Question Entailment
Visual
BQA Contents
Scientific
Clinical
Consumer Health
Examination
BQA Formats
Yes/no
Extraction
Generation
Multichoice
Retrieval

Methodology (RQ3)



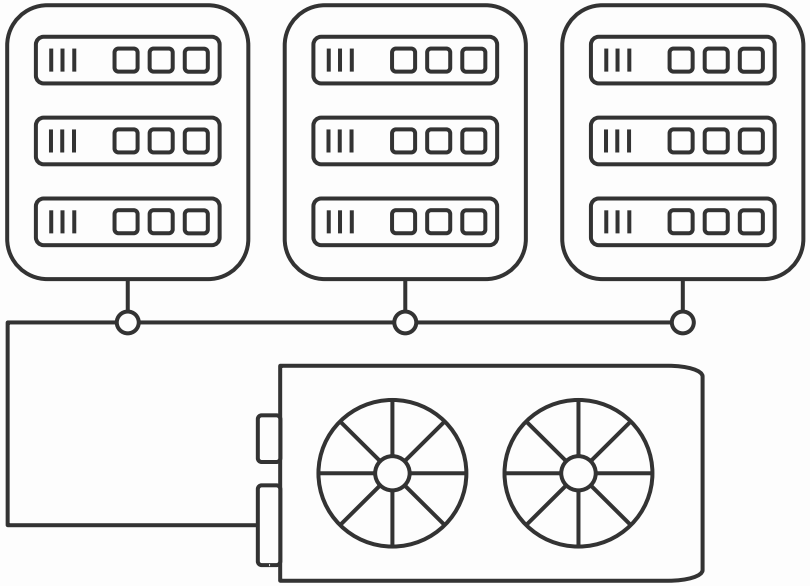


Timeline

The Gantt chart provides a visual representation of the proposed work plan, outlining the tasks and milestones described before.

Pleiades

High Performance Computing (HPC) cluster operated by **IEETA**. The system features both traditional CPU nodes and GPU-accelerated computing nodes, making it versatile for different types of computational workloads.



Thanks!

Do you have any questions?

djbf@ua.pt
danielferreira011102.pages.dev