

Configuration Manual

MSc Research Project MSc in Artificial Intelligence

Madni Ali Hussain Student ID: 23158859

School of Computing National College of Ireland

Supervisor:

Paul Stynes

National College of Ireland



MSc Project Submission Sheet

School of Computing

Student Name:	Madni Ali Hussain		
Student ID:	23158859		
Programme:	MSC in Artificial Intelligence	Year:	2024
Module:	MSc Practicum/Internship part 2		
Lecturer: Submission Due Date:	Paul Stynes		
	12 December 2024		
Project Title:	A Retrieval-Augmented Generation Framework for Medical Question Answer		

Word Count: 812 Page Count: 2

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

<u>ALL</u> internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Madni Ali Hussain

Date: 12 December 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	\checkmark
Attach a Moodle submission receipt of the online project	\checkmark
submission, to each project (including multiple copies).	•
You must ensure that you retain a HARD COPY of the project,	\checkmark
both for your own reference and in case a project is lost or mislaid. It is	
not sufficient to keep a copy on computer.	

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Configuration Manual

Madni Ali Hussain Student ID: 23158859

1 Installation

- 1. Install Conda
 - a. If not installed download from Anaconda
 - b. Verify installation with conda -version
- 2. Setup a Conda Environment
 - a. conda create --name <env_name> python=<version>
 - b. activate environment conda activate <env_name>
- 3. Install dependencies
 - a. conda install -- file requirements.txt
 - **b.** verify installation **conda install --file requirements.txt**
- 4. Install and Setup Ollama
 - a. Download Ollama from Ollama
 - **b.** Download embeddings models by running
 - i. ollama pull mxbai-embed-large
 - ii. ollama pull nomic-embed-text
 - iii. ollama pull all-minilm
- 5. Install Node
 - a. Install Node by downloading Node

2 Execution and Configuration

- 1. Database Setup
- **Database Choice**: We recommend using a PostgreSQL database for ease of use, specifically Supabase for this configuration.
- **Extension Installation**: Install the pgvector extension by executing the following SQL command in your database query editor: CREATE EXTENSION vector;
- Table Setup: Initialize the database structure by running the SQL scripts provided:

```
create table
public.answer_sections (
    id bigint generated always as identity not null,
    answer_id bigint not null,
    content text not null,
    embedding public.vector null,
    section_order integer null,
    constraint answer_sections_pkey primary key (id),
    constraint answer_sections_answer_id_fkey foreign key (answer_id) references
questions (id)
    ) tablespace pg_default;
```

```
create table
  public.questions (
    id serial not null,
    content text not null,
    embedding public.vector null,
    created_at timestamp with time zone null default now(),
    type text null,
    question text null,
    answer text null,
    constraint questions_pkey primary key (id)
    ) tablespace pg_default;
```

There are two folders in the code artifact

- 2. Backend
 - a. For backend we have already installed all the dependencies now we need to run **conda activate <env_name>**
 - b. Then run py medical-rag-application.py, this will run our backend
- 3. Frontend
 - a. Go into the frontend folder and npm install
 - b. And run npm run dev for running the frontend
 - c. Go to http://localhost:3000
 - d. Upload Data: Navigate to the "Upload Data" tab to upload the Med Quad data CSV file. This step is necessary only if you are not using the pre-configured Supabase instance.
 - e. Go to the "Ask Question" tab to submit questions and receive responses from the backend.

3 Evaluation

• RAG Technique Evaluation: Each RAG technique has a separate folder containing the necessary scripts for execution and evaluation. For example, to evaluate the Multi-Query RAG technique: For Example, for Multi Query RAG

- Navigate to the Multi-Query-RAG folder.
- Execute the script with python multi-query-rag.py
- For performance evaluation, run python app.py