Final Report: The Science of Weight Loss PLP

1. Introduction

The project goal was to design and build an interactive Personal Learning Portal (PLP) enabled by an LLM with a Retrieval-Augmented Generation (RAG) pipeline. This will help users understand the actual science behind losing weight and maintaining lost weight. Today, social media is rife with misinformation and conflicting advice from influencers. Scientific information is either too complex for the average person to understand or not interesting enough for the public to decode. The public also faces a flood of conflicting, often inaccurate, nutritional information, making it difficult to find reliable guidance. This project aims to solve this by providing evidence-based answers from a curated set of scientific sources. The PLP is a RAG system powered by the Gemini API. It uses a hand-selected corpus of 12 high-quality scientific documents (with a multitude of topics) to generate trustworthy, educational responses within an interactive Streamlit application. The project was guided by seven key questions covering the science, diet, exercise, psychology, and long-term maintenance of weight loss.

2. System Architecture and Design

Overall Architecture:

The user interacts with a Streamlit front-end. The request is handled by a modular, class-based Python backend. The main file is a RAG engine which contains the RAG class. The RAG engine retrieves relevant text chunks from a ChromaDB vector store. The chunks and the user's question are sent to the Gemini API (with a gemini-2.5-flash model) for synthesis. The final, teacher-like response is displayed to the user.

Technology Stack:

- **Language/Frameworks:** Python, Streamlit, LangChain
- **LLM & Embeddings:** Google Gemini API (gemini-2.5-flash, embedding-001)
- **Vector Database:** ChromaDB (local persistence)
- **Data Processing:** PyMuPDF(fitz) for document ingestion

PLP Interface Enhancements: The following two LMS inspired features were chosen for the PLP -

1. **Guided Learning Path:** An interactive sidebar on the side helps the user click and explore different topics. It acts like a curriculum
2. **Interactive Topic Completion:** .The topics on the right can be clicked and accessed. Once a topic is clicked, the symbol changes to "✅" allow the user to track their progress.

3. Corpus Curation Strategy

- **Selection Criteria:** The knowledge base was built by prioritizing high-quality, freely available, text-based sources.
    - Emphasis was placed on systematic reviews, meta-analyses, and official clinical guidelines to ensure scientific consensus.
    - Inclusion of authoritative articles from quality institutions like Harvard and the Mayo Clinic to provide practical, accessible explanations.
    - Other sources like podcasts, youtube videos, etc were omitted to keep the vectorstore relevant and efficient.
- **Content Coverage:** The final corpus of 12 documents (sources) was chosen to comprehensively address every learning question, from the core science of energy balance to the psychological aspects of habit formation.
- **Inclusivity:** A source was especially included to address the effectiveness of weight management interventions in underrepresented ethnic minority groups to combat bias in the model knowing that a one-size-fits-all approach is insufficient.

4. Evaluation Methodology

- **Quantitative Evaluation (RAGAs):**
    - The RAGAs framework was used to generate objective scores for the system's performance.
    - **Faithfulness:** This metric measures how factually consistent the generated answer is with the retrieved source documents.
    - **Answer Relevancy:** This metric measures how on-topic and relevant the answer is to the user's original question.
    - The full suite of RAGAs metrics were not used because key retrieval-focused metrics, like **context_recall** and **context_precision**, require a manually-created "ground truth" answer for each question, which was beyond the scope.
- **Qualitative Evaluation (User Reflection):**
    - A set of 7 sample queries (the core learning questions) was run through the system.
    - Each generated answer was manually reviewed and analyzed for its clarity, tone, accuracy, and overall effectiveness as a learning tool.

5. Results and Analysis

The effectiveness of the Personal Learning Portal was evaluated using a two-faced approach: a quantitative assessment with the RAGAs framework and a qualitative human-in-the-loop reflection. This provided a more nuanced and comprehensive view of the system's performance. This shows that there are both strengths and also areas for improvement.

- **Quantitative Findings:**

#	Question	faithfulness	answer_relevancy
1	How does weight loss fundamentally work in the body?	0.87	0.86
2	What are the components of a diet that supports sustainable weight loss?	0.77	0.90
3	What are the recommended types and amounts of exercise for losing weight?	0.97	0.87
4	What is a safe and sustainable rate of weight loss?	0.90	0.92
5	What other lifestyle habits like sleep and stress are important?	0.71	0.84
6	What are effective, science-backed strategies for managing hunger and cravings?	0.46	0.91
7	What distinguishes successful long-term weight maintainers from those who regain weight?	0.74	0.90

- **Qualitative Findings:**

Answer Relevancy: The system showed excellent performance in understanding and addressing the user's query (probably because of the use of a better Gemini model). The answer_relevancy scores are consistently high, with five of the seven questions scoring above 0.87. This shows that the RAG pipeline is very effective at retrieving the correct documents and generating an answer that is on-topic and directly relevant to the user's learning goal.

Faithfulness: The faithfulness is a bit more nuanced - the scores here show an interesting finding. Many answers were highly faithful to the source material (0.97), but some questions (like question 6) were significantly lower (0.46). This is probably due to the "Expert Educator" persona type prompting. The prompt was successful in creating a teacher-like and comprehensive tone (which is really important in this context) but sometimes it itself led the model to add explanatory details or synthesize concepts on its own in a way that was not strictly present in the retrieved source text which lowered the faithfulness score.

6. Conclusion and Future Work

In conclusion, a functional, interactive, and most importantly - an evidence-based PLP based on a RAG pipeline for the science of weight loss was created which can now answer questions using high quality sources. The system was then evaluated using RAGAs for faithfulness and relevance which led to some interesting findings. The evaluation showed a trade off problem in RAG systems based on these kinds of LLMs, which is that there is a trade-off between a creative, teacher-like and human-like persona and strict factual grounding (which directly corresponds to faithfulness).

The limitation of this system though is that the corpus is limited to 12 documents (of which two are particularly comprehensive but still not broad enough to cover all possible questions and latest up to date research). Some of the references are older and need updating to align with latest information and advice. The system relies on a single LLM (gemini 2.5 flash which is a simple model optimized for speed) and a basic prompt strategy. In the future, expanding the corpus with more diverse and specialized sources would be the best way to go.