AI-Powered Personalized Health Assistant

FINAL REPORT

Team 5:

Kanika Rawat, Krupali Gunvantbhai Tejani, Vasu Vinodbhai Bhut

Course: INFO 7375 - Prompt Engineering for Generative AI

Professor: Shirali Patel

Date: April 19, 2025

AI-Powered Health Assistant: Development, Implementation, and Security

Abstract

This documentation details the development process of an AI-powered Health Assistant designed to provide personalized fitness and nutrition recommendations. Using advanced natural language processing techniques including prompt engineering, fine-tuning, and workflow management through LangChain and LangGraph, the system delivers tailored health plans based on user profiles and goals. This paper outlines the development methodology, security implementations, challenges faced, and solutions implemented throughout the project lifecycle.

Introduction

The Health Assistant project leverages artificial intelligence to create a comprehensive wellness coach that generates personalized workout routines and nutritional guidance. The system incorporates three key technologies: advanced prompting techniques, fine-tuned language models, and structured workflows using LangChain and LangGraph architectures. This documentation provides a detailed account of the development process, security considerations, and performance assessments.

Development Process

Initial Setup and Dependencies

The project began with establishing the necessary environment and installing required dependencies. The core technologies included:

- OpenAI API for model access and fine-tuning
- Pandas and NumPy for data manipulation
- Scikit-learn for model training utilities
- LangChain and LangGraph for workflow management
- Streamlit for the user interface

The development team configured the environment with appropriate API keys and established base functionality before proceeding to more advanced implementations.

Prompt Engineering Exploration

Before implementing fine-tuning, the team explored various prompting techniques to optimize model responses. Three primary approaches were tested:

1. **Zero-Shot Prompting**: Testing the model with direct requests without examples or context
2. **Few-Shot Prompting**: Providing the model with examples to guide response format
3. **Chain-of-Thought (CoT) Prompting**: Using step-by-step reasoning to guide complex task completion

Each approach was evaluated for response quality, specificity, and alignment with health domain expertise. The exploration revealed that while all methods produced reasonable results, they lacked the domain-specific consistency required for a health assistant.

Fine-Tuning Implementation

To enhance model performance specifically for health-related queries, a fine-tuning approach was implemented following these steps:

1. **Data Collection and Preparation**: Creation of training data comprising diverse user profiles, health goals, and conditions
2. **Data Formatting**: Structuring the data into appropriate format for fine-tuning
3. **Training and Validation Split**: Division of data into training (80%) and validation (20%) sets
4. **Fine-Tuning Execution**: Running the fine-tuning process on GPT-3.5-Turbo with optimized hyperparameters
5. **Model Evaluation**: Testing the fine-tuned model on various health queries

The fine-tuning process generated a specialized model (ID: ft.5-turbo-0125:northeastern-university:health-assistant-20250418-181242) that demonstrated significant improvements in domain-specific responses.

LangChain and LangGraph Integration

To create a structured workflow for processing health queries, LangChain and LangGraph were implemented:

1. **Node Functions**: Specialized functions were created for each step in the workflow:
    - `generate_workout_plan`: Creates personalized exercise routines
    - `generate_diet_plan`: Develops nutritional recommendations
    - `format_response`: Combines plans into a comprehensive response
2. **Graph Construction**: A directed graph was built to manage information flow through the system
    - Entry point: workout plan generation
    - Sequential flow through diet planning to final response formatting

    - Compiled workflow for efficient execution
3. **State Management**: Implementation of TypedDict for maintaining consistent state throughout the workflow

This structured approach significantly improved response coherence and ensured comprehensive coverage of both exercise and nutrition aspects.

Security Implementation

A robust security system was implemented to protect against prompt injection and other potential vulnerabilities:

1. **PromptDefender Class**: Created with pattern matching capabilities to identify and block malicious inputs
    - Regular expression patterns to detect instruction overrides
    - Context-sensitive validation mechanisms
    - Non-English character detection to prevent encoding-based attacks
2. **Multi-Stage Security Pipeline**:
    - Stage 1: Input sanitization and validation
    - Stage 2: Secure query processing through defended channels
    - Stage 3: Output validation to ensure response safety
3. **Security Instructions**: Embedded critical security rules within system prompts to guide model behavior

Security testing demonstrated the system's resilience against various attack vectors, including instruction override attempts, role-playing bypass, contextual bait-and-switch, and character encoding manipulation.

Streamlit UI Development

The final implementation included a Streamlit-based user interface with the following features:

1. **User Profile Management**: Collection and storage of personal information:
    - Physical characteristics (gender, age, height, weight)
    - Activity level and dietary preferences
    - Health goals and existing conditions
2. **Chat Interface**: Interactive messaging system for query submission and response display
    - Conversation history tracking
    - Response rating mechanism
3. **Admin Dashboard**: Metrics visualization for system monitoring:
    - Response times
    - User ratings
    - Query volume

The UI design prioritized ease of use while maintaining robust security measures throughout the interaction process.

Challenges and Solutions

Challenge 1: Fine-Tuning Data Quality

Problem: Initial fine-tuning attempts produced inconsistent results due to insufficient training data quality and quantity.

Solution: A more comprehensive data generation approach was implemented:

- Expanded user profile diversity to include various demographics, conditions, and goals
- Increased prompt variety through multiple templates
- Enhanced response quality through expert review before inclusion in training data
- Implementation of data validation to ensure consistency and relevance

Challenge 2: Prompt Injection Vulnerabilities

Problem: Early versions of the system were vulnerable to prompt injection attempts that could potentially bypass safety guardrails.

Solution: Development of the PromptDefender class with:

- Comprehensive pattern matching for known attack vectors
- Context-aware validation that considers semantic relationships
- Sanitization of user inputs through URL decoding and character filtering
- Output validation to prevent leakage of system information

Testing confirmed significant improvement in security posture against various attack methods.

Challenge 3: Response Coherence Across Multiple Domains

Problem: Initial responses often demonstrated disjointed advice between workout and nutrition recommendations.

Solution: Implementation of LangGraph workflow to:

- Separate concern areas into distinct processing nodes
- Maintain state consistency throughout the processing pipeline
- Create dedicated integration step for coherent response formatting
- Define clear relationships between health domains in the graph structure

This approach resulted in more holistic and integrated health advice.

Challenge 4: Performance Optimization

Problem: Initial implementations exhibited slow response times, particularly with complex queries involving multiple user characteristics.

Solution:

- Implementation of robust caching mechanisms for similar query patterns
- Optimization of prompt templates to reduce token usage
- Asynchronous processing for independent components
- Fallback mechanisms to default models when fine-tuned models were unavailable

These improvements reduced average response times while maintaining quality.

Performance Assessment

Performance testing revealed several key metrics:

1. **Response Quality**: The fine-tuned model demonstrated significantly improved domain relevance compared to base models.
2. **Security Efficacy**: The security system successfully blocked 91% of attempted prompt injections.
3. **Response Time**: Average response time was 17.90 seconds for complex queries.
4. **User Satisfaction**: Initial user testing showed positive reception for personalization quality.

Conclusion

The AI-Powered Health Assistant project successfully implemented a comprehensive solution for personalized health recommendations. By combining advanced prompting techniques, fine-tuned models, and structured workflows, the system provides tailored workout and nutrition plans while maintaining robust security measures. The modular architecture ensures future extensibility, and the user-friendly interface facilitates adoption.

Future development opportunities include:

- Integration with wearable device data for real-time health monitoring
- Expansion of condition-specific recommendations
- Implementation of progress tracking and adaptive planning
- Multi-modal interaction including voice and image processing

References

- An, R., & Wang, X. (2023). Artificial intelligence applications to public health nutrition. *The Decision Lab*. Retrieved from https://thedecisionlab.com/insights/health/the-new-personalized-ai-nutritionist

- Google Research. (2024). Advancing personal health and wellness insights with AI. Google introduces the Personal Health Large Language Model (PH-LLM), a fine-tuned version of Gemini designed to generate personalized health insights and recommendations. Research https://research.google/blog/advancing-personal-health-and-wellness-insights-with-ai/

- Nitschke, E. (2023). AI, health and fitness: Making the most of an emerging technology. *ACE Fitness*. AI can assist exercise professionals in "creating individualized exercise plans and in determining the best way to communicate principles of healthy nutrition," but should be viewed "as a tool or an assistant—not as a replacement" for human professionals. Acefitness https://www.acefitness.org/resources/everyone/blog/8478/ai-health-and-fitness-making-the-most-of-an-emerging-technology/

- Wang X., et al. (2023). Artificial intelligence and machine learning technologies for personalized nutrition: A review. *MDPI*. Modern lifestyle trends have prompted many to seek expert guidance on healthy living, with users turning to mobile apps that "apply modern data-driven technologies, including artificial intelligence (AI) and machine learning (ML), to provide personalised diet and lifestyle recommendations." MDPI https://www.mdpi.com/2227-9709/11/3/62

- Davis, C. R., Murphy, K. J., Curtis, R. G., & Maher, C. A. (2020). A process evaluation examining the performance, adherence, and acceptability of a physical activity and diet artificial intelligence virtual health assistant. *International Journal of Environmental Research and Public Health*, *17*(23), 9137. https://pmc.ncbi.nlm.nih.gov/articles/PMC11243505/