

Final Project Proposal: Mental Health Detection & Support System

1. Project Description

This project, titled "Synapse AI: Early Detection and Conversational Support for Mental illness," aims to facilitate the **early and private detection** of potential mental health conditions using a two-pronged approach.

The **primary deliverable** is a **Machine Learning (ML) classification model** designed to analyze user-provided data to classify the patient's likelihood of experiencing a mental illness, resulting in a binary output (**Yes/No**).

The **secondary deliverable** is a sophisticated **AI Chatbot** powered by a **Retrieval-Augmented Generation (RAG) system**. This chatbot, built on strong, pre-vetted documentation and clinical guidelines, serves two critical functions:

- 1. **Support for Ambiguous Cases:** Providing personalized, contextual assistance and non-diagnostic guidance to patients whose classification scores fall into a neutral or ambiguous zone.
- 2. **General Wellness Support:** Offering a safe, conversational space for all users to naturally discuss their feelings, symptoms, and concerns, thereby enhancing user engagement and providing a supportive safety net.

The core innovation lies in the secure and seamless integration of these two systems to provide a high-accuracy service with humanistic conversational support.

2. Group Members & Roles

Member	Assigned Role	Key Responsibilities
Sherif Karam	--	Design and implementation of the RAG pipeline, document chunking, vector database setup, and prompt engineering for the chatbot.
Hussen Jeballi	--	Core model selection (e.g., transformer models), training, hyperparameter tuning.
Salma Ahmed	--	Data acquisition, cleaning, feature engineering, and model evaluation metrics analysis.
Habiba Yousef	--	Project oversight, timeline management, stakeholder communication, final documentation.
Ahmed Hazem	--	API development, cloud deployment (AWS/GCP/Azure), backend infrastructure, security, and latency optimization.
Nouran Shawkat	--	Development of the user interface (web/mobile), ensuring an intuitive and secure experience for both the ML screening and the chatbot interaction.

3. Team Leader: Sherif Karam

4. Objectives

- 1. **Achieve High-Fidelity Classification:** Develop and train the ML model to achieve a clinically useful level of accuracy, specifically minimizing the **False Negative (FN)** rate to ensure early detection is not missed.
- 2. **Establish Secure RAG System:** Build a robust RAG pipeline, indexing high-quality, non-diagnostic mental health literature to provide accurate, contextually relevant, and safe conversational responses.
- 3. **Ensure Scalable Deployment:** Deploy the integrated system (ML API + RAG Chatbot) on a reliable cloud platform, guaranteeing low latency and high performance.
- 4. **Validate User Experience:** Design a frontend that is intuitive, empathetic, and promotes natural interaction, leading to high user satisfaction and retention.

5. Tools & Technologies

Category	Tools & Technologies	Rationale
Programming Language	Python	Industry standard for Data Science, ML, and NLP development.
ML/Deep Learning	Scikit-learn, Pandas, and NumPy	Essential libraries for data manipulation, feature engineering, and implementing classical ML algorithms
Natural Language Processing	Hugging Face (Transformers) (till now)	Access to pre-trained models for text processing and feature extraction for the ML classifier.
RAG/Chatbot	--	Frameworks for orchestrating the RAG pipeline; ChromaDB for fast, efficient vector search over documentation.
API & Backend	--	Lightweight, high-performance Python frameworks for serving the ML model as an API.
Deployment & Cloud	--	Cloud platform for scalable and reliable deployment of the application and database (e.g., Firebase, S3/Cloud Storage).

6. Milestones & Deadlines

Milestone	Key Deliverables	Deadline
Phase 1: Data Acquisition & Baseline Modeling	Finalized data source, Preprocessing pipeline complete, Initial (low performance) ML model trained and evaluated.	--
Phase 2: Core ML Development & Optimization	Finalized high-performance ML model, Error Analysis (FP/FN) complete, Finalized feature set.	--
Phase 3: RAG System Development	Indexed and chunked knowledge base, Operational RAG pipeline (proof-of-concept), Basic chatbot integration.	--
Phase 4: API & Deployment	-- --	--
Phase 5: Final Testing, UI & Presentation	Completed Frontend UI, End-to-end system testing, Final performance audit, Project documentation, and presentation materials.	--

7. KPIs (Key Performance Indicators)

1. Data Quality

Metric	Target Value	Rationale
Percentage of missing values handled:	--	Critical for preserving data integrity and avoiding model bias.
Data accuracy after preprocessing:	--	Verification that feature engineering and scaling/encoding is implemented correctly.
Dataset diversity (representation of different categories):	--	Ensuring the training data covers a wide range of relevant demographics and conditions to promote generalization.

2. Model Performance

Metric	Target Value	Rationale
Model accuracy (Accuracy/F1-Score):	--	---
Model prediction speed (Latency):	--	A near-instantaneous response is critical for a positive user experience.
Error rate (False Negative Rate):	--	The most critical metric—we must minimize missing a detection, as a False Negative can have serious consequences.

3. Deployment & Scalability

Metric	Target Value	Rationale
API uptime:	--	Ensures reliable, near-constant service availability.
Response time per request (end-to-end):	--	Measures the time from user clicking "submit" to seeing the result, including both ML and RAG-based systems.
Real-time processing speed (RAG Context Retrieval):	--	Ensures the chatbot can retrieve and integrate documentation context quickly enough for a natural conversation flow.

4. Business Impact & Practical Use

Metric	Target Value	Rationale
Reduction in manual effort (initial screening):	--	The AI system handles the initial, time-consuming screening process, freeing up human professionals.
Expected cost savings:	--	Achieved through automation of the initial screening process.
User satisfaction (post-interaction survey):	--	Reflecting the system’s empathetic UI, helpful RAG responses, and reliable ML performance.