# 1. System Architecture

[User Input]

↓

[Knowledge Retrieval (FAISS Vector Search)]

↓

[Prompt Construction based on Task Type]

↓

[Text Generation (Gemini 1.5 Flash Model)]

↓

[Educational Content Output]

# 2. Implementation Details

EduMind integrates Retrieval-Augmented Generation (RAG) and Prompt Engineering techniques. The system uses Sentence Transformers to embed mental health knowledge base content, FAISS for efficient vector search, and Google Generative AI (Gemini 1.5 Flash) for educational text generation.
The front-end user interface is built using Streamlit, providing a simple and intuitive experience for users to enter queries, select content types, and receive generated educational outputs.

Key components:

- Knowledge Base: The knowledge base (kb.csv) was constructed by using GPT to extract and summarize key concepts from the glossary provided by

MQ Mental Health. This ensured the educational materials were accurate, structured, and aligned with reputable mental health references.

- Embedding Model: all-MiniLM-L6-v2 for sentence encoding.
- Vector Database: FAISS for fast semantic search.
- Prompt Engineering: Dynamically crafted prompts tailored to user-selected tasks.
- Text Generation Model: Google Gemini-1.5 Flash 8B model
- User Interface: Implemented using Streamlit.

## 3. Performance Metrics

The performance metrics reported here are based on rough estimates from manual testing during development. Retrieval generally took around 1 second, and content generation using the Gemini model typically completed in about 3 seconds. Output relevance was observed to be high in most cases, but no formal automated benchmarking was performed.

## 4. Challenges and Solutions

During the development of EduMind, two major challenges emerged.

The first challenge was building a suitable knowledge base. After deciding to focus on mental health educational content, I initially searched Kaggle for relevant datasets but found few that were high-quality and directly applicable. I eventually discovered the MQ Mental Health Glossary[1], which offered excellent content, but unfortunately, there was no API or downloadable dataset available. To solve this, I used GPT to help extract and summarize key entries from the glossary, constructing a customized knowledge base (kb.csv) tailored to the project's needs.

The second challenge was selecting an appropriate language model. I initially attempted to use large open-source models like DeepSeek through Hugging Face pipelines, but found them too resource-intensive for practical deployment. I also experimented with smaller models like Qwen and DistilGPT2; however, their performance was not sufficient for generating high-quality educational outputs. To overcome this, I shifted to an API-based solution by integrating Google’s Gemini model, which provided strong performance and seamless integration within the Streamlit framework.

## 5. Future Improvements5.

Future work could focus on expanding the RAG knowledge base to cover a wider range of mental health topics with more comprehensive and higher-quality content.Another improvement would be adding multilingual generation capabilities to make the system accessible to a global audience.

[1] https://www.mqmentalhealth.org/glossary/

## 6. Ethical Considerations

EduMind is designed solely for educational purposes and does not provide professional medical advice. Additionally, the system does not collect or store any user data during operation.