



**ORTA DOĞU TEKNİK ÜNİVERSİTESİ**  
**MIDDLE EAST TECHNICAL UNIVERSITY**

# **METU Personal Tutor**

**Berk Açikel**  
**Yusuf Sami Lök**  
**Computer Engineering Department**  
[berk.acikel@metu.edu.tr](mailto:berk.acikel@metu.edu.tr)  
[yusuf.lok@metu.edu.tr](mailto:yusuf.lok@metu.edu.tr)

**02/01/2025**

# Outline

- Problem definition
- Our method
- Data collection
- Retrieval mechanisms
- Our solution for retrieval
- Website

## **Problem Definition**

- METU has a vast amount of public data available for students, faculty, and staff.
- Information is often scattered across multiple formats (e.g., PDFs, documents, websites).
- Students and faculty face challenges when searching for specific answers.

**Key Problem:** Finding the right answer in the least amount of time is difficult and inefficient.

# Our Method

- Relevant data is collected with the retrieval mechanism.
- The question is appended to retrieved documents and sent to LLM.
- Mental health analysis is done by sending the history to LLM.

**Used Model:** Llama-3-70b with Groq API

# Dataset Collection

## Sources:

- Collected data from all websites under the domain [metu.edu.tr](https://metu.edu.tr).
- Data was extracted in a structured format to ensure relevant information could be easily accessed.

## Data Types:

- Documents, webpages, PDFs, and other METU-related public information.

# Retrieval Mechanisms

## BM25

### What is BM25?

- A popular algorithm for ranking documents based on their relevance to a query.
- Built on **TF-IDF** (Term Frequency-Inverse Document Frequency), with added improvements for handling term frequency and document length.

### How It Works:

- Scores documents by considering:
  - **Term Frequency (TF)**: How often a term appears in a document.
  - **Inverse Document Frequency (IDF)**: How rare or important the term is across all documents.
  - **Document Length**: Normalizes scores to avoid favoring longer documents.

# Retrieval Mechanisms

## BM25

### Strengths:

- Simple and effective for many information retrieval tasks.
- Robust to irrelevant terms and noisy data.

### Limitations:

- Doesn't understand context or semantic meaning.
- Works best with exact matches between query terms and document terms.

# Retrieval Mechanisms

## Vector Embeddings

- Documents with embeddings that have highest cosine similarity with the query are retrieved.

**Used embedding model:** multi-qa-mpnet-base-dot-v1

- Effective for semantic understanding but may miss certain exact matches.
- Slower than BM25.



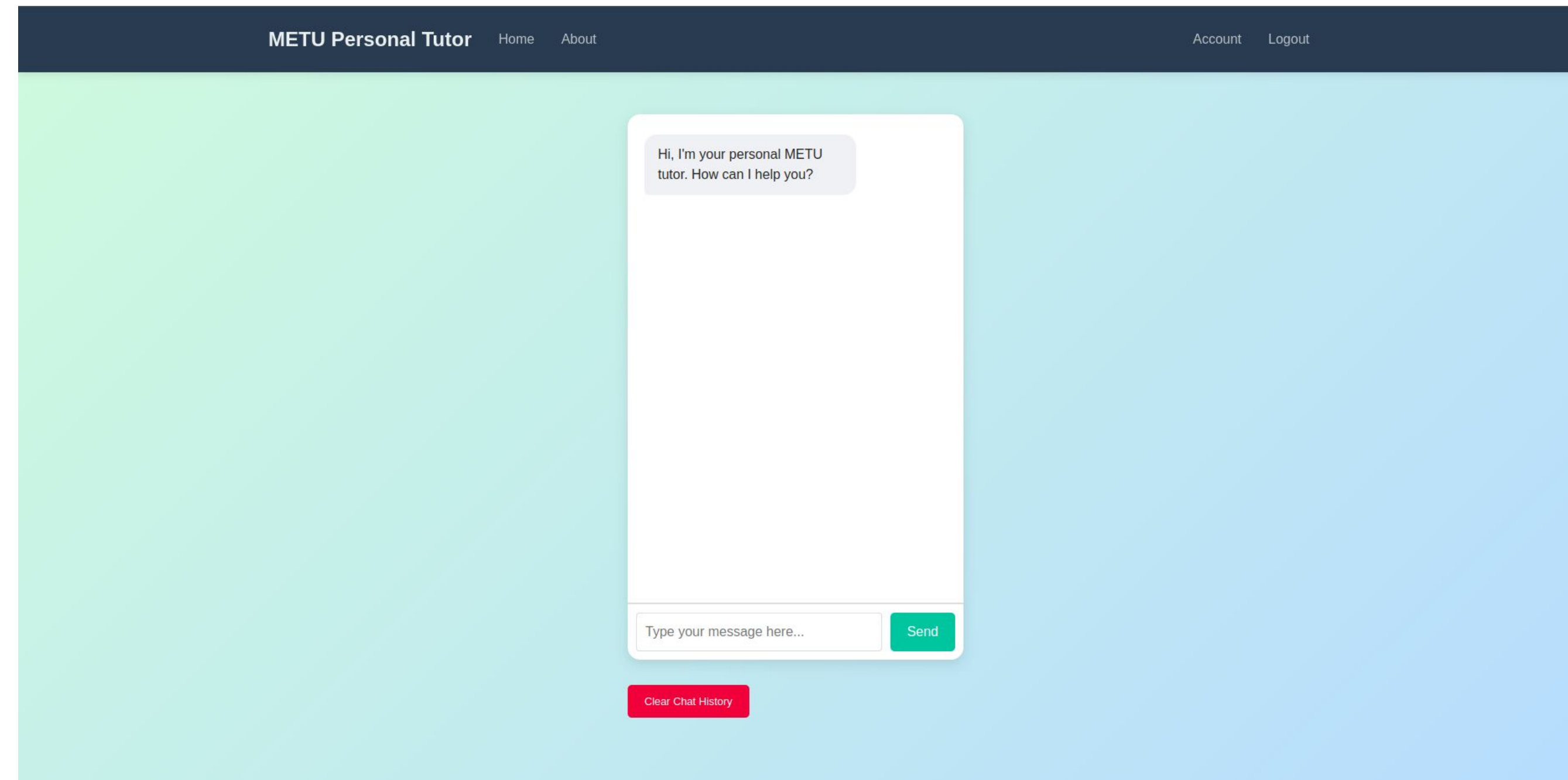
# Our Solution For Retrieval

- Both retrieval methods are applied.
- Highest ranked documents from both methods are sent to the LLM.
- Documents with both highest exact matches and semantic matches are retrieved.
- Since both methods are applied, slower than usage of a single method.

# METU Personal Tutor Website

## Features:

- User-friendly interface powered by Flask.
- Basic authentication for secure access.
- Chatbot interface for direct interaction with users.
- Mental health analysis based on user interactions with the chatbot.





**ORTA DOĞU TEKNİK ÜNİVERSİTESİ**  
**MIDDLE EAST TECHNICAL UNIVERSITY**

**Thanks**

**Berk Açikel**  
**Yusuf Sami Lök**

**02/01/2025**