

# ***MEDCOACH***

## Project Proposal

Elevate medical training with AI-powered mentor that immerses in realistic patient cases and delivers instant expert feedback to sharpen diagnostic skills.

Maya Kimhi & Mai Werthaim

# ***Problem statement***

## **Project Motivation**

Medical students and interns often rely on static teaching methods that lack the complexity of real-world clinical scenarios. They need an interactive, case-based platform to sharpen their diagnostic reasoning and decision-making skills.

By harnessing the power of large language models, we can simulate authentic patient interactions, generate diverse medical questions, and provide immediate feedback—ultimately bridging knowledge gaps and enhancing the learning experience.

# *Problem statement*

## Problem Definition



### *Inputs*

#### **Named Entity Recognition (NER)**

Medical questions or patient case descriptions from a dataset (MedQuad).

#### **Question Answering, Semantic Evaluation**

Student responses or queries about diagnoses and treatment.



### *Outputs*

#### **Medical Text Generation**

A clinical case scenario generated by an LLM based on a medical QA dataset (includes age, symptoms, etc.).

#### **Question Answering**

Feedback comparing student answers to reference answers.

# ***Problem statement***

## **Problem Challenges**

- Evaluating open-ended responses which requires semantic and medical understanding beyond keyword matching.
- Accurately interpreting specialized medical terminology and context
- Generating medically accurate and coherent clinical scenarios.
- Providing meaningful feedback.

# *Training and Test Data*

**MedQuad (Kaggle):** A Public dataset that contains over 16,000 patient questions from real-life situations and their answers.



MedQuad dataset is divided into 3 columns:

- Question Type – The type of medical question (String), there are 31 distinct types
- Question – The medical question posed by the patient (String).
- Answer – The expert response to the medical question (String).

**Synthetic Data Generation:** Using LLM to create case scenarios or variations of existing questions from the MedQuad dataset for broader coverage.

- Augmentation
  - LLM to identify explicit entities.
  - LLM to rephrase omitting explicit entity reference but keeping the context.

# *Training and Test Data*

## **Example**

**MedCoach:** Mr. John Miller, a 42-year-old male, presented to the emergency department with a 12-hour history of progressively worsening neurological symptoms. Approximately 24 hours prior, he had attended a family picnic where he consumed home-canned vegetables that were later suspected to be improperly sterilized. Initially, Mr. Miller noticed a subtle blurring of his vision and a sensation of heaviness around his eyes. Within a few hours, symptoms: double vision, drooping eyelids, and a noticeable difficulty in speaking as his speech became slurred. difficulty swallowing, a persistently dry mouth, and generalized muscle weakness. What is Mr. Miller's diagnosis?

**Student:** Did Mr. Miller undergo a physical examination?

**MedCoach:** On physical examination, the clinical team observed bilateral ptosis, impaired cranial nerve responses, and decreased muscle strength in both his upper and lower limbs. What Mr. Miller has?

**Student:** Botulism

**MedCoach:** Correct

# *Evaluation*

## *Student Answer Assessment*

### **Comparison with Baseline**

Evaluating model performance by comparing it to a simple baseline (TF-IDF + cosine similarity).

### **Classification Metrics**

Accuracy, Precision, Recall, F1-score

## *Generated Clinical Cases*

### **Medical Plausibility & Completeness**

BERTScore between generated cases and original data to gauge linguistic quality.

### **Cases Diversity**

Embedding distance to measure semantic differences between generated cases



# *THANK you*

Do you have any questions for us?