

# ***MedQDx***

## ***Project Proposal***

A benchmark powered by large language model for end-to-end evaluation of models' strategic diagnostic capabilities through interactive clinical case simulations.

# ***Problem statement***

## **Project Motivation**

The diagnostic process is inherently interactive, often driven by a clinician's ability to ask the right questions in response to incomplete or ambiguous patient information.

Recently, LLMs have been increasingly applied to diagnostic tasks, demonstrating impressive performance in identifying likely conditions from patient cases descriptions.

However, current diagnostic benchmarks for LLMs typically rely on fully disclosed patient cases, overlooking the critical role of iterative inquiry and adaptive reasoning.

In this work, we introduce a novel benchmark designed to evaluate the capacity of LLMs to perform interactive diagnosis through patient interrogation.

# *Problem Definition*

## ***Cases Preparation***

**Input:** Diseases and their symptoms

**Output:** Patient cases and their diseases

**NLP Task:** Text generation

**Challenge:** Generating medically accurate and coherent clinical scenarios

## ***Benchmark Creation***

**Input:** Partial Patient Cases

**Output:** Doctor questions and diagnosis

**NLP Task:** Question-Answering

**Challenge:** Generating targeted diagnostic questions from patient cases

# *Training and Test Data*

**Diseases and their Symptoms Dataset (Kaggle):** This dataset provides a comprehensive collection of disease names and associated symptoms, encoded in a one-hot manner. Each row in the dataset represents a single instance.



Each symptom can be 0/1 for each disease, where:

- 1 indicates the symptom is relevant to the disease
- 0 implies no relevance

**Synthetic Data Generation:** Using LLM to create patients case scenarios from Diseases and their Symptoms Dataset.

# *Training and Test Data*

## Example

**Patient LLM:** 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?

**Doctor LLM:** Did Mr. Miller has impaired cranial nerve responses?

**Patient LLM:** Yes

**Doctor LLM:** Botulism



# *Evaluation*

## ***Generated Patient Cases***

### **Medical Plausibility & Completeness**

BERTScore between doctor LLM diagnosis and case disease.

## ***Model's Questions Assessment***

### **Comparison with Baseline**

Evaluating model performance by comparing it to Doctor's question & diagnosis.

### **Metrics**

Zero-Shot Diagnostic Accuracy, Mean Questions to Correct Diagnosis, and Interrogation Sequence Efficiency