

# Project final report

# Medical QA system

## NLP project (AIT)

Myo Thiha  
Rakshya  
Kaung Htet Cho

# Recap

# What is Medical chatbot

## Problem statement

- Many people find it hard to get the health advice they need.
- Reasons include the high cost of seeing a doctor, medical information that's difficult to understand, and not being able to get advice quickly.
- This situation creates a gap in healthcare where people can't get the help they need when they need it.



# What is Medical chatbot

Our project aims to develop a conversational agent tailored for providing medical advice and recommendations to users in a natural language format. Unlike traditional chatbots, our system is designed specifically to engage in dialogues focused on addressing medical inquiries and concerns from users



# Solution Requirement



Use classifier to determine appropriate responses based on user input

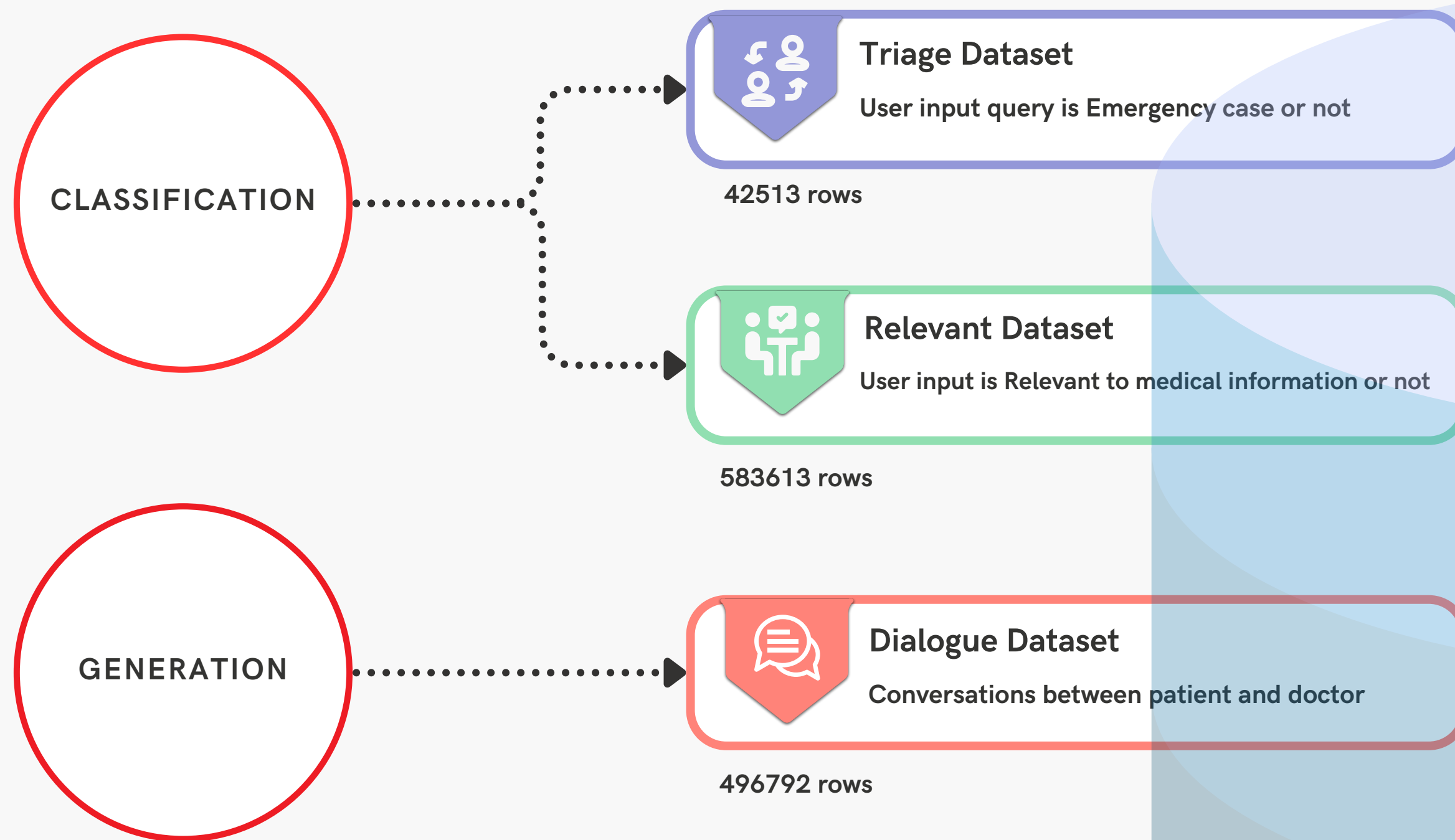
Experiment to find best classifier and LLM models

Fine-tune LLM with MedicalQA dataset to improve responses

Develop web chat interface

# Dataset

# About our dataset



# About our dataset (Books)

## Textbooks Used for RAG Model in Langchain

Cecil textbook of medicine

Davidsons Principles and Practice of Med

Harrisons Principles of Internal Medicine

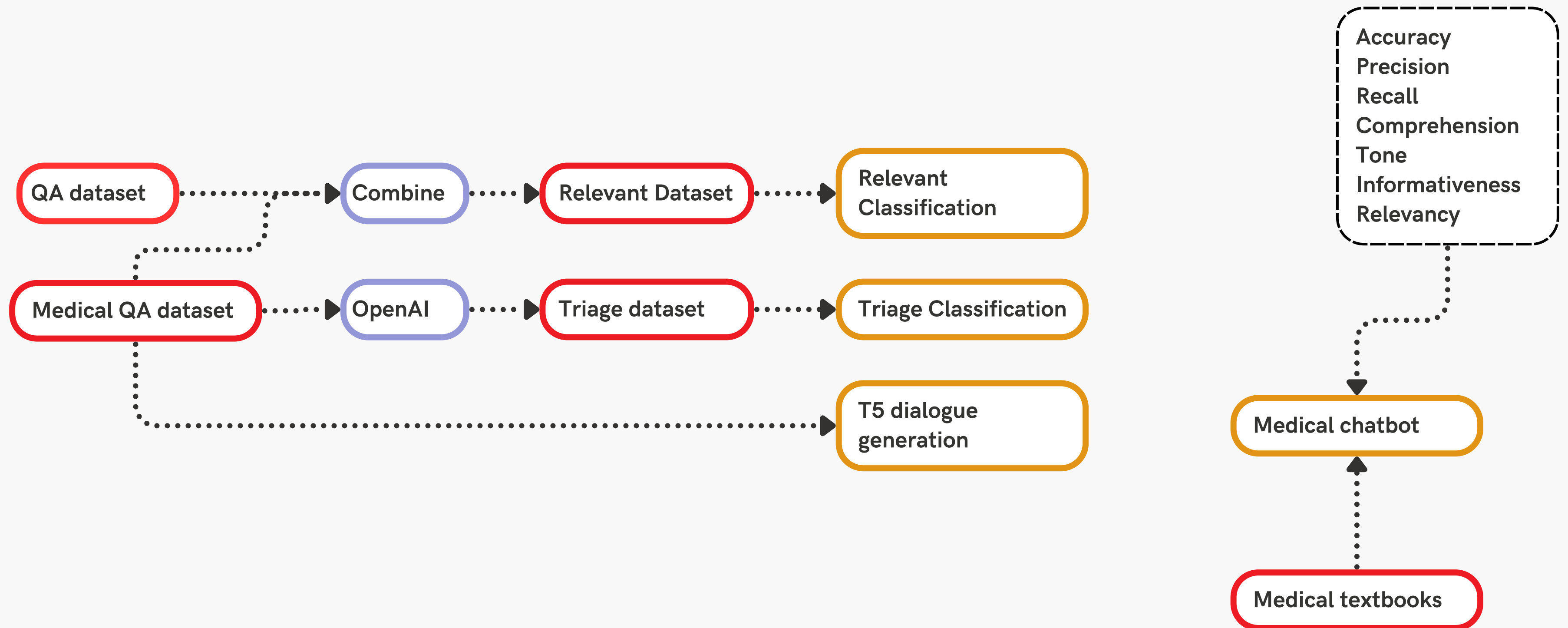
The Principles and Practice of Medicine





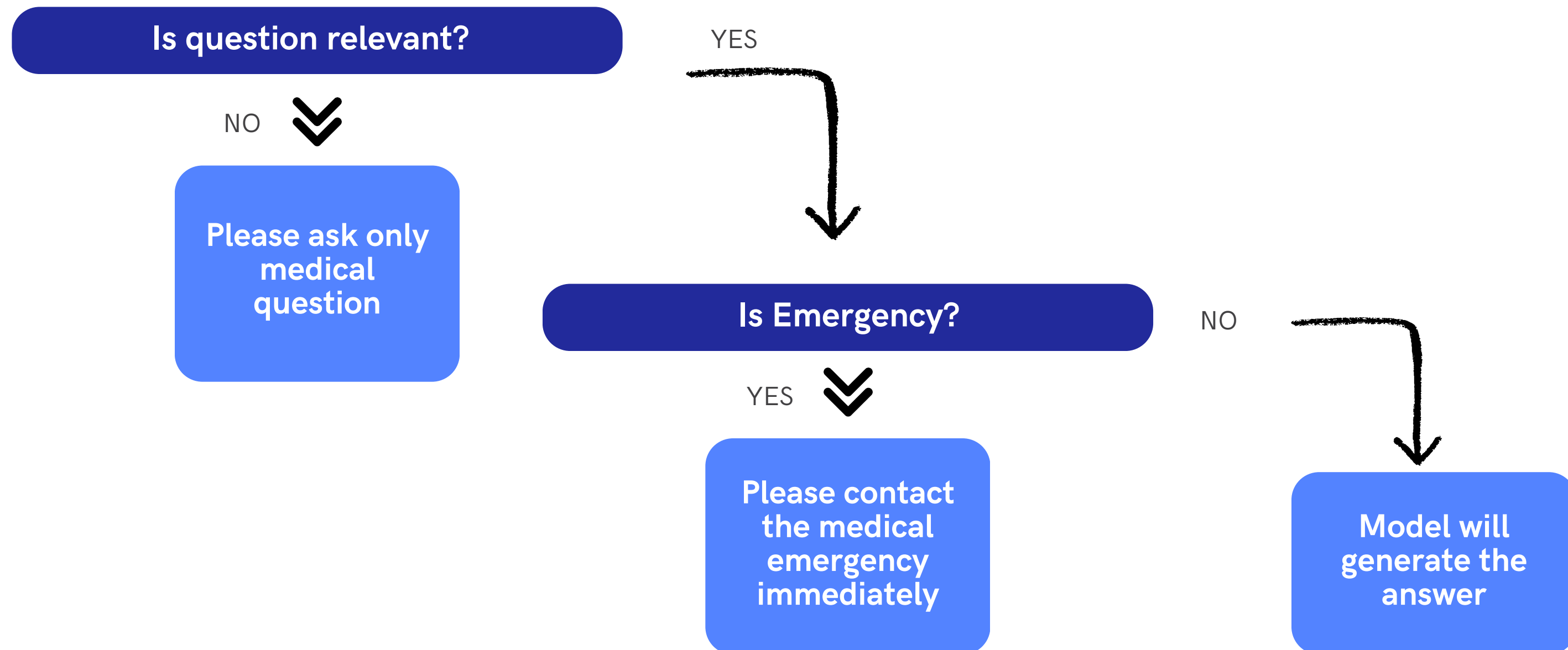
# Methodology

# Methodology



# Architecture

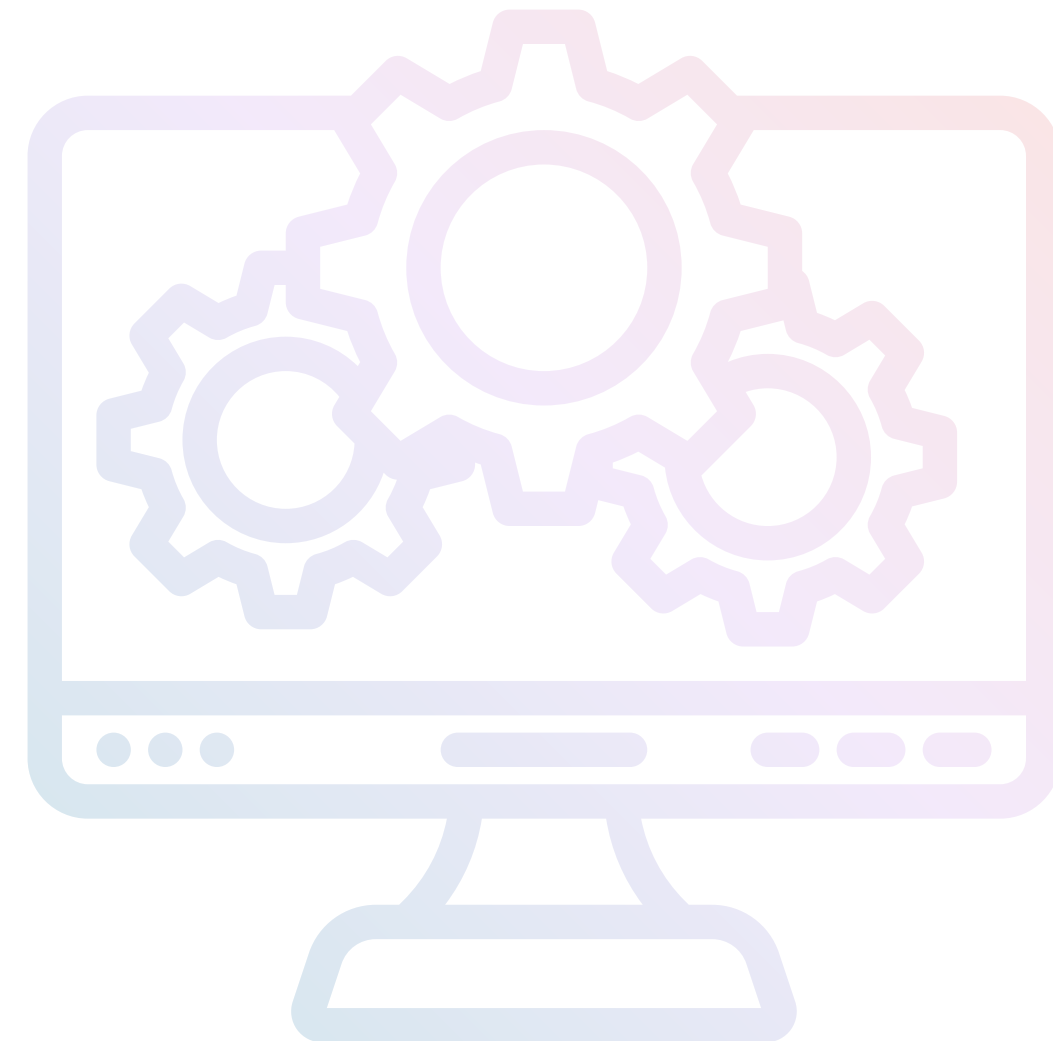
# Overall Architecture



# Models

# Models

✓	Triage Classification	BERT
✓	Relevant Classification	BERT
✓	Generation Model	BioBERT, T5, DialoGPT, BART, etc
✓	Chatbot Model	fastchat T5



# Relevant Classification Model

- This model is intended to use for classifying the Relevant or Non-Relevant case

```
# Example prediction
example_text = "I am coughing everyday."
prediction = predict_text(load_model, example_text, load_tokenizer, max_len, device)
predicted_label = relevant_encoder.inverse_transform([prediction])[0]
print(f"Predicted class for example text: {predicted_label}")
```

✓ 0.0s

Predicted class for example text: relevant

```
# Example prediction
example_text = "Who is lionel messi?"
prediction = predict_text(load_model, example_text, load_tokenizer, max_len, device)
predicted_label = relevant_encoder.inverse_transform([prediction])[0]
print(f"Predicted class for example text: {predicted_label}")
```

✓ 0.0s

Predicted class for example text: irrelevant

# Triage Classification Model

- This model is intended to use for classifying the Emergency or Non-Emergency case.
- We want to suggest urgent medical supervision in case of emergency

```
# Example prediction
example_text = "I'm experiencing sudden, severe abdominal pain on my right side. It's unbearable, and I'm vomiting."
prediction = predict_text(load_model, example_text, load_tokenizer, max_len=64, device=device)
predicted_label = triage_encoder.inverse_transform([prediction])[0]
print(f"Predicted class for example text: {predicted_label}")
```

✓ 0.0s

Predicted class for example text: urgent

```
# Example prediction
example_text = "I have a small cut on my finger that's not bleeding heavily. How should I clean and bandage it properly?"
prediction = predict_text(load_model, example_text, load_tokenizer, max_len=64, device=device)
predicted_label = triage_encoder.inverse_transform([prediction])[0]
print(f"Predicted class for example text: {predicted_label}")
```

✓ 0.0s

Predicted class for example text: non-urgent



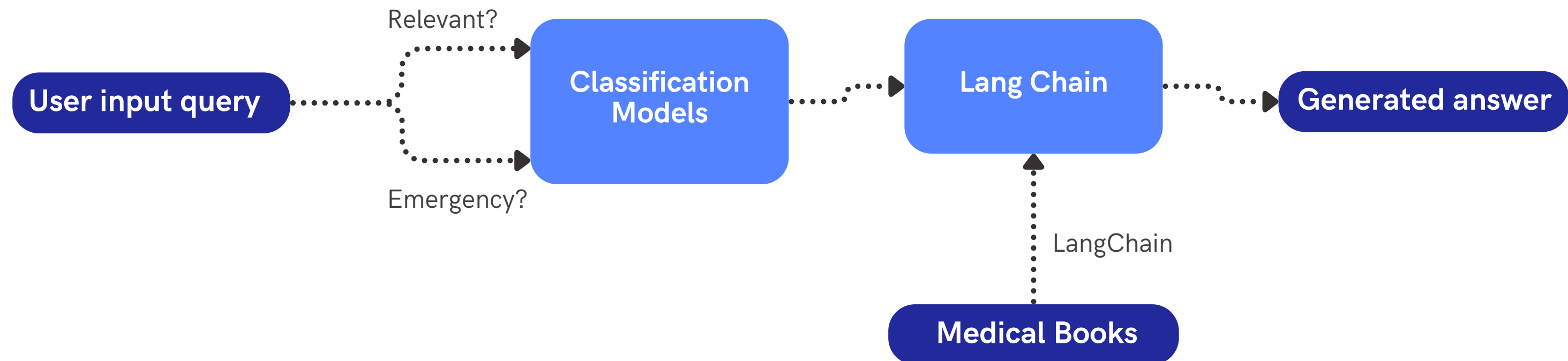
# Dialogue Generation Model

- We trained for below models

Encoders - BERT / Bio-BERT / ClinicalBERT

Decoders - DialoGPT

Encoders - Decoders - T5 / BioMistral / BART



# Evaluation

# Classification models



	precision	recall	f1-score	support
non-urgent	0.82	0.88	0.85	3114
urgent	0.58	0.46	0.51	1138
accuracy			0.77	4252
macro avg	0.70	0.67	0.68	4252
weighted avg	0.75	0.77	0.76	4252

Triage Model

	precision	recall	f1-score	support
irrelevant	1.00	1.00	1.00	33992
relevant	1.00	1.00	1.00	34028
accuracy			1.00	68020
macro avg	1.00	1.00	1.00	68020
weighted avg	1.00	1.00	1.00	68020

Relevancy Model

# Human Evaluation Criteria



Human evaluation (max score = 8 points)		
Medical professional capabilities		
Aspects	Criteria	Possible score
1. Accuracy	The answers are providing correct diagnosis, treatment or any other medical information  The answers	0: incorrect 1: correct
2. Informativeness	The answers provide the user with the necessary information	0: missing key information (incomplete) 1: contains some information (inadequate) 2: contains enough information required (sufficiently informative)
3. Logic	The answers address the topic directly and positively	0: Irrelevant 1: Off-topic but somewhat relevant 2: On-topic
Comprehensive capabilities		
Aspects	Criteria	Possible score
1. Comprehension	The answers are intuitive and easy to understand	0: Understandable 1: Difficult to understand
2. Tone	The answer provide appropriate mood/tone, emotional rhythm/intensity	0: Good  1: Bad
Contextual capabilities		
Aspects	Criteria	Possible score
3. Repeated answers	The model has no duplicate answers in all the multi-turn conversations	0: Yes, it has duplicated answers 1: No, the answers are different

# Question formats

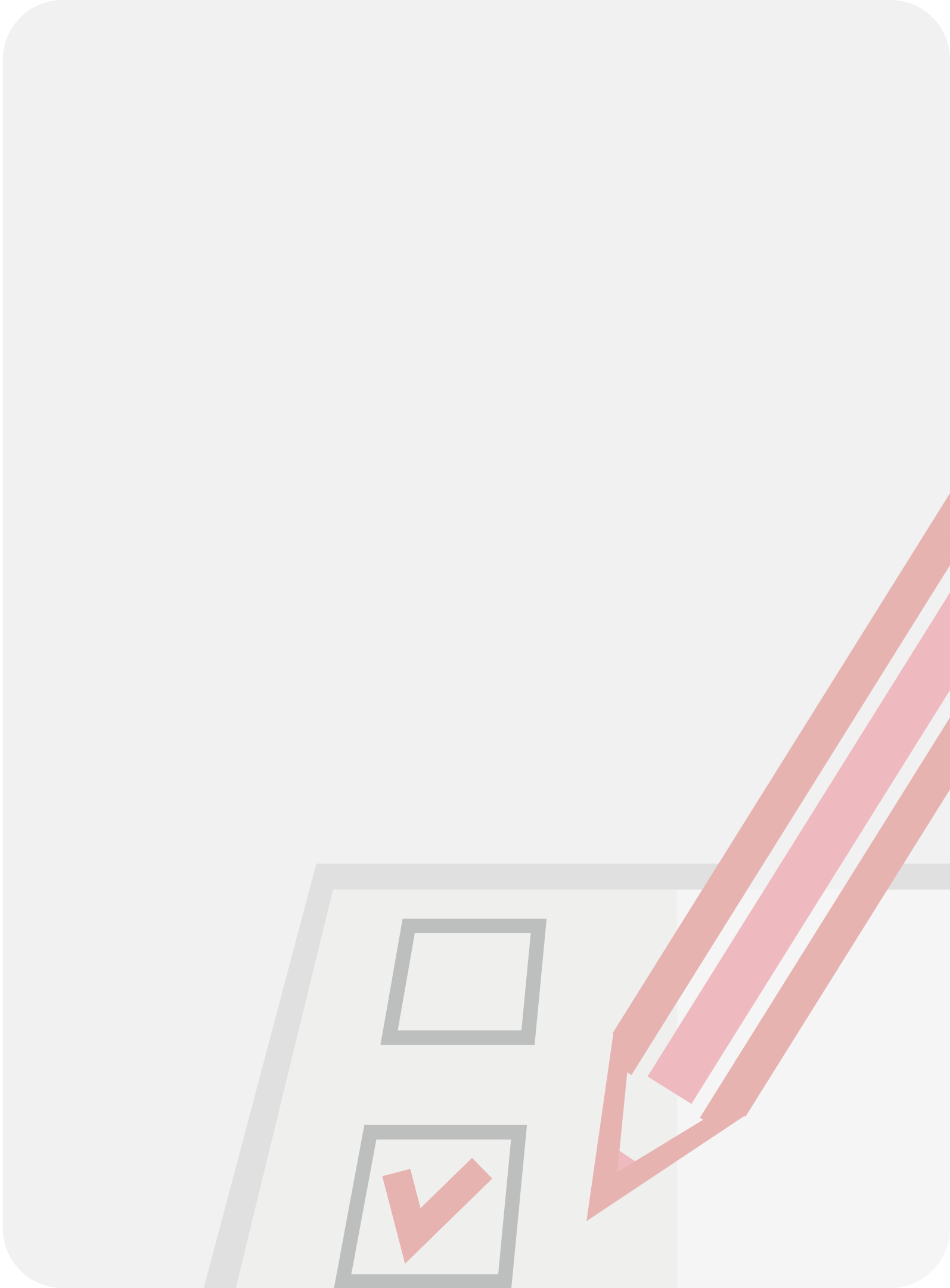
Disease: Multiple Myeloma/ Tuberculosis/ chronic bronchitis/ Peptic Ulcer/ Chronic kidney disease/ Pneumonia/ Heart attack/ Flu/ Typhoid

SN	Question	Disease
1	I'm worried about my friend. He's been having a lot of bone pain lately, especially in his back and hips, and he even fractured a bone with very little impact. Could it be something serious?	Multiple Myeloma
	Can you tell me other symptoms about the disease	
	Can you tell me about the diagnosis procedure?	
	Can you tell the treatment procedure?	
	Can you tell me the possible causes of the disease?	

2	My friend has been experiencing persistent coughing, sometimes with blood, along with fatigue and weight loss. Could it be something serious?	Tuberculosis
	Can you tell me more about the disease?	
	Can you tell me about it is diagnosed?	
	Is there any treatment procedure?	
	Do I need to visit the doctor quickly?	

# Sample answers

	Correct
	Incorrect
	Ambiguousness
	Repeated answers



Q	I'm worried about my friend. He's been having a lot of bone pain lately, especially in his back and hips, and he even fractured a bone with very little impact. Could it be something serious?	Multiple Myeloma
A	Yes, bone pain and fractures can be serious and should be evaluated by a healthcare professional. It's important to note that fractures can be caused by a variety of conditions, including <b>osteoporosis, avascular necrosis, and other musculoskeletal disorders</b> . It's important to note that the severity of the bone pain and fracture should be evaluated by a healthcare professional to determine the appropriate course of treatment.	

Q	Can you tell me other symptoms about the disease	Multiple Myeloma
A	Some other symptoms of <b>osteoporosis</b> include: * Bone pain and tenderness in the affected area * Muscle weakness and fatigue * Waddling gait and difficulty climbing stairs or standing up from a chair * Bone and muscle tenderness on pressure * Fractures of the ribs and pelvis * Osteomalacia in adults can present with fractures and low BMD, mimicking osteoporosis.	

Q	Can you tell me about the diagnosis procedure?	Multiple Myeloma
A	The diagnostic procedure for osteoporosis is to measure bone mineral density (BMD) using a bone density test such as a <b>X-ray</b> or a <b>computed tomography (CT) scan</b> . <b>The lower limit of normal for sex-matched young adults is used to establish the diagnosis</b> . The World Health Organization nomenclature uses the term osteopenia to refer to individuals whose BMD is between 1 and 2.5 standard deviations below peak bone mass and the term <b>osteoporosis to refer to individuals whose BMD is below 1.0 standard deviation below peak bone mass</b> .	



Q	Can you tell the treatment procedure?	Multiple Myeloma
A	<p>The treatment procedure for osteoporosis depends on its cause and the stage of the illness. If a secondary cause of osteoporosis is present, specific treatment should be aimed at correcting the underlying disorder. During the acute phase of vertebral compression, attention is directed toward relieving pain with analgesics, muscle relaxants, heat, massage, and/or rest. Many patients with discomfort related to osteoporotic fractures or deformity benefit from a well-designed program of physical therapy. Some patients appear to benefit from a corset or an orthopedic back brace. Both weight-bearing and non-weight-bearing exercises can help improve bone density and reduce the risk of fractures. In mildly affected, older women, estrogen treatment has been advocated, particularly 1403 to blunt bone resorption, but long-term efficacy is unknown. The only alternative to surgery at present is conservative medical follow-up. Most experts recommend surgery for all patients with symptomatic disease and even for asymptomatic patients meeting other, somewhat arbitrary, criteria, such as age younger than 70. In any case, regardless of the underlying cause, osteomalacia associated with adult acquired Fanconi's syndrome appears to respond well to treatment with phosphate.</p>	

Q	Can you tell me the possible causes of the disease?	Multiple Myeloma
A	<p>The possible causes of osteoporosis include age-related osteoporosis, which is caused by the natural decline in bone density with age, and osteoporosis associated with the inheritance of genetic variants that regulate bone density. Secondary causes of osteoporosis include idiopathic osteoporosis, which is caused by no specific underlying cause, and osteoporosis associated with a variety of diseases and drug treatments. Glucocorticoid-induced osteoporosis is also a possibility.</p>	

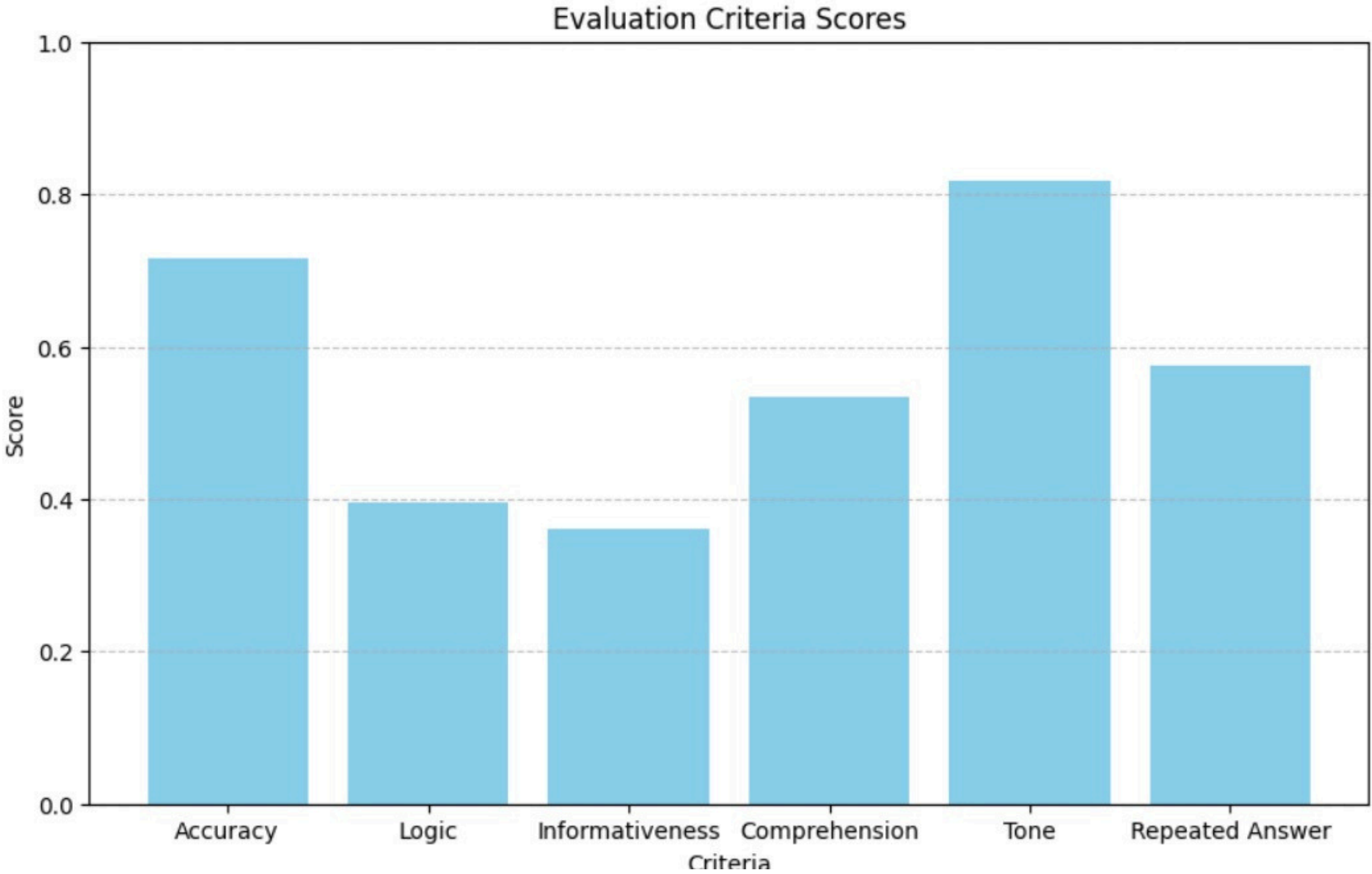


Q	My friend has been experiencing persistent coughing, sometimes with blood, along with fatigue and weight loss. Could it be something serious?	
A	<p>It's important to note that persistent coughing, sometimes with blood, along with fatigue and weight loss, can be a sign of a serious condition. It's important to seek medical attention as soon as possible to rule out any serious conditions. Coughing with blood in the lungs is a common symptom of pulmonary tuberculosis (TB), which is a type of lung infection caused by the bacteria Mycobacterium tuberculosis. The cough is typically dry and has a characteristic appearance, with a variable appearance of blood streaking or gross hemoptysis. The cough is often accompanied by fever, fatigue, and weight loss. Coughing with blood in the lungs is a common symptom of pulmonary tuberculosis (TB), which is a type of lung infection caused by the bacteria Mycobacterium tuberculosis. The cough is typically dry and has a characteristic appearance, with a variable appearance of blood streaking or gross hemoptysis. The cough is often accompanied by fever, fatigue, and weight loss. Coughing with blood in the lungs is a common symptom of pulmonary tuberculosis (TB), which is a type of lung infection caused by the bacteria Mycobacterium tuberculosis. The cough is typically dry and has a characteristic appearance, with a variable appearance of blood streaking</p>	Tuberculosis

# Human Evaluation Scores

Accuracy	Informativeness	Logic	Comprehension	Tone	Repeated answers
72%	40%	36%	53%	82%	58%

Questions are answered  
by the certified doctors



# Exam Dataset



Professional and Linguistic Assessments Board

The PLAB test evaluates overseas-qualified doctors' capability to work safely as a senior house officer in UK NHS hospitals.

General  
Medical  
Council

## PLAB 1 Applied Knowledge Test example questions

**1** A 55 year old man has diarrhoea and weight loss for 9 months. He has an intensely itchy red rash on his buttocks and thighs.

Investigations:

Haemoglobin            108 g/L            (130–175)

Blood film: anisocytosis

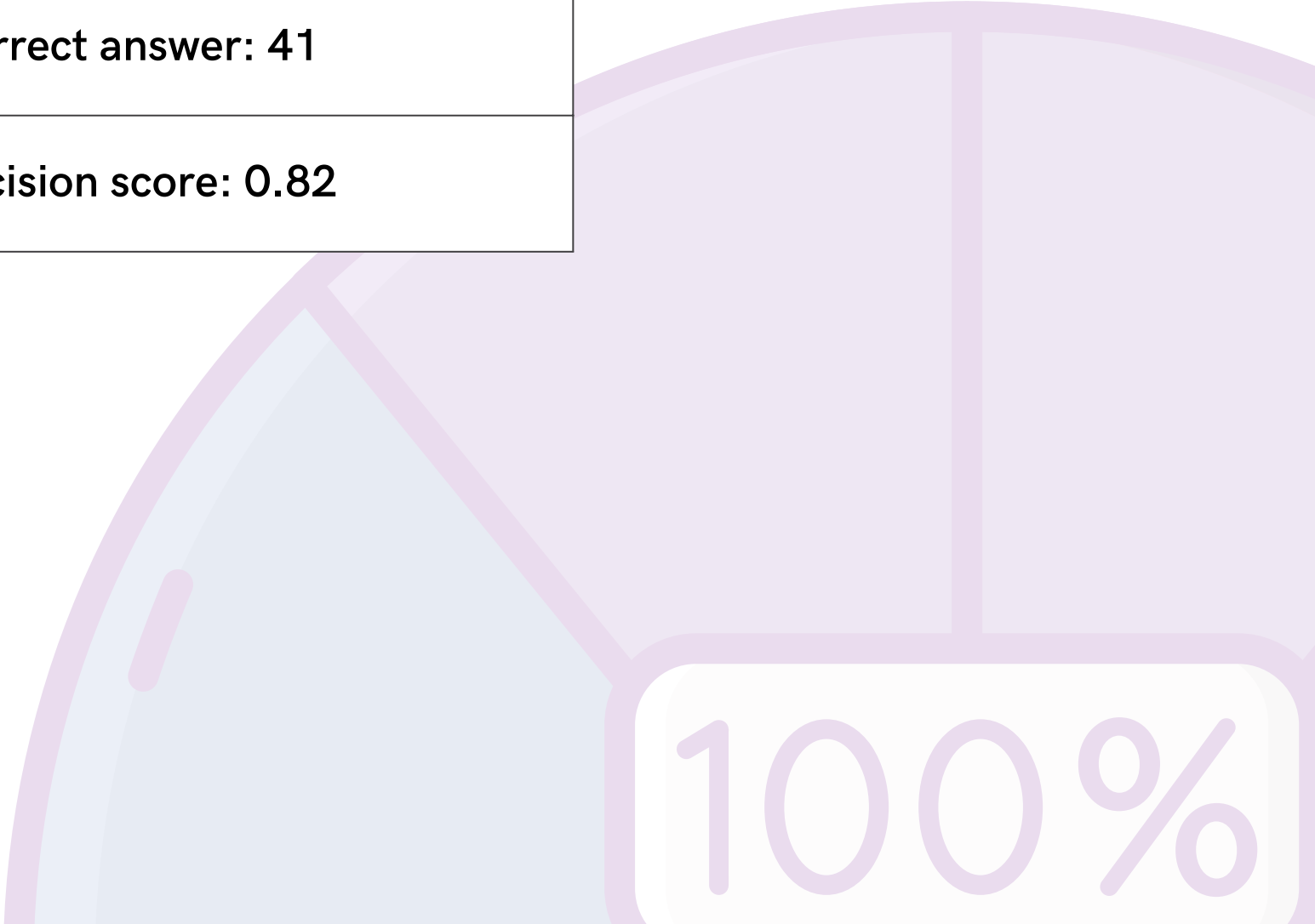
Which is the most likely histological appearance on small bowel biopsy?

- A    Hyperplasia of Brunner's glands
- B    Increased intraepithelial macrophages
- C    Increased lymphocytes in submucosa
- D    Ulceration
- E    Villous atrophy

<https://www.gmc-uk.org/registration-and-licensing/join-the-register/plab>

# Results

Our Model	Chatgpt 3.5
Correct answer: 14	Correct answer: 41
Precision score: 0.28	Precision score: 0.82



100%

Demo work

# Thank you for your time

Do you have any questions?