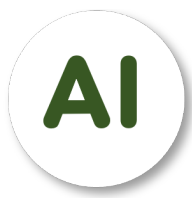


# NLP - Competition

# Vietnamese Medical QA

AI VIET NAM  
Nguyen Quoc Thai



# CONTENT

<b>1</b>	<b>Task Description</b>
<b>2</b>	<b>Retriever</b>
<b>3</b>	<b>Classifier - Generator</b>
<b>4</b>	<b>Improvement</b>
<b>5</b>	<b>Source Code</b>

# 1 – Task Description



I/O

- Task: Multiple Choices Question Answering
- Domain: Medical

## NOTE

**Corpus:** Ban tổ chức cung cấp dữ liệu về các loại bệnh phổ biến, mỗi bệnh có từ 1 đến 2 bài viết, bao gồm các thông tin liên quan đến nguyên nhân, triệu chứng, cách phòng tránh bệnh...

**Lưu ý:** Một lựa chọn được coi là ĐÚNG nếu xuất hiện thông tin tương ứng tại các bài viết trong corpus **được cung cấp bởi ban tổ chức**. Những lựa chọn không có thông tin trong corpus (mặc dù về mặt kiến thức có thể đúng) sẽ KHÔNG được coi là đáp án chính xác.

Question	Expected Answer
Đâu là triệu chứng của bệnh van tim? A. Khó thở B. Tăng cân nhanh chóng C. Vàng da D. Rụng tóc	1100
Tháng 8 vừa qua, An và Bình đi khám sức khỏe. An được chẩn đoán cận thị độ 3, Bình được chẩn đoán gan nhiễm mỡ. Làm thế nào để Bình hạn chế và giảm bớt bệnh của mình? A. Tăng cường tiêu thụ cồn B. Ăn nhiều thực phẩm chứa cholesterol C. Giảm cân, tập thể dục đều đặn và duy trì chế độ ăn uống lành mạnh D. Hút thuốc lá	0010

# 1 – Task Description



## I/O

- Task: Multiple Choices Question Answering
- Domain: Medical

## LIMITATION

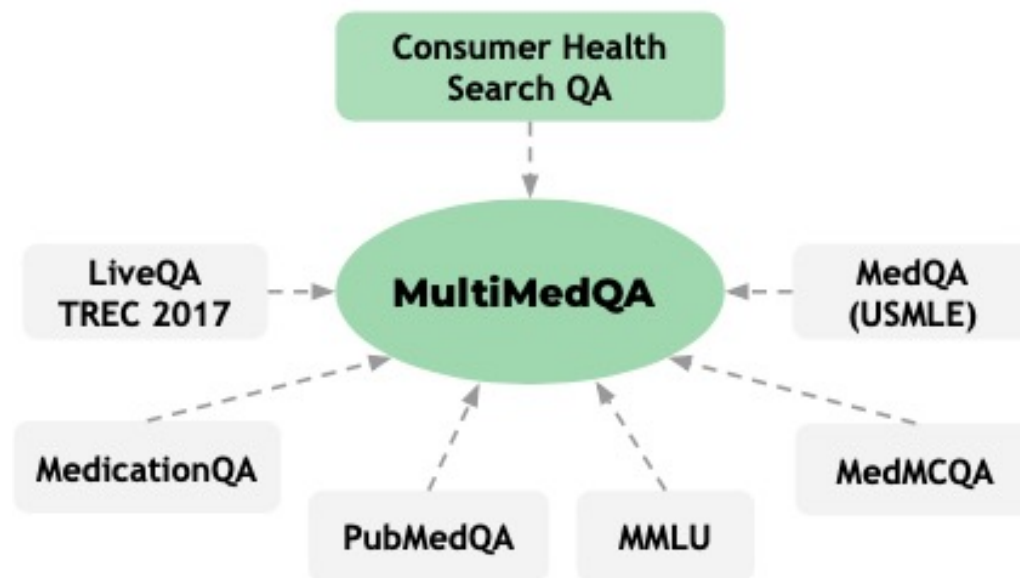
- Các đội được phép sử dụng các **pre-trained model và dataset được công bố public**. Việc sử dụng pre-trained model cần được thông báo với ban tổ chức. Các pre-trained model và dataset private không được phép sử dụng
- Thời gian inference mỗi test case không quá 10 giây
- Môi trường inference không có internet
- Cấu hình máy inference:
  - GPU: NVIDIA Tesla T4 16GB GDDR6 PCIe 3.0
  - CPU: Intel SkyLake 6151 3.0GHz
  - RAM: 32GB

# 1 – Task Description



## MultiMedQA Dataset

### Medical Question Answering Benchmark



# 1 – Task Description



## MultiMedQA Dataset

Dataset	Format	Size (dev/test)	Domain
MedQA (USMLE)	Q + A (4-5 Choices)	11450 / 1273	General medical knowledge in US medical licensing exam
MedMCQA (AIIMS/NEET)	Q + A (4 Choices and Explanations)	187K / 6.1K	General medical knowledge in Indian medical entrance exams
PubMedQA	Q + Context + A (Yes/No/Maybe) (Long Answer)	500 / 500 #QA pairs: Labeled: 1k Unlabeled: 61.2k Synthetic: 211.3k	Biomedical scientific literature
MMLU	Q + A (4 Choices)	123 / 1089	Medical knowledge covering anatomy, clinical knowledge, college medicine, medical genetics, professional medicine, and college biology
LiveQA TREC-2017	Q + Long Answer (Librarian Answers)	634 / 104	General medical knowledge sought by consumers
Medication QA	Q + A (Long Answer)	NA / 674	Medication knowledge frequently sought by consumers
HealthSearchQA (Ours)	Q + Manual Expert Evaluation	3375	General medical knowledge searched for by consumers

# 1 – Task Description



## MultiMedQA Dataset

MedQA (2021) [33]

**Format:** Q + A, multiple choice, open domain

**Size (Dev/Test):** 11450 / 1273

**Question:** A 65-year-old man with hypertension comes to the physician for a routine health maintenance examination. Current medications include atenolol, lisinopril, and atorvastatin. His pulse is 86/min, respirations are 18/min, and blood pressure is 145/95 mm Hg. Cardiac examination reveals end diastolic murmur. Which of the following is the most likely cause of this physical examination?

**Answer:** (A) Decreased compliance of the left ventricle (B) Myxomatous degeneration of the mitral valve (C) Inflammation of the pericardium (D) Dilation of the aortic root (E) Thickening of the mitral valve leaflets

# 1 – Task Description



## MultiMedQA Dataset

MedMCQA (2022) [64]

**Format:** Q + A, multiple choice, open domain

**Size (Dev/Test):** 187000 / 6100

**Question:** Which of the following ultrasound findings has the highest association with aneuploidy?

**Answer:** (A) Choroid plexus cyst (B) Nuchal translucency (C) **Cystic hygroma** (D) Single umbilical artery

**Explanation:** All the above mentioned are ultrasound findings associated with increased risk of aneuploidy although the highest association is seen with cystic hygroma. Nuchal translucency and cystic hygroma are both measured in the first trimester. Trisomy 21 is the most common aneuploidy associated with increased NT and cystic hygroma while monosomy X presents as second-trimester hygroma.



# 1 – Task Description



## MultiMedQA Dataset

PubMedQA (2019) [34]

**Format:** Q + A + context, multiple choice, closed domain

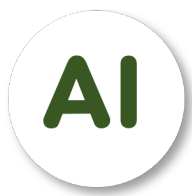
**Size (Dev/Test):** 500 / 500

**Question:** Double balloon enteroscopy: is it efficacious and safe in a community setting?

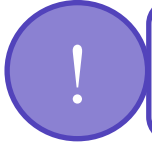
**Context:** From March 2007 to January 2011, 88 DBE procedures were performed on 66 patients. Indications included evaluation anemia/gastrointestinal bleed, small bowel IBD and dilation of strictures. Video-capsule endoscopy (VCE) was used prior to DBE in 43 of the 66 patients prior to DBE evaluation. The mean age was 62 years. Thirty-two patients were female, 15 were African-American; 44 antegrade and 44 retrograde DBEs were performed. The mean time per antegrade DBE was  $107.4 \pm 30.0$  minutes with a distance of  $318.4 \pm 152.9$  cm reached past the pylorus. The mean time per lower DBE was  $100.7 \pm 27.3$  minutes with  $168.9 \pm 109.1$  cm meters past the ileocecal valve reached. Endoscopic therapy in the form of electrocautery to ablate bleeding sources was performed in 20 patients (30.3%), biopsy in 17 patients (25.8%) and dilation of Crohn's-related small bowel strictures in 4 (6.1%). 43 VCEs with pathology noted were performed prior to DBE, with findings endoscopically confirmed in 32 cases (74.4%). In 3 cases the DBE showed findings not noted on VCE.

**Answer:** Yes

**Long Answer:** DBE appears to be equally safe and effective when performed in the community setting as compared to a tertiary referral center with a comparable yield, efficacy, and complication rate.



# 1 – Task Description



## MultiMedQA Dataset

MMLU (2020) [29]

**Format:** Q + A, multiple choice, open domain

Anatomy

**Size (Dev/Test):** 14 / 135

**Question:** Which of the following controls body temperature, sleep, and appetite?

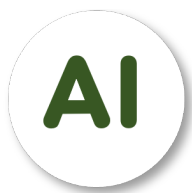
**Answer:** (A) Adrenal glands (B) **Hypothalamus** (C) Pancreas (D) Thalamus

Clinical Knowledge

**Size (Dev/Test):** 29 / 265

**Question:** The following are features of Alzheimer's disease except:

**Answer:** (A) short-term memory loss. (B) confusion. (C) poor attention. (D) **drowsiness.**



# 1 – Task Description



## MultiMedQA Dataset

LiveQA (2017) [1]

---

**Format:** Q + long answers, free text response, open domain

**Size (Dev/Test):** 634/104

**Question:** Could second hand smoke contribute to or cause early AMD?

**Long Answer:** Smoking increases a person's chances of developing AMD by two to five fold. Because the retina has a high rate of oxygen consumption, anything that affects oxygen delivery to the retina may affect vision. Smoking causes oxidative damage, which may contribute to the development and progression of this disease. Learn more about why smoking damages the retina, and explore a number of steps you can take to protect your vision.

# 1 – Task Description



## MultiMedQA Dataset

**HealthSearchQA** We curated our own additional dataset consisting of 3375 commonly searched consumer questions, referred to as “HealthSearchQA”. The dataset was curated using seed medical conditions and their associated symptoms. We used the seed data to retrieve publicly-available commonly searched questions generated by a search engine, which were displayed to all users entering the seed terms. We publish the dataset as an open benchmark for consumer medical question answering and hope this will be a useful resource for the community, as a dataset reflecting real-world consumer concerns.

HealthSearchQA (Ours)

**Format:** Q only, free text response, open domain

**Size:** 3375

**Question:** How serious is atrial fibrillation?

**Question:** What kind of cough comes with Covid?

**Question:** Is blood in phlegm serious?

# 1 – Task Description



## MultiMedQA Dataset

**MedicationQA** The MedicationQA dataset [2] consists of commonly asked consumer questions about medications. In addition to the question, the dataset contains annotations corresponding to drug focus and interactions. Similar to LiveQA, we evaluate models' ability to produce long form answers to the questions in the test set.

MedicationQA (2017) [2]

**Format:** Q + long answers, free text response, open domain

**Size (Dev/Test):** NA/674

**Question:** *Question:* how does valium affect the brain?

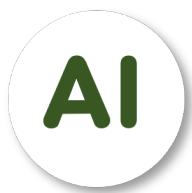
**Focus (Drug):** Valium

**Question Type:** Action

**Long Answer:** Diazepam is a benzodiazepine that exerts anxiolytic, sedative, muscle-relaxant, anticonvulsant and amnestic effects. Most of these effects are thought to result from a facilitation of the action of gamma aminobutyric acid (GABA), an inhibitory neurotransmitter in the central nervous system.

**Section Title:** CLINICAL PHARMACOLOGY

**URL:** <https://dailymed.nlm.nih.gov/dailymed/drugInfo.cfm?setid=554baee5-b171-4452-a50a-41a0946f956c>



# 1 – Task Description

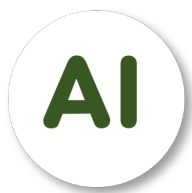
## FrenchMedMCQA Dataset

# Answers	Training	Validation	Test	Total
1	595	164	321	1,080
2	528	45	97	670
3	718	71	141	930
4	296	30	56	382
5	34	2	7	43
Total	2171	312	622	3,105

Table 1: FrenchMedMCQA dataset distribution.

```
{
  "id": "6979d46501a3270436d37b98cf351439fbcbec8d5890d293dabfb8f85f723904",
  "question": "Cocher la (les) proposition(s) exacte(s) : Le métronidazole :",
  "answers": {
    "a": "Est un dérivé du pyrazole",
    "b": "Peut induire un effet antabuse",
    "c": "Peut être administré par voie parentérale intraveineuse",
    "d": "Peut être utilisé dans certaines parasitoses à protozoaires",
    "e": "Est inefficace dans les infections à germes anaérobies"
  },
  "correct_answers": [
    "b",
    "c",
    "d"
  ],
}
```

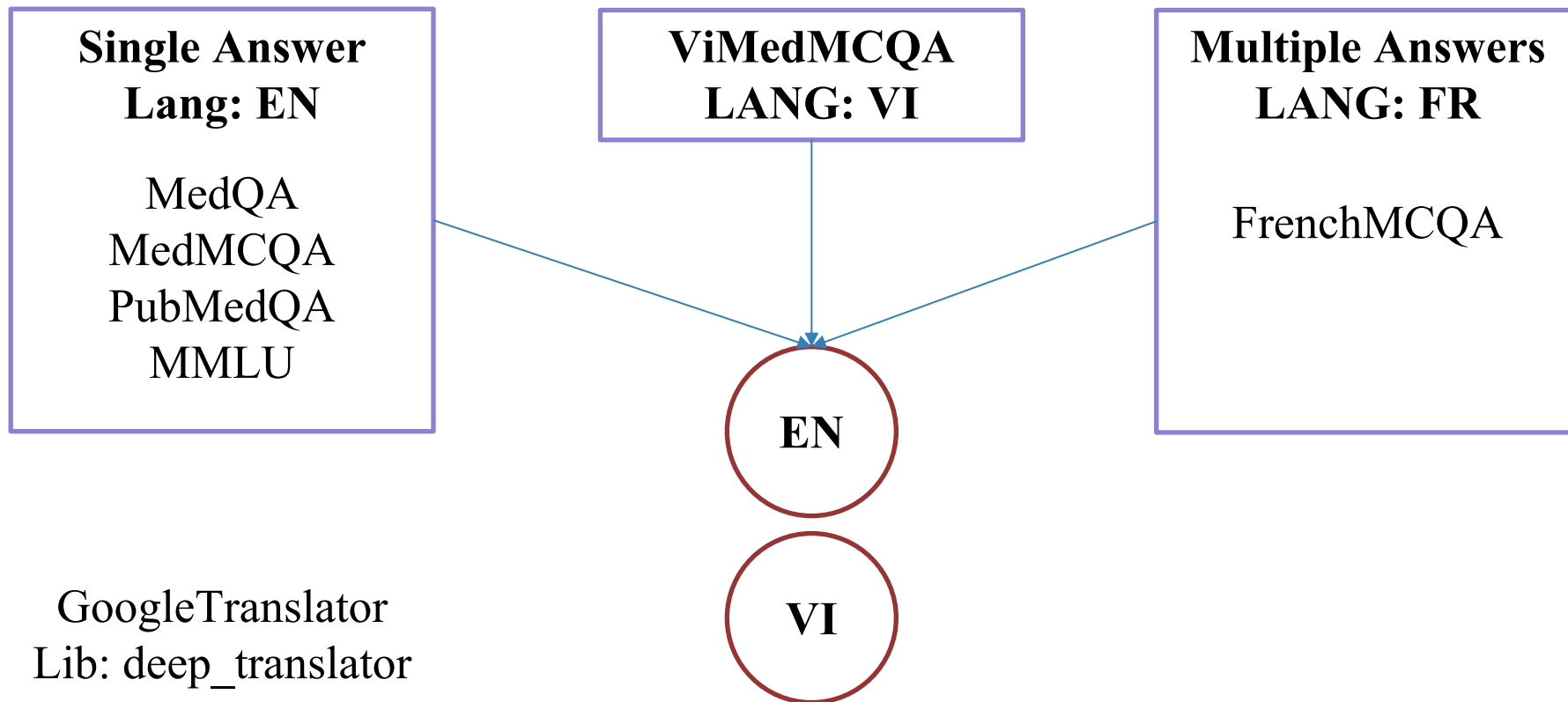




# 1 – Task Description



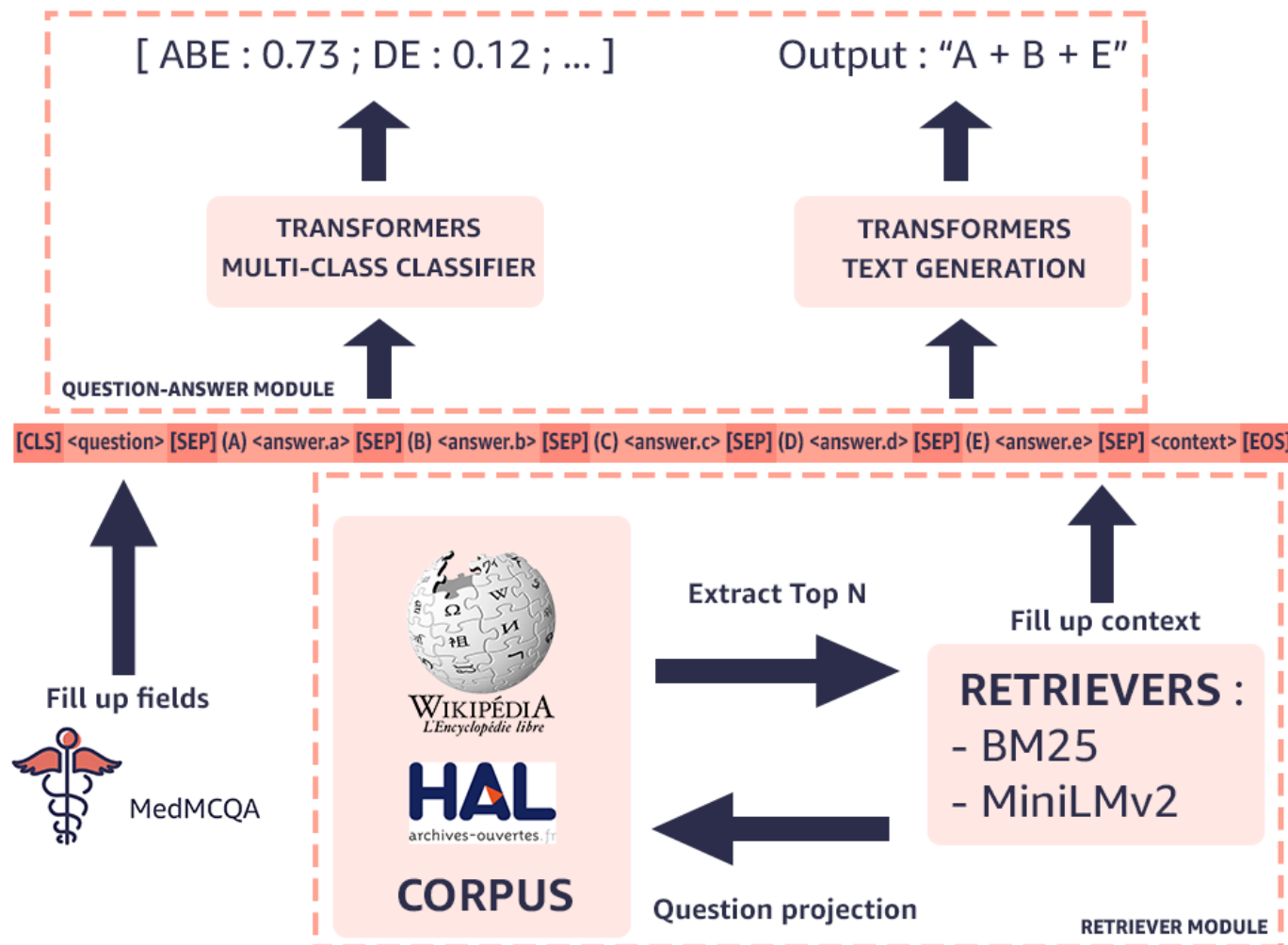
## MCQA Dataset



# 1 – Task Description

## Stage 2: QA Module Classifier - Generator

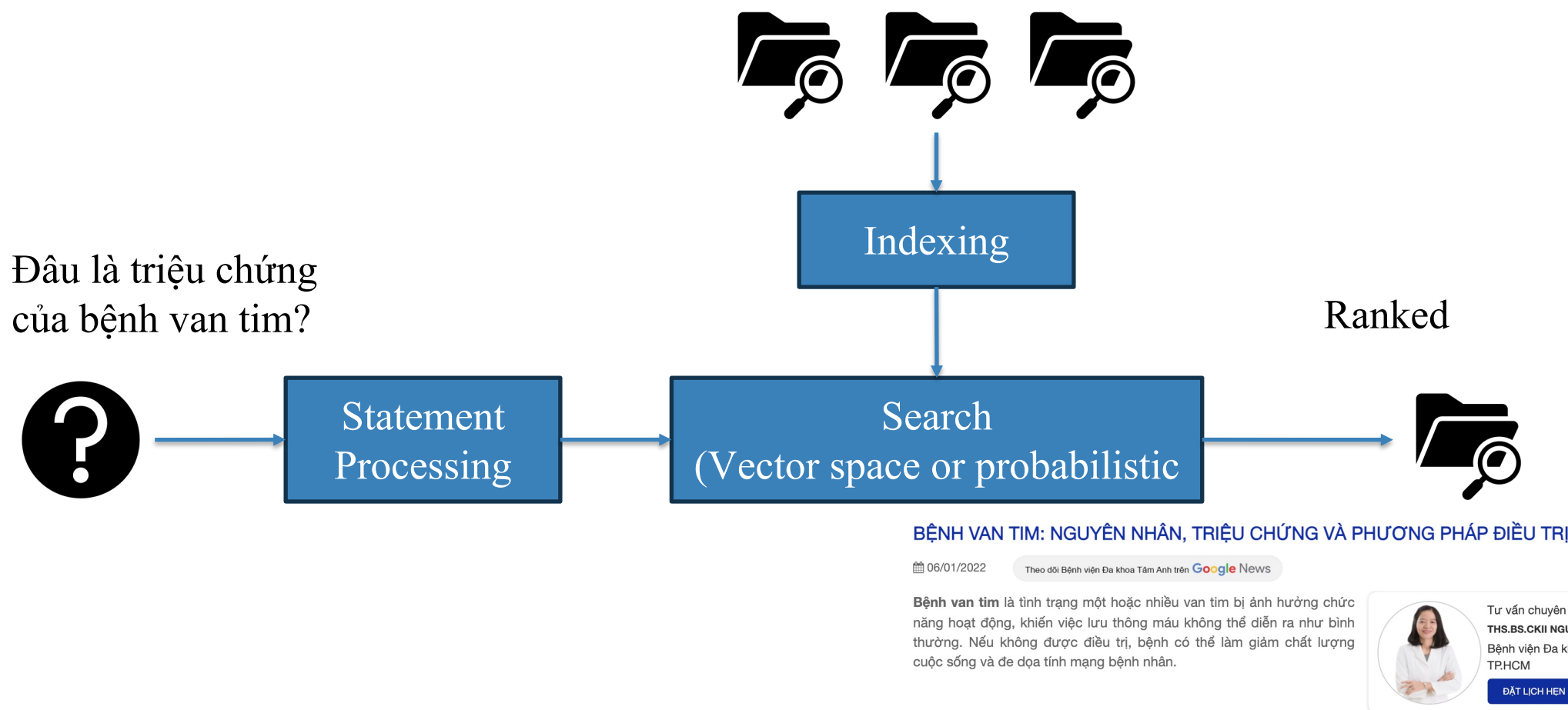
## Stage 1: Retriever







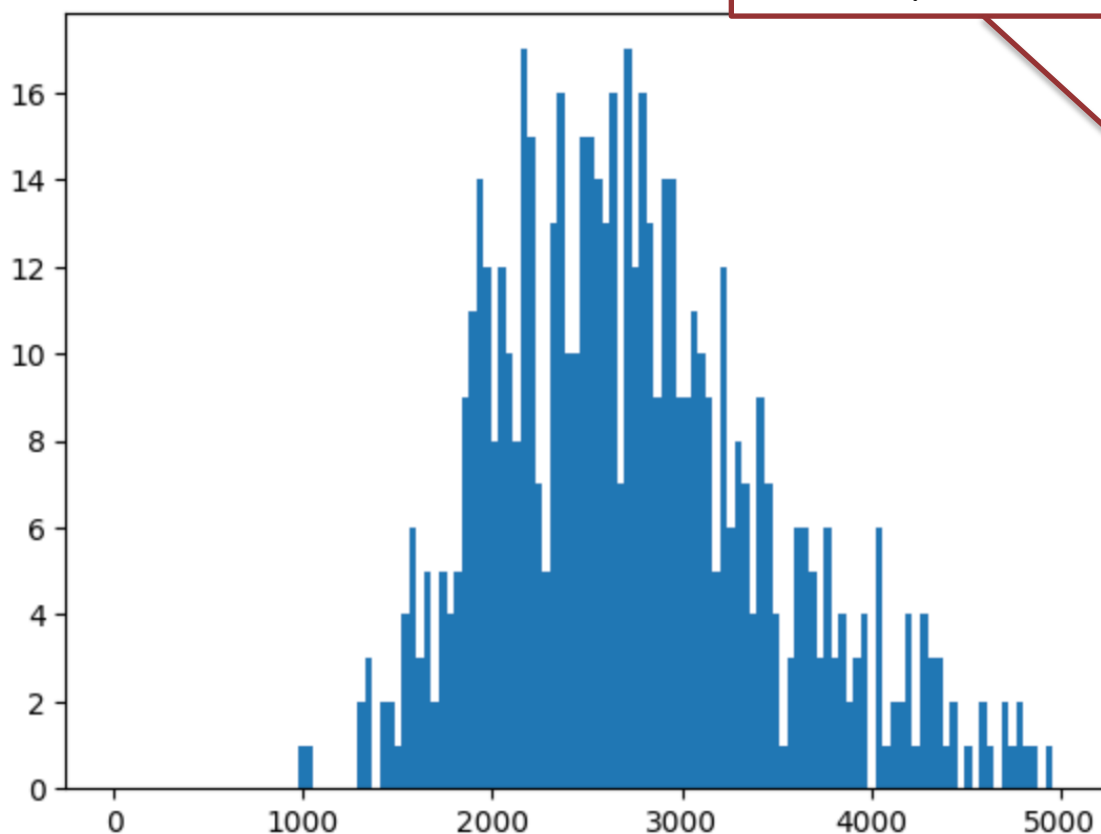
## Document – Passage Retriever





## EDA

## ➤ Document Length



Đây là triệu chứng  
của bệnh van tim?

## Mục lục

- Bệnh van tim là gì?
- Các loại bệnh van tim thường gặp
- Nguyên nhân gây bệnh van tim
- Yếu tố nguy cơ của bệnh
- Các triệu chứng của bệnh van tim
- Biến chứng của bệnh van tim
- Các bệnh lý van tim được chẩn đoán như thế nào?
- ✓ • Bệnh van tim được điều trị như thế nào?
  - 1. Dùng thuốc
  - 2. Thủ thuật nong van tim
  - 3. Phẫu thuật sửa van tim
  - 4. Phẫu thuật thay van tim
- Lưu ý đối với người bệnh van tim
- ✓ • Lời khuyên dành cho người bệnh van tim



## Document – Passage Retriever

### ➤ Document Length

Đâu là triệu chứng  
của bệnh van tim?

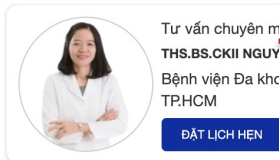
Document Retriever: BM25

#### BỆNH VAN TIM: NGUYÊN NHÂN, TRIỆU CHỨNG VÀ PHƯƠNG PHÁP ĐIỀU TRỊ

06/01/2022

Theo dõi Bệnh viện Đa khoa Tâm Anh trên Google News

**Bệnh van tim** là tình trạng một hoặc nhiều van tim bị ảnh hưởng chức năng hoạt động, khiến việc lưu thông máu không thể diễn ra như bình thường. Nếu không được điều trị, bệnh có thể làm giảm chất lượng cuộc sống và đe dọa tính mạng bệnh nhân.



Passage Retriever

DPR: Bi-Encoder

#### Mục lục

- Bệnh van tim là gì?
- Các loại bệnh van tim thường gặp
- Nguyên nhân gây bệnh van tim
- Yếu tố nguy cơ của bệnh
- Các triệu chứng của bệnh van tim
- Biến chứng của bệnh van tim
- Các bệnh lý van tim được chẩn đoán như thế nào?
- Bệnh van tim được điều trị như thế nào?



## BM25

➤ rank\_bm25

Common words  
less important

Repetitions of query  
words => good

$$\text{BM25}(D, Q) = \sum_{i=1}^n \text{IDF}(q_i, D) \frac{f(q_i, D) \cdot (k_1 + 1)}{f(q_i) + k_1 \cdot (1 - b + b \cdot |D|/d_{avg})}$$

More words in common with  
the query => good

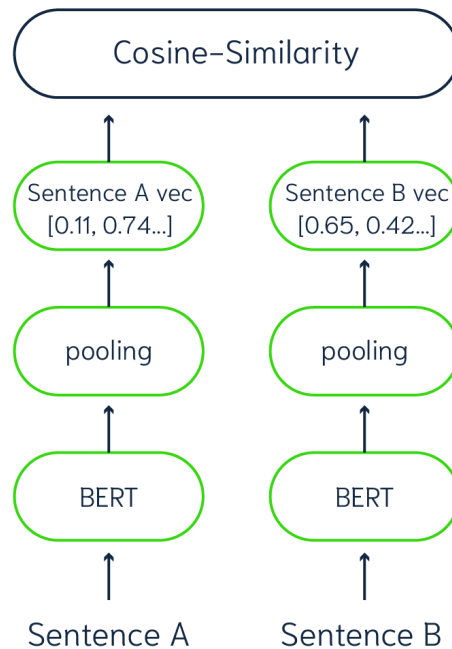
Repetitions less important  
than different query words

But more important if document  
is relatively long (average)

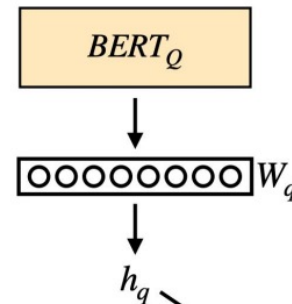


## DPR: Bi-Encoder

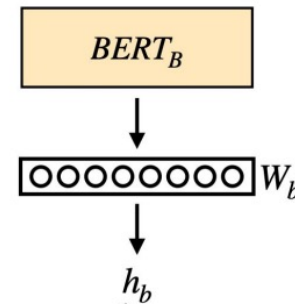
## ➤ Bi-Encoder



Question  $q$   
What does the zip  
in zip code stand for?

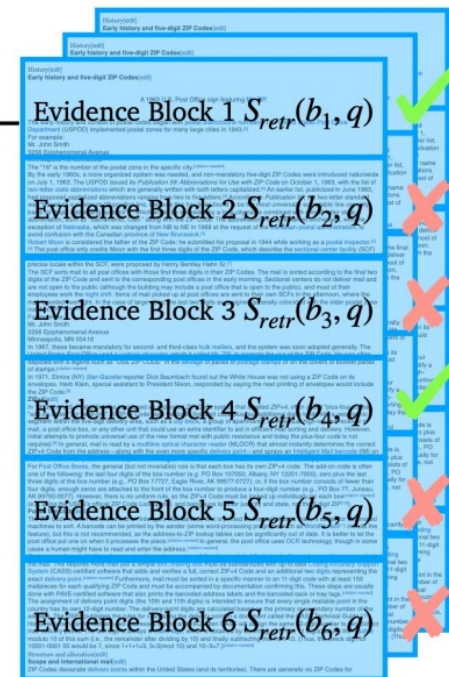


Each evidence block  $b$



$$S_{retr}(b, q) = h_q^T h_b$$

All of Wikipedia: select top K



# 3 – Classifier - Generator



## QA Module

### QUESTION

Đâu là triệu chứng của bệnh van tim

### CHOICES

- A. Khó thở
- B. Tăng cân
- C. Vàng da
- D. Rụng tóc

### CONTEXT

Các triệu chứng của bệnh van tim

- Khó thở
- Suy nhược hoặc chóng mặt
- Khó chịu ở ngực
- Đánh trống ngực
- Sung mắt cá chân, bàn chân hoặc bụng (chứng phù nề).
- Tăng cân nhanh chóng

Model

### ANSWER

1100  
A B

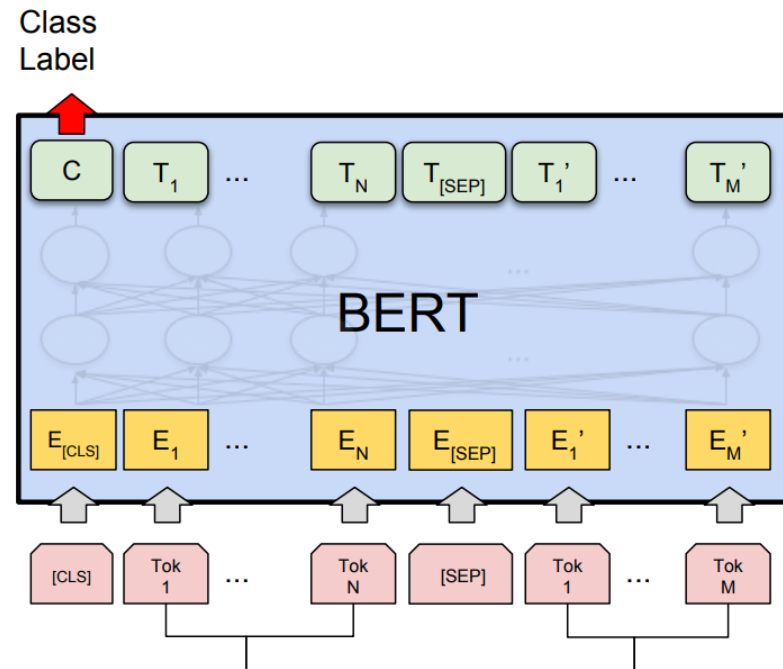
# 3 – Classifier - Generator



## Classifier

### ➤ Multi-Class Text Classification

Label: {‘A’:0, ‘B’:1,..., ‘A B’: 3,... ‘A B C’: 5}



Input: “Context”: context. “Question”. “Choices”: A – B – C – D - E

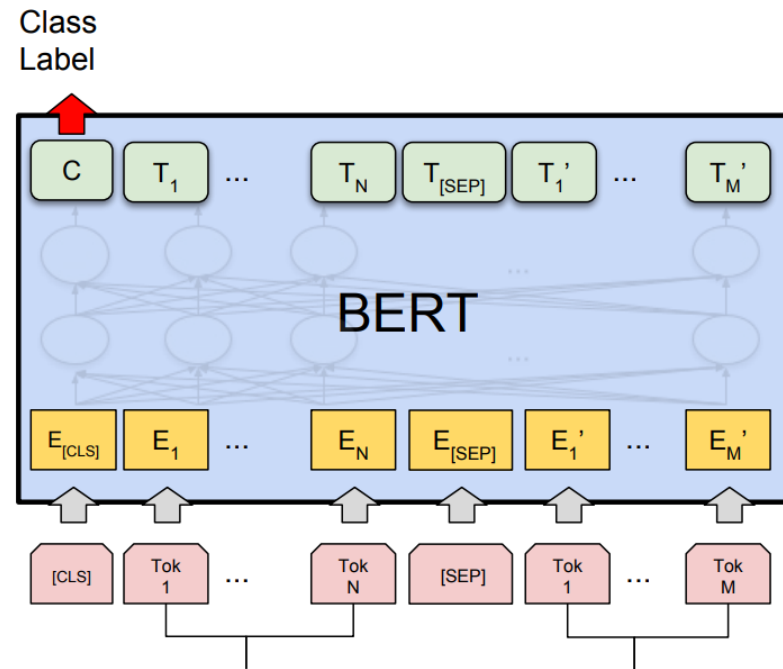
# 3 – Classifier - Generator



## Classifier

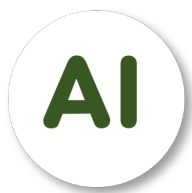
### ➤ Multi-Label Text Classification

Label-One hot Encoding (A: [1, 0, 0, 0 0],... 'A B'

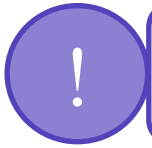


Input: “Context”: context. “Question”. “Choices”: A – B – C – D - E





## 3 – Classifier - Generator



### Classifier

#### ➤ Classification Model

Label-One hot Encoding (A: [1, 0, 0, 0 0],... 'A B'

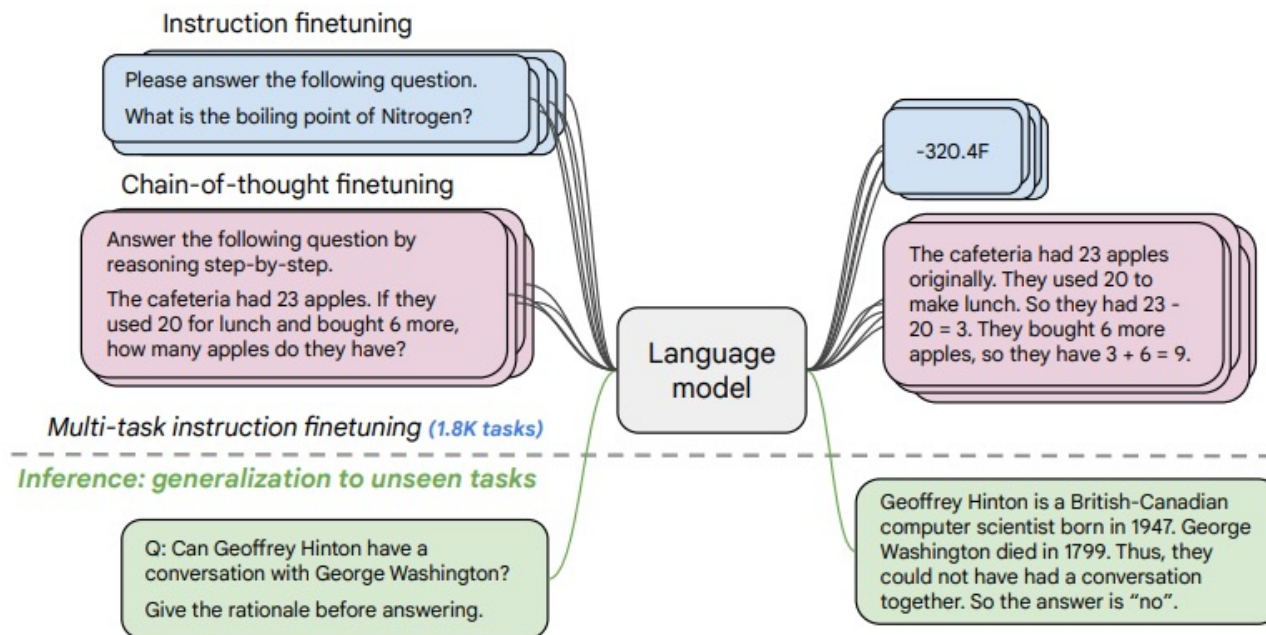
	Without Context		Wiki w/ BM25		HAL w/ BM25		Wiki w/ MiniLMv2		HAL w/ MiniLMv2	
Architecture	Hamming	EMR	Hamming	EMR	Hamming	EMR	Hamming	EMR	Hamming	EMR
BioBERT V1.1	36.19	15.43	<b>38.72</b>	16.72	33.33	14.14	35.13	16.23	34.27	13.98
PubMedBERT	33.98	14.14	34.00	13.98	35.66	15.59	33.87	14.79	35.44	14.79
CamemBERT-base	36.24	16.55	34.19	14.46	34.78	15.43	34.66	14.79	34.61	14.95
XLM-RoBERTa-base	37.92	17.20	31.26	11.89	35.84	16.07	32.47	14.63	33.00	14.95
BART-base	31.93	15.91	34.98	<b>18.64</b>	33.80	17.68	29.65	12.86	34.65	18.32

Input: “Context”: context. “Question”. “Choices”: A – B – C – D - E

# 3 – Classifier - Generator



## Generator



Input: “Context”: context. “Question”.  
“Choices”: A – B – C – D - E

Output: “A B”  
“A. Khó thở và B. Tăng cân”

## 4 – Improvement



## Difficulty of Question

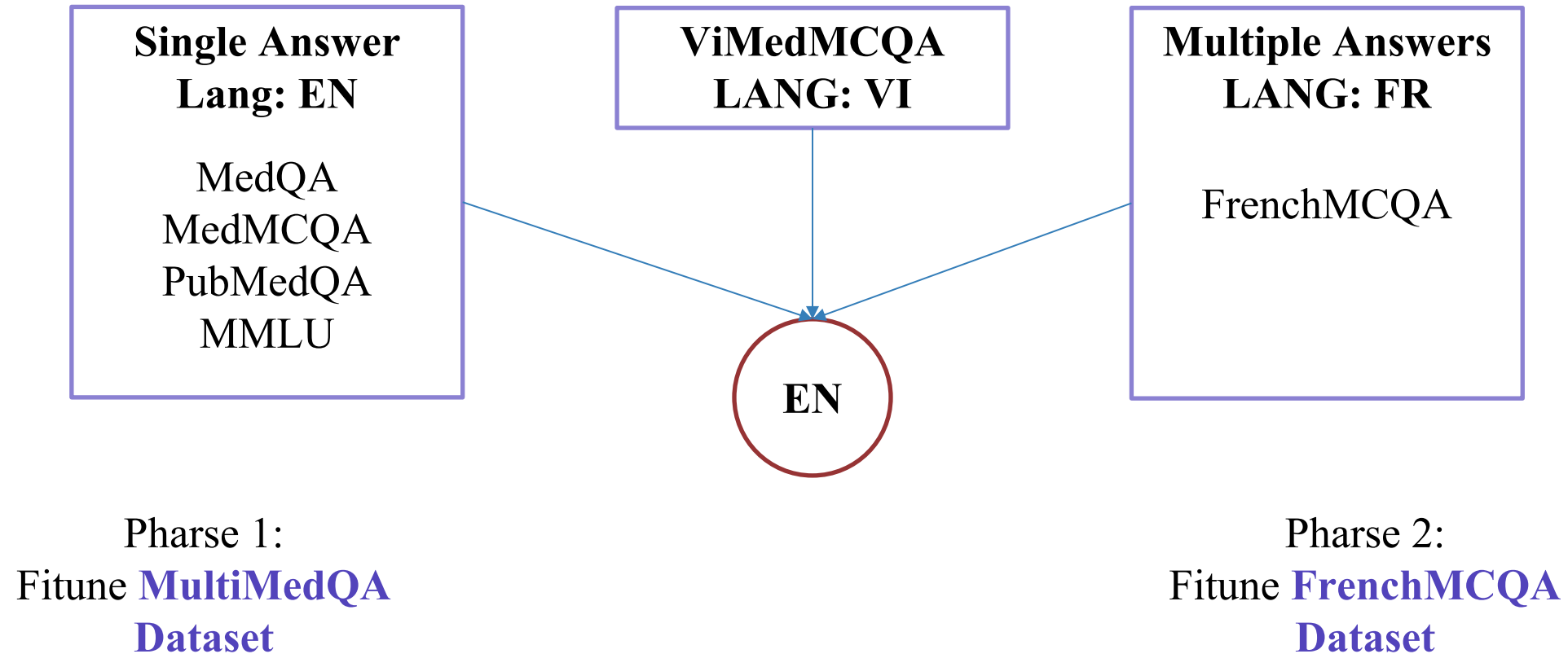
Độ khó	Số bài viết liên quan	Mức độ ngữ nghĩa	Ví dụ
1	1	Đơn giản, câu hỏi dạng hỏi trực tiếp vào vấn đề	<p>Đâu là triệu chứng của bệnh van tim?</p> <p>A. Khó thở B. Tăng cân nhanh chóng C. Vàng da D. Rụng tóc</p>
2	$\geq 2$	Đơn giản, câu hỏi dạng hỏi trực tiếp vào vấn đề	<p>Viêm tai ngoài và mất thính lực một bên tai đều có thể được điều trị bằng những loại thuốc nào?</p> <p>A. Thuốc kháng sinh hoặc corticosteroid B. Thuốc giảm đau và thuốc ngủ C. Thuốc hạ huyết áp và thuốc kháng histamin</p>
3	1	Phức tạp hơn, có thể lồng ghép những thông tin gây nhiễu	<p>Tháng 8 vừa qua, An và Bình đi khám sức khỏe. An được chẩn đoán cận thị độ 3, Bình được chẩn đoán gan nhiễm mỡ. Làm thế nào để Bình hạn chế và giảm bớt bệnh của mình?</p> <p>A. Tăng cường tiêu thụ cồn B. Ăn nhiều thực phẩm chứa cholesterol C. Giảm cân, tập thể dục đều đặn và duy trì chế độ ăn uống lành mạnh D. Hút thuốc lá</p>
4	$\geq 2$	Phức tạp hơn, có thể lồng ghép những thông tin gây nhiễu	<p>Tiểu Đường là nữ ca sĩ nổi tiếng ở Trung Quốc. Thời gian gần đây giọng hát của cô đi xuống trầm trọng do những triệu chứng khó chịu như đau họng, rát họng và khó nuốt. Tiểu Đường có thể đã mắc căn bệnh gì?</p> <p>A. Viêm họng B. Tiểu đường C. Ung thư vòm họng</p>

Clause Detection  
NER

# 4 – Improvement

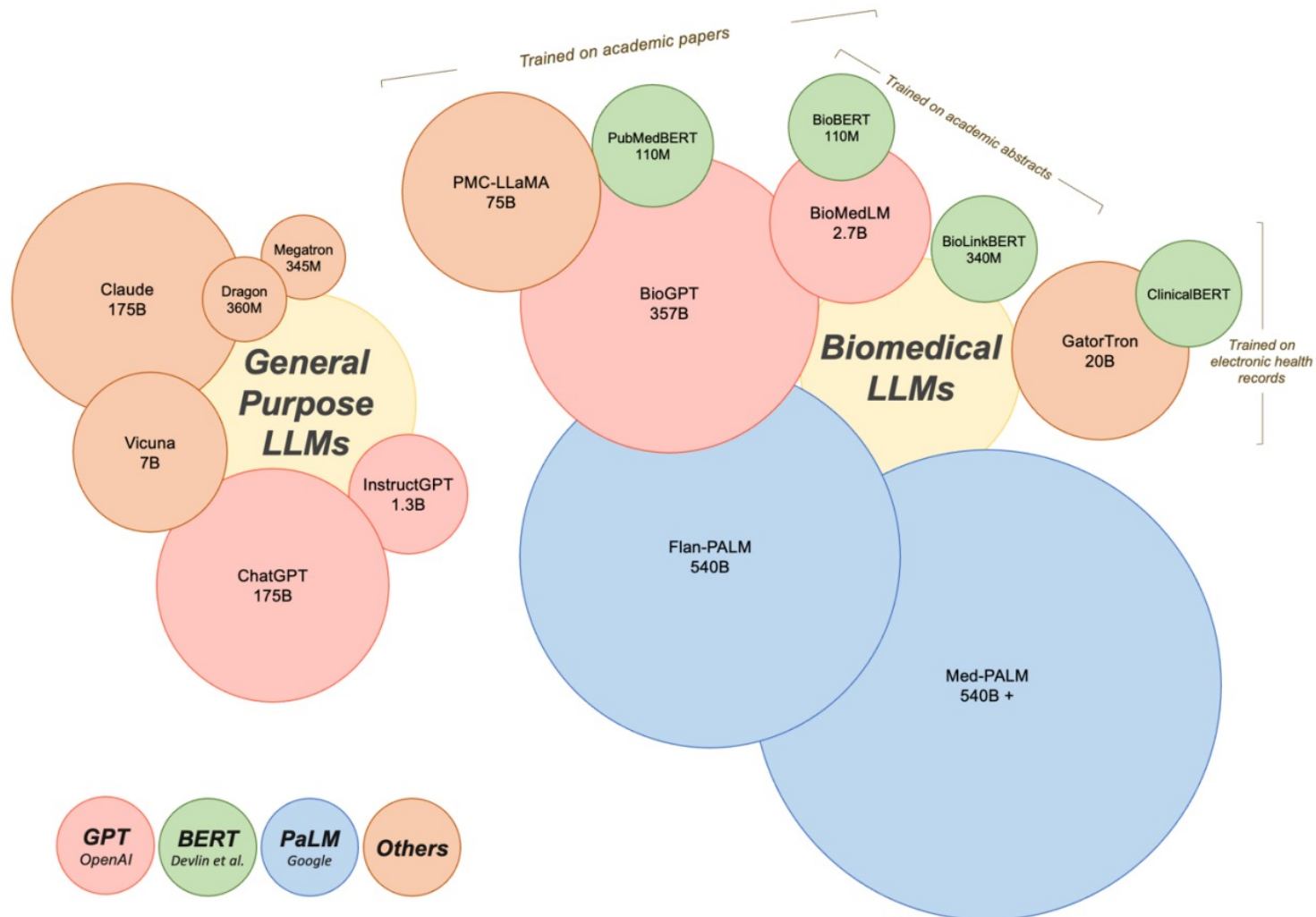


## MultiMedQA Dataset





## Prompting Techniques



# 4 – Improvement



## Prompting Techniques

**You are a helpful medical knowledge assistant. Provide useful, complete, and scientifically-grounded answers to common consumer search queries about health.**

**Question:** How do you treat skin redness?

**Complete Answer:** It depends on the cause of the skin redness. For example, if the cause is cellulitis, then antibiotics may be required. However, this might be inappropriate for other causes of redness such as eczema. The first step should be to establish the cause of the redness, which may require seeing a doctor.

**You are a helpful medical knowledge assistant. Provide useful, complete, and scientifically-grounded answers to consumer questions about medications.**

**Question:** Why take simvastatin?

**Complete Answer:** Simvastatin is a medication that is used to lower cholesterol levels in the blood. Simvastatin is typically used to treat high cholesterol and to prevent heart disease, and to reduce the risk of heart attacks and strokes in those with the greatest risk of these happening (or re-occurring). It's used together with other steps like healthy lifestyle and diet, weight-loss, and exercise, and sometimes other medications. Simvastatin may also be used to decrease the amount of...

**You are a helpful medical knowledge assistant. Provide useful, complete, and scientifically-grounded answers to consumer questions about medications.**

**Question:** How can i split a 3mg or 4mg waefin pill so i can get a 2.5mg pill?

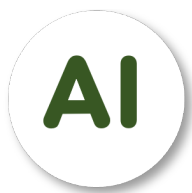
**Complete Answer:** You should not attempt to split a 3mg or 4mg warfarin pill to obtain a 2.5mg dose. If you are prescribed warfarin in a 2.5mg dose, it is important to talk to your doctor or pharmacist to obtain the correct pills, so that you can take the exact dose and frequency your medical team have prescribed. Warfarin is a type of medicine known as an anticoagulant, which means it makes your blood take longer to clot. The exact amount required to achieve the desired effect can vary.

→  
Instruction  
Prompt Tuning

Flan-PaLM

Prompt parameters  
aligned with the  
medical domain

Med-PaLM



AI VIET NAM

@aivietnam.edu.vn

## 5 – Source Code



Code



AI VIET NAM

@aivietnam.edu.vn

# Thanks!

## Any questions?