

‘한국형 의료 모델 만들기’ 선행연구

SAI NLP팀

의료 AI

의료 문진을 위한 거대 언어 모델(LLM) 활용 방안 탐색

비용 효율적인 데이터 수집, 모델 학습 방안 연구


소아청소년과 복통 관련 50개 질병

Meta LLaMA 2 이용

LLAMA2란? : Meta사에서 출시한 대규모 언어모델

 **Meta**
LLaMA 2

- 파라미터 70억 or 130억 개
- 오픈소스
- 상용 PC 구동 가능

 **OpenAI**
GPT 4

- 파라미터 1.7 조 개(추정)
- 이용량에 비례한 과금 구조
- openAI의 서비스 의존적

chatGPT 이용 학습 데이터 생성

USMLE (미국 의사국가고시) 형식 학습 데이터 생성

소아청소년과 복통 관련 50개 질병에 대한 가상
환자 시나리오 생성

LLM을 이용한 데이터 생성 (a.k.a. 'Textbook')
기법 활용



ChatGPT

Case 1: Acute Appendicitis

A 13-year-old male presents with severe pain in the lower right quadrant of his abdomen. The pain started mildly around the navel but worsened over 24 hours. He mentions feeling nauseous and had a low-grade fever last night. Upon examination, he exhibits tenderness in the McBurney's point and a slight increase in heart rate.

Diagnosis: Acute appendicitis. The classic migration of pain and local tenderness, combined with systemic symptoms, strongly suggest appendicitis.

Case 2: Gastroenteritis

A 7-year-old female complains of abdominal pain, especially around the belly button, for two days. She also reports watery diarrhea and occasional vomiting. She recently attended a



Message





가상 환자 시나리오




- 소아청소년과 보통 50개 질병에 대한 가상 환자 시나리오
- 질병 당 약 100건
- 전체 4,970건

Dataset Viewer Auto-converted to Parquet </> API Go to dataset viewer

Split

train (4.97k rows)

Search this dataset

question string · lengths	answer string · lengths	explanation string · lengths
		
120 780	3 99	20 544
A 12 yearold girl presented with a 6 month history of abdominal pain...	the most likely diagnosis is Irritable Bowel Syndrome IBS	Given the patients chronic abdominal pain change in bowel...
An 8 yearold boy complained of recurrent abdominal pain over the...	Irritable Bowel Syndrome IBS	The boys recurrent abdominal pain stressrelated exacerbations and...
A 15 yearold girl came in with a 4 month history of intermittent...	Irritable Bowel Syndrome IBS	The patients chronic abdominal pain change in bowel habits and family...
A 10 yearold boy presented complaining of abdominal pain that...	Irritable Bowel Syndrome IBS	The patients symptoms family history and physical findings...
A 17 yearold girl presented with a 2 year history of recurring...	Irritable Bowel Syndrome IBS	The patients chronic abdominal pain change in bowel habits and family...
A 13 yearold boy came in with a history of intermittent abdominal...	Irritable Bowel Syndrome IBS	The boys chronic abdominal pain altered bowel habits and family...

< Previous 1 2 3 ... 50 Next >



가상 환자 시나리오 예시

증상

A 14 yearold boy presented with a 1 year history of recurring abdominal pain which he described as sharp and sometimes crampy. He also complained of frequent loose stools and occasional constipation. He mentioned I feel better after going to the bathroom. His mother was diagnosed with IBS Physical examination showed mild lower abdominal tenderness.

14세 남자아이가 날카롭고 때때로 경련이 일어나는 듯한 복통이 1년째 반복되고 있다고 호소했습니다. 또한 잦은 묽은 변과 가끔 변비가 있다고 호소했습니다. 그는 화장실에 다녀오면 기분이 나아진다고 말했습니다. 그의 어머니는 경미한 하복부 압통이 나타나는 과민성 대장 증후군 진단을 받았습니다.

정답

His vital signs were within the normal range The patients symptoms and family history suggest Irritable Bowel Syndrome IBS

활력 징후는 정상 범위 내에 있었습니다. 환자의 증상과 가족력은 과민성 대장 증후군 IBS를 시사합니다.

논문 요약 데이터

- 소아청소년과 보통 50개 질병
의학 논문 요약 데이터
- 질병 당 약 3건
- 전체 143건



Dataset Viewer

Auto-converted to Parquet API Go to dataset viewer

Split

train (143 rows)

Search this dataset

Disease string · lengths	Explanation string · lengths
	
Abdominal hernia (inguinal, umbilical, etc.)	abdominal hernias in pediatric patients 1. common and less common clinical manifestationsymptoms - common...
Abdominal hernia (inguinal, umbilical, etc.)	abdominal hernias in pediatric patients 1. common and less common clinical manifestationsymptoms - common...
Abdominal hernia (inguinal, umbilical, etc.)	abdominal hernias in pediatric patients 1. common and less common clinical manifestationsymptoms - common...
Abdominal migraines	abdominal migraines in pediatric patients 0-18 years 1. common and less common clinical manifestationsymptoms...
Abdominal migraines	abdominal migraines in pediatric patients 0-18 years 1. common and less common clinical manifestationsymptoms...
Abdominal migraines	abdominal migraines in pediatric patients 0-18 years 1. common and less common clinical manifestationsymptoms...

< Previous 1 2 Next >



논문 요약 데이터 예시

****Diabetic Ketoacidosis (DKA) in Pediatric Patients (0-18 years)****

1. ****Common and Less Common Clinical Manifestation/Symptoms****

****Common:****

– Polyuria (increased urination)

...

2. ****Natural History of the Disease****

– DKA is an acute complication that can occur in the course of type 1 diabetes.

...

3. ****Epidemiology****

– DKA is the leading cause of morbidity and mortality in children with type 1 diabetes. ...

4. ****Etiology****

...

모델 학습

미세 조정 프롬프트 (Stanford Alpaca Prompt)

###Instruction:

Use the Input below to create an instruction, which could have been used to generate the response using an LLM.

Input:

{sample['symptom']}

Response:

{sample['long_answer']}

모델 작동 예시

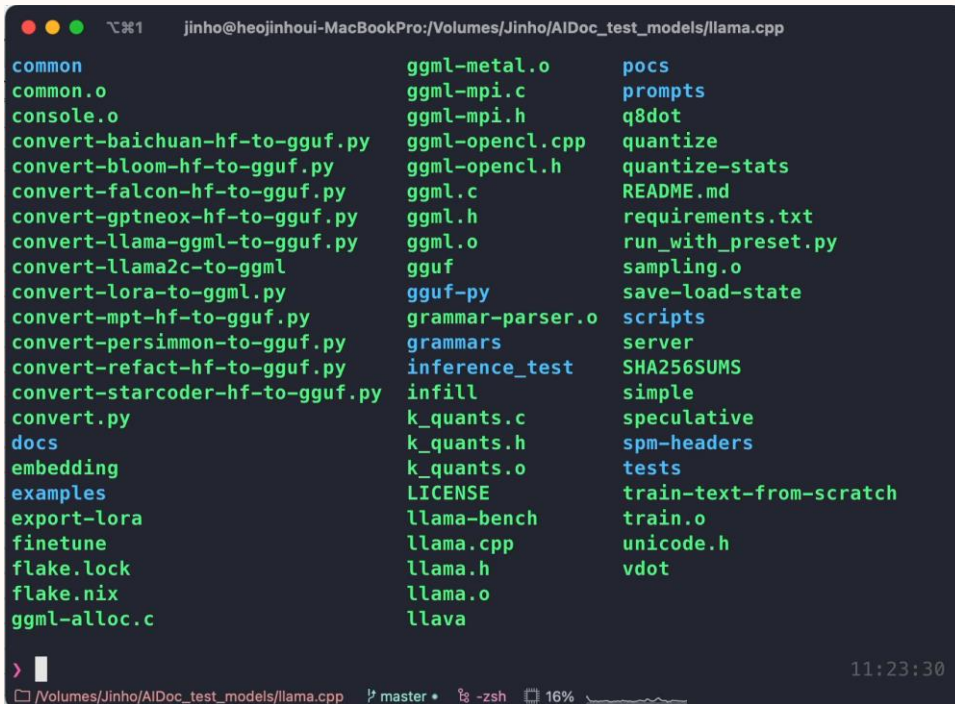
7b-knowledge

q8_0 모델 추론 영상

Apple M1 pro

16GB RAM

20 tokens/second



```
jinho@heojinhoui-MacBookPro: /Volumes/Jinho/AIDoc_test_models/llama.cpp

common
common.o
console.o
convert-baichuan-hf-to-gguf.py
convert-bloom-hf-to-gguf.py
convert-falcon-hf-to-gguf.py
convert-gptneox-hf-to-gguf.py
convert-llama-ggml-to-gguf.py
convert-llama2c-to-ggml
convert-lora-to-ggml.py
convert-mpt-hf-to-gguf.py
convert-persimmon-to-gguf.py
convert-refact-hf-to-gguf.py
convert-starcoder-hf-to-gguf.py
convert.py
docs
embedding
examples
export-lora
finetune
flake.lock
flake.nix
ggml-alloc.c
ggml-metal.o
ggml-mpi.c
ggml-mpi.h
ggml-opencl.cpp
ggml-opencl.h
ggml.c
ggml.h
ggml.o
gguf
gguf-py
grammar-parser.o
grammars
inference_test
infill
k_quants.c
k_quants.h
k_quants.o
LICENSE
llama-bench
llama.cpp
llama.h
llama.o
llava
pocs
prompts
q8dot
quantize
quantize-stats
README.md
requirements.txt
run_with_preset.py
sampling.o
save-load-state
scripts
server
SHA256SUMS
simple
speculative
spm-headers
tests
train-text-from-scratch
train.o
unicode.h
vdot

> |
```

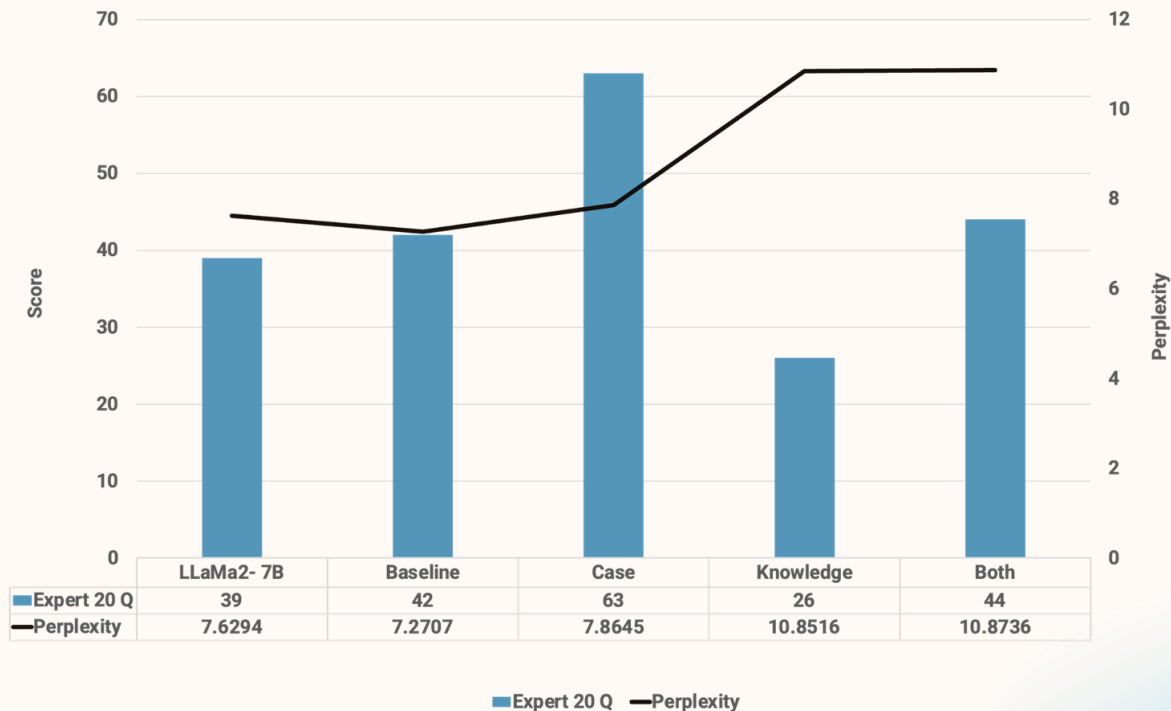
11:23:30

/Volumes/Jinho/AIDoc_test_models/llama.cpp master - zsh 16%

모델 평가

■ **Expert 20 QA**
전문가 출제 가상 소아
복통 환자 20례
모델 응답 전문가 채점

■ **Perplexity**
모호한 질문에 대한
응답 분산 지표
(낮을 수록 성능 높음)



결과 분석

1) 가상 환자 시나리오 데이터 학습 적합

서비스 제공 시 생성되는 데이터 이용, 모델 성능 개선 가능
경량 모델로도 선제적인 진단을 통한 신속한 판단 가능함 실험적 확인

2) 의학 지식 데이터 학습 시 성능 저하

데이터 부족 원인 추정
오버피팅 발생