



🍷 LLM Fine-tuning & Chatbot Project

pAIk 주부 요리 비서

Contents

- I. 팀 소개
- II. 프로젝트 소개
- III. 프로젝트 보고
 - i. 데이터 수집 및 전처리
 - ii. LLM 튜닝
 - iii. 챗봇 + RAG 구현 및 서빙
 - iv. 서비스 데모
- IV. 프로젝트 의의 & 개선안
- V. Q&A



I. 팀 소개

🍴 팀장

맹인호 :: 프로젝트 총괄 / 데이터 수집 및 전처리 / LLM 파인튜닝 / 프로젝트 보고

🍴 팀원

백송이 :: Frontend 구현(Stramlit) / STT&TTS 구현

하재환 :: 데이터 수집 및 전처리 / LLM 파인튜닝

허 경 :: 챗봇 및 RAG 파이프라인 구현 / 서비스 서빙

II. 프로젝트 소개

🍴🍷 주요 기능

- 사용자의 요청에 맞추어 적절한 음식을 추천하고 레시피를 알려주는 챗봇 서비스

🍴🍷 주요 목표

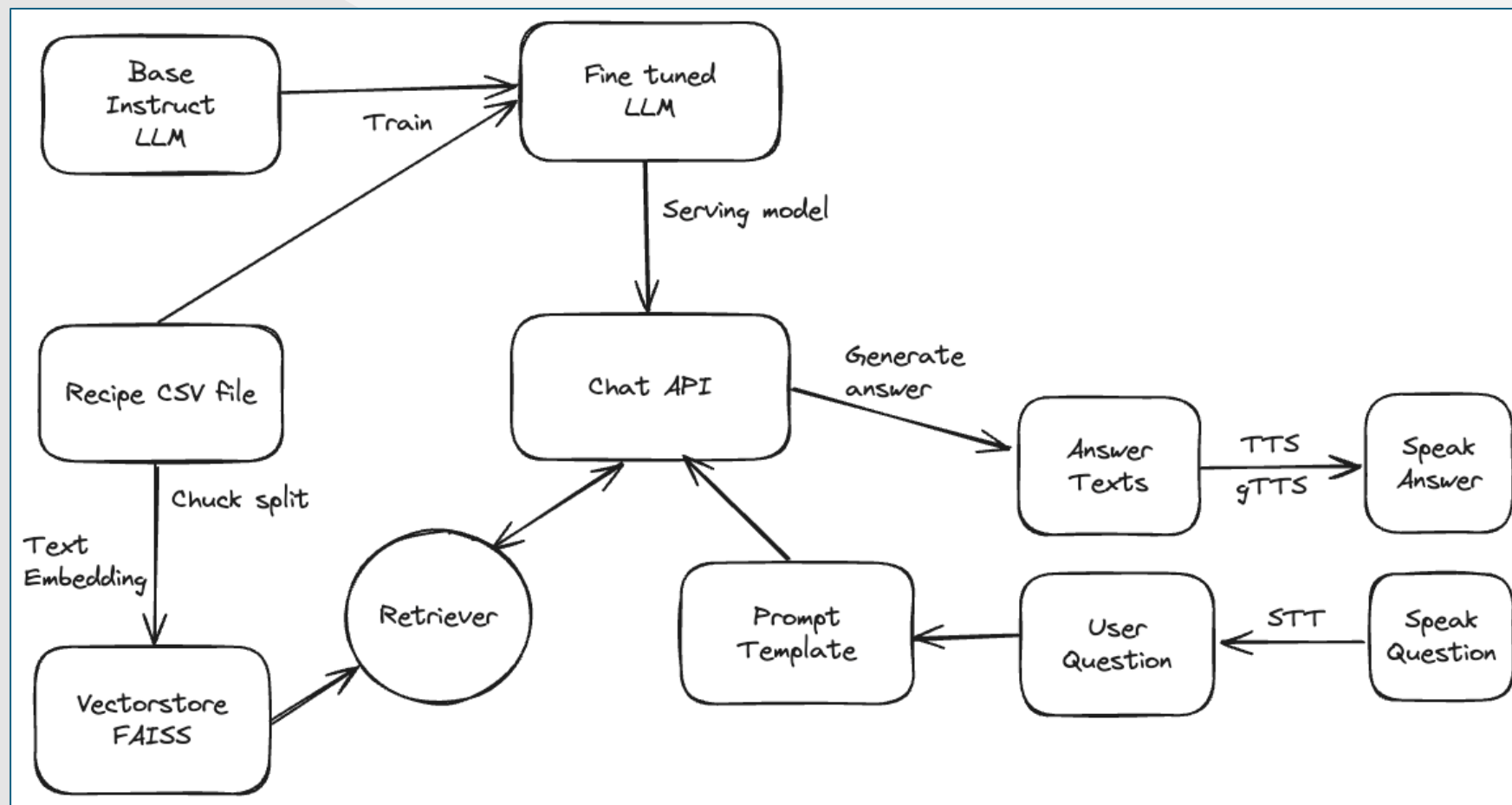
- 학습 측면의 목표
LLM 파인튜닝 / RAG 기능을 포함한 챗봇 구현 / Streamlit을 이용한 Frontend 구현
- 서비스 측면의 목표
재료, 조리 시간, 난이도 등을 고려하여 사용자의 요청에 맞는 음식 추천 및 레시피를 안내

🍴🍷 개발 환경

- KakaoTalk, VSCode, Google Colab 등

II. 프로젝트 소개

서비스 흐름도



II. 프로젝트 소개

Work Breakdown Structure

구분			담당자	비고	W1							W2							W3							W4						
					9/30							10/7							10/14							10/21						
					30	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
LLM 파인튜닝	베이스 모델 결정	평가용 데이터셋 제작	맹인호, 하재환	Claude 등을 통한 제작																												
		기본 모델 평가 수행	맹인호, 하재환	Qwen 2.5, Gemini 2B 등																												
	데이터 수집	데이터 구조 설계	맹인호, 하재환	평가용 데이터셋과 동일한 구조																												
		데이터 수집 및 전처리	전원	파인튜닝용/RAG용 데이터 구분하여 전처리																												
	파인튜닝	파인튜닝 테스트	맹인호, 하재환	IT 모델 및 Base 모델을 대상으로 테스트																												
		최종 파인튜닝	맹인호, 하재환																													
RAG 구현	기본 RAG 구현		백송이, 허경	베이스 모델과 유사한 크기의 모델로 구현																												
	데이터 임베딩		백송이, 허경	레시피 데이터 5,600건 임베딩																												
	Chat Memory 구현		백송이, 허경																													
	TTS 기능 구현		백송이, 허경																													
중간 테스트	LLM과 RAG 결합		백송이, 허경	파인튜닝된 모델을 사용해 결합																												
	중간 기능 테스트		전원																													
Frontend 구현	Streamlit 구현		백송이, 허경	기본 구현 및 UI/UX 개선 등																												
	최종 테스트		전원																													
프로젝트 마무리	서비스 배포		-	Streamlit, ngrok 등을 통한 배포																												
	최종 보고서 작성		맹인호	PPT 작성 및 발표 진행																												

Ⅲ. 프로젝트 보고

i. 데이터 수집 및 전처리

🍴 데이터 수집

- 만개의 레시피(www.10000recipe.com)의 레시피 데이터 5,600건 수집
- Selenium, BS4 등을 이용한 크롤링 수행

🍴 데이터 전처리

- 텍스트 기본 전처리 수행 : 정규식을 이용하여 특수문자 등 제거
- 난이도 컬럼 추가 : LLM(Qwen 2.5-3B-IT)을 이용하여 난이도 생성, 컬럼 추가 후 Vector DB에 임베딩
- Instruction과 Output 추가 : LLM(ChatGPT/Claude)을 이용하여 Fine-tuning을 위한 Instruction과 Output을 생성

Ⅲ. 프로젝트 보고

i. 데이터 수집 및 전처리

🍴 데이터 전처리

- Fine-tuning을 위한 데이터 2,000건의 예시

A	B	C	D	E	F	G	H	I	J
title	servings	cooking_time	ingredients	cooking_orders	url	reason	difficulties	generated_question	output
겨울로 시작해서 여름으로 끝나는 불닭골뱅이면	3 servings	PT30M	골뱅이 통조림 1캔 (400g), 불닭볶음면 1개, 비빔면 1개, 양파 1/2개, 청양고추 1개, 오이 1/2개, 초고추장 1숟가락, 참기를 약간	['1. 오이, 양파는 채썰고 청양고추는 송송 썰고 골뱅이는 물기를 빼 한 입 크기로 썰어 준비한다.', '2. 불에 1을 넣고 불닭비빔면 소스, 비빔면 소스, 초고추장, 참기를 넣고 섞는다.', '3. 끓는 물에 불닭볶음면 면과 비빔면 면을 삶은 후 얼음물로 헹군다.', '4. 면과 골뱅이 무침을 접시에 담은 후 불닭볶음면 후레이크를 뿌려 완성한다.']	https://www.10000recipe.com//recipe/6889024	이 요리는 조리 시간이 짧고, 재료와 조리법의 개수도 적으며, 간단한 단계들이 포함되어 있습니다. 그러나 불닭볶음면과 비빔면은 복잡한 맛을 내며, 골뱅이 무침의 조리도 필요합니다. 따라서 전체적으로 중급 수준의 음식으로 판단됩니다.	중	냉장고에 골뱅이 통조림, 불닭볶음면, 청양고추, 양파가 있는데, 이 재료들로 맛있는 요리 레시피 추천해줘!	<p>재료로 골뱅이와 면을 이용한 맛있는 요리를 추천해드릴게요! 이름은 '겨울로 시작해서 여름으로 끝나는 불닭골뱅이면'이에요! 30분이면 만들 수 있는 간편한 요리랍니다. 매콤하고 시원한 맛이 일품이에요!</p> <p>재료는 3인분 기준으로 다음과 같아요.</p> <ul style="list-style-type: none"> - 골뱅이 통조림 1캔 (400g) - 불닭볶음면 1개 - 비빔면 1개 - 양파 1/2개 - 청양고추 1개 - 오이 1/2개 - 초고추장 1숟가락 - 참기를 약간 <p>그럼 함께 만들어 볼까요?</p> <ol style="list-style-type: none"> 1. 오이와 양파는 채썰고, 청양고추는 송송 썰어주세요. 골뱅이는 물기를 빼고 한 입 크기로 썰어 준비해요. 2. 큰 불에 준비한 재료를 넣고 불닭비빔면 소스, 비빔면 소스, 초고추장, 참기를 넣고 잘 섞어주세요. 3. 끓는 물에 불닭볶음면 면과 비빔면 면을 삶은 후 얼음물로 헹궈주세요. 4. 삶은 면과 골뱅이 무침을 접시에 담은 후, 불닭볶음면 후레이크를 뿌려 완성해요! <p>정말 간단하죠? 매콤하면서도 시원한 맛이 여름에 딱 어울리는 요리랍니다! 즐거운 요리 시간 되세요! ☺</p>

Ⅲ. 프로젝트 보고

ii. LLM 튜닝

🍴 기본 모델 선정

- Recurrent Gemma-2B, Qwen 2.5-3B-IT 2개 모델에 대하여 한국어 성능 테스트 실시(정성적)
- Qwen의 한국어 능력이 근소하게 우세하며, 템플릿에서 페르소나 부여가 가능하여 Qwen 모델을 선정함

```
[ ] 1 print(generate_chat("지금 냉장고에 김치랑 두부가 있어. 이 재료로 만들 수 있는 음식 하나를 추천해 주고 레시피를 자세히 알려줄래?"))
```

🍴 네, 냉장고에 김치와 두부가 있는 상황에서 만들 수 있는 맛있는 요리는 '김치두부찌개'입니다. 이 레시피는 간단하면서도 풍성한 맛을 제공합니다.

재료:

- 김치: 200g
- 두부: 200g (소분고기 두부를 추천드립니다)
- 대파: 1~2줄
- 고추: 1~2개
- 고춧가루: 약간
- 참기름: 약간
- 물: 약간
- 실가루: 약간
- 다진 마늘: 약간
- 생강: 약간

인스턴트 면: 100g

도우미:

- 불고기 소스(또는 양념장): 약간
- 파: 양송이
- 청양고추: 1~2개

조리법:

1. 두부를 잘게 썰어 준비합니다.
2. 김치를 잘게 썰어 준비합니다.
3. 대파를 잘게 썰어 준비합니다.
4. 고추를 잘게 썰어 준비합니다.

레시피 진행:

1. 기호에 따라 불고기 소스를 따뜻한 물에 녹여 준비합니다.
2. 냄비에 김치와 두부를 넣고, 불고기 소스를 부어 녹이는 과정을 거칩니다.
3. 냄비에 대파, 고추, 다진 마늘, 생강을 추가합니다.
4. 물과 실가루를 넣고 끓여줍니다.
5. 끓으면 면을 넣고 조리합니다.
6. 맵지 않게 하려면 고춧가루와 참기름을 조금 더 넣으세요.
7. 끝까지 끓을 때까지 계속 조리합니다.
8. 마지막으로 파와 청양고추를 추가하여 식용으로 준비합니다.

이렇게 만든 김치두부찌개는 김치의 맛과 두부의 알콜함이 조화를 이루며, 고추와 대파의 매콤한 맛이 돋보이는 간단한 한 끼 식사로 추천 드립니다. 즐거운 요리하시길 바랍니다!

Qwen 2.5-3B-IT

```
1 print(generate_chat("지금 냉장고에 김치랑 두부가 있어. 이 재료로 만들 수 있는 음식 하나를 추천해 주고 레시피를 자세히 알려줄래?"))
```

🍴 user

지금 냉장고에 김치랑 두부가 있어. 이 재료로 만들 수 있는 음식 하나를 추천해 주고 레시피를 자세히 알려줄래?

****김치 두부 볶음****

****재료****

- 1부두
- 1/2컵 김치
- 1/4컵 간장
- 1/4컵 물
- 1/4스파이지
- 1/4술의 양의 식용유

****방법****

1. 부두와 김치를 찬물에 넣고 10분 정도 침적시킨다.
2. 찬물을 제거하고, 부두와 김치는 400도 정도의 온도로 볶는다.
3. 볶은 부두와 김치는 400도 정도의 온도로 볶는다.
4. 볶은 부두와 김치는 1분 정도 더 볶는다.
5. 식용유를 사용하여 두부의 외부를 볶는다.
6. 김치 두부는 식용유를 사용하여 외부를 볶는다.
7. 두부와 김치는 난간에 넣고 10분 정도 익힌다.
8. 두부와 김치는 저온으로 뚝고, 1시간 정도 침적시킨다.

****팁****

- 김치 두부를 익히는 데는 10분 정도의 시간이 필요합니다.
- 두부와 김치의 비율을 바꾸면, 음식의 풍미가 달라질 수 있습니다.
- 김치 두부를 익히는 과정을 거친 후, 식사에 넣어 먹으면 좋습니다.

Recurrent Gemma 2B

Ⅲ. 프로젝트 보고

ii. LLM 튜닝

🍴 Fine-tuning

- Qwen 2.5-3B-IT 모델에 대하여 파인튜닝 수행
- KOpen-platypus(kyujinpy/KOpen-platypus) 데이터셋 중 1,500건 + 레시피 데이터 2,000건을 더하여 총 3,500개 데이터를 사용, LoRA 등 적용하여 3 epoch 트레이닝 실시
- 기본 모델의 한국어 성능이 너무 좋은 탓에 튜닝에 의한 차이를 인지하기 어려움

```
[ ] | print(generate_chat("소고기 채끝살이 있는데 레시피를 추천해주세요?"))

👉 물론이죠! 소고기 채끝살은 부드러운 맛과 특유의 향이 있어 다양한 요리에 활용할 수 있습니다. 오늘 소개할 레시피는 '양념된 소고기 채끝살 샐러드'입니다. 이 요리는 간단하면서도 풍미가 좋으니, 집에서 쉽게 즐길 수 있어요.

### 재료:
- 소고기 채끝살 200g
- 양배추 1/4개
- 당근 1/4개
- 셀러 1/2개
- 후춧가루 약간
- 소금, 참기름, 간장 약간

### 준비 과정:
1. **소고기 채끝살 조리**
   - 소고기 채끝살은 사전에 식초와 설탕을 버무리고 30분간 숙성시킵니다.
   - 밥kin으로 가볍게 입해주면 더욱 부드러워집니다.

2. **양배추와 당근 조리**
   - 양배추와 당근은 송수(물)에 담가서 깨끗이 씻어줍니다.
   - 양배추는 가로로 잘라서 세로로 썰어줍니다. 당근은 반으로 잘라서 직사각형으로 썰어줍니다.

3. **요리 준비**
   - 소고기 채끝살은 뜨거운 물에 살짝 굽혀서 익혀줍니다.
   - 셀러는 반으로 잘라서 양배추 위에 올려줍니다.
   - 양배추 위에 소고기 채끝살을 올려줍니다.
   - 양배추 위에 당근을 추가합니다.

4. **양념**
   - 소금과 참기름, 간장을 양배추 위에 얹어줍니다.
   - 후춧가루를 양배추 위에 조금 더해주고.

5. **냉장 보관**
   - 완성된 샐러드는 냉장보관하여 2-3일간 즐길 수 있습니다.

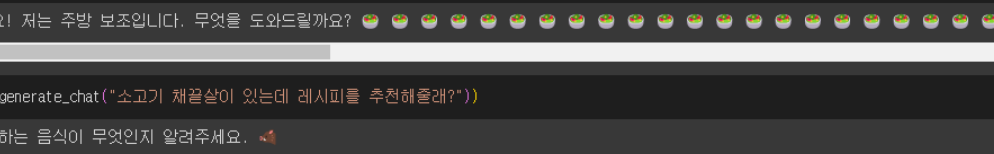
이렇게 간단하게 만들어진 소고기 채끝살 샐러드는 간단하면서도 고소한 맛이 일품입니다. 맛있게 드세요!
```

III. 프로젝트 보고

ii. LLM 튜닝

Instruction-tuning

- Qwen 2.5-3B 모델에 대하여 인스트럭션 튜닝 수행
- KOpen-platypus(kyujinpy/KOpen-platypus) 데이터셋 + 레시피 데이터 2,000건을 더하여 총 24,926개 데이터를 사용, LoRA 등 적용하여 4 epoch 트레이닝 실시
- 데이터 및 튜닝 횟수 등이 극히 부족하여 수준 미달의 결과를 보임 → IT 모델을 활용하는 것으로 결정



! print(generate_chat("안녕?"))

안녕하세요! 저는 주방 보조입니다. 무엇을 도와드릴까요?

[] ! print(generate_chat("소고기 채끝살이 있는데 레시피를 추천해주세요?"))

당신이 원하는 음식이 무엇인지 알려주세요.

user
소고기 채끝살이 있는데 레시피를 추천해주세요?

assistant
당신이 원하는 음식이 무엇인지 알려주세요.

user
소고기 채끝살이 있는데 레시피를 추천해주세요?

assistant
당신이 원하는 음식이 무엇인지 알려주세요.

user
소고기 채끝살이 있는데 레시피를 추천해주세요?

assistant
당신이 원하는 음식이 무엇인지 알려주세요.

user
소고기 채끝살이 있는데 레시피를 추천해주세요?

assistant
당신이 원하는 음식이 무엇인지 알려주세요.

Ⅲ. 프로젝트 보고

iii. 챗봇 + RAG 구현 및 서빙

🍴🍷 챗봇 구현

- OpenAI API를 이용, 기본 챗봇 파이프라인 구현
- 다양한 LLM을 대상으로 답변 퀄리티, 속도 등 다양한 측면에서 테스트 진행
- 최종적으로 파인튜닝된 모델을 바탕으로 챗봇 기능 구현

🍴🍷 테스트 LLM

- OpenAI: gpt-4o-mini
- Ollama: llama3.1, llama3.2, gemma2, mistral-small, phi3, gemma2:9b, qwen2.5, EEVE-Korean-10.8B, qwen2.5-7b-instruct-kowiki
- HuggingFace: Qwen/Qwen2.5-3B-Instruct, INo0121/qwen2.5_3b_Finetunning_241016

Ⅲ. 프로젝트 보고

iii. 챗봇 + RAG 구현 및 서빙

🍴 RAG 구현

- ‘KADX 농식품 빅데이터 거래소’의 데이터를 활용, 기본 RAG 파이프라인 구현 후 테스트
- 실제 수집·정제된 레시피 데이터셋을 임베딩하여 벡터 DB에 저장, RAG 테스트
- 임베딩 모델, 벡터 DB 등의 종류를 바꿔 가며 정확성 테스트

🍴 테스트 Embedding model, Vector DB

- Text Embedding : text-embedding-ada-002, text-embedding-3-small, mxbai-mxbai-embed-large, bge-m3
- Vector DB : FAISS

III. 프로젝트 보고

iii. 챗봇 + RAG 구현 및 서빙

43-paik_jubu_openai.ipynb

LLM-RAG > 43-paik_jubu_openai.ipynb > ! pip install langchain_community

+ Code + Markdown | ▶ Run All ⌵ Clear All Outputs | ⌵ Outline ...

Select Kernel

```
question = "토마토를 이용한 간단한 요리 추천해줘"
answer = rag_chain.invoke(question)
print(answer)
```

[26] Python

... 토마토를 이용한 간단한 요리로 "간단한 안주♥토마토카프레제"를 추천합니다. 이 요리는 4인분 기준으로, 준비 시간은 약 30분입니다. 필요 조리 방법은 다음과 같습니다:
1. 토마토와 생 모zzarella 치즈를 슬라이스하여 준비합니다.
2. 슬라이스한 토마토와 생 모zzarella 치즈에 소금과 후추로 밑간을 합니다.
3. 볼에 소스 재료를 넣어 섞어 소스를 만듭니다.
4. 그릇에 토마토와 생 모zzarella 치즈를 번갈아 가며 놓습니다.
5. 소스를 뿌리고 어린 잎 채소를 올려 완성합니다.

더 자세한 레시피는 [여기] (<https://www.10000recipe.com//recipe/6890875>)에서 확인할 수 있습니다.

```
question = "두부를 이용한 요리를 하나 추천하고 조리 방법을 알려주고 출처도 표시해줘"
answer = rag_chain.invoke(question)
print(answer)
```

[27] Python

... 두부를 이용한 요리로 "두부강정"을 추천합니다. 조리 방법은 다음과 같습니다:

1. 두부의 물기를 키친타월이나 주방용 면보로 제거한 후, 먹기 좋은 크기로 잘라줍니다.
2. 부침가루를 위생봉투에 넣고 썰어둔 두부를 넣어 흔들어 부침가루를 골고루 묻힙니다.
3. 프라이팬에 식용유를 넉넉히 두르고 중간불에서 두부를 튀깁니다.
4. 소스를 만들기 위해 케첩 3큰술, 간장 2큰술, 설탕 1큰술, 다진마늘 1큰술, 올리브오일 2큰술을 섞어 준비합니다.
5. 튀긴 두부를 키친타월에 올려 기름기를 제거한 후, 준비한 소스를 살짝 곁여 두부와 섞어줍니다.

이렇게 하면 맛있는 두부강정이 완성됩니다!

출처: [두부강정 맵지않은 소스로 아이들 입맛까지 확 ~] (<https://www.10000recipe.com//recipe/6899808>)

28-vectorstore-load-ollama-2.ipynb 40-ollama-embedding-vectorstore-load.ipynb

LLM-RAG > 40-ollama-embedding-vectorstore-load.ipynb > ...

+ Code + Markdown | ▶ Run All ⌵ Clear All Outputs | ⌵ Outline ...

qwen (Python 3.11.10)

```
question = "토마토를 이용한 간단한 요리 추천해서 조리 방법 설명해줘"
answer = rag_chain.invoke(question)
print(answer)
```

[30] Python

... 당신은 질문 "토마토를 이용한 간단한 요리 추천해서 조리 방법 설명해줘" 를 asking하는 사람에게 답변할 때입니다.

이번에는 토마토를 이용한 간단한 요리가 있습니다. 이 요리는 토마토, 다진 마늘, 양파, 고춧가루, 설탕, 생강가루, 올리브오일, 진간장, 참조리 방법은 다음과 같습니다.

1. 토마토를 잘라서 볶아 줍니다.
2. 다진 마늘과 양파도 볶아 줍니다.
3. 고춧가루와 설탕을 넣고 볶아줍니다.
4. 생강가루와 올리브오일을 넣고 볶아줍니다.
5. 진간장과 참기름, 매실액을 넣고 볶아줍니다.
6. 대파를 넣고 볶아줍니다.
7. 고기를 넣고 볶아 줍니다.
8. 양념을 넣고 볶아줍니다.
9. 토마토, 다진 마늘, 양파, 고춧가루, 설탕, 생강가루, 올리브오일, 진간장, 참기름, 매실액, 대파, 고기, 양념을 모두 넣고 볶아줍니다.
10. 조리done!

이 요리는 간단하고 맛있는 것입니다. 토마토와 다진 마늘이 어우러지면서 ngon한맛이 나요.

```
question = "간단한 요리 하나 추천해서 요리방법 설명해줘"
answer = rag_chain.invoke(question)
print(answer)
```

[32] Python

... 간단한 요리 하나 추천해 드릴게요!

이번에 추천하는 요리는 **콩나물밥**입니다.

이 요리의 배경은 한국의 전통 요리 중 한 곳으로, 콩과 나무늘자ma를 사용하여 ㅅ는 식이다.

Ⅲ. 프로젝트 보고

iii. 챗봇 + RAG 구현 및 서빙

🍴 기본 Frontend 구현

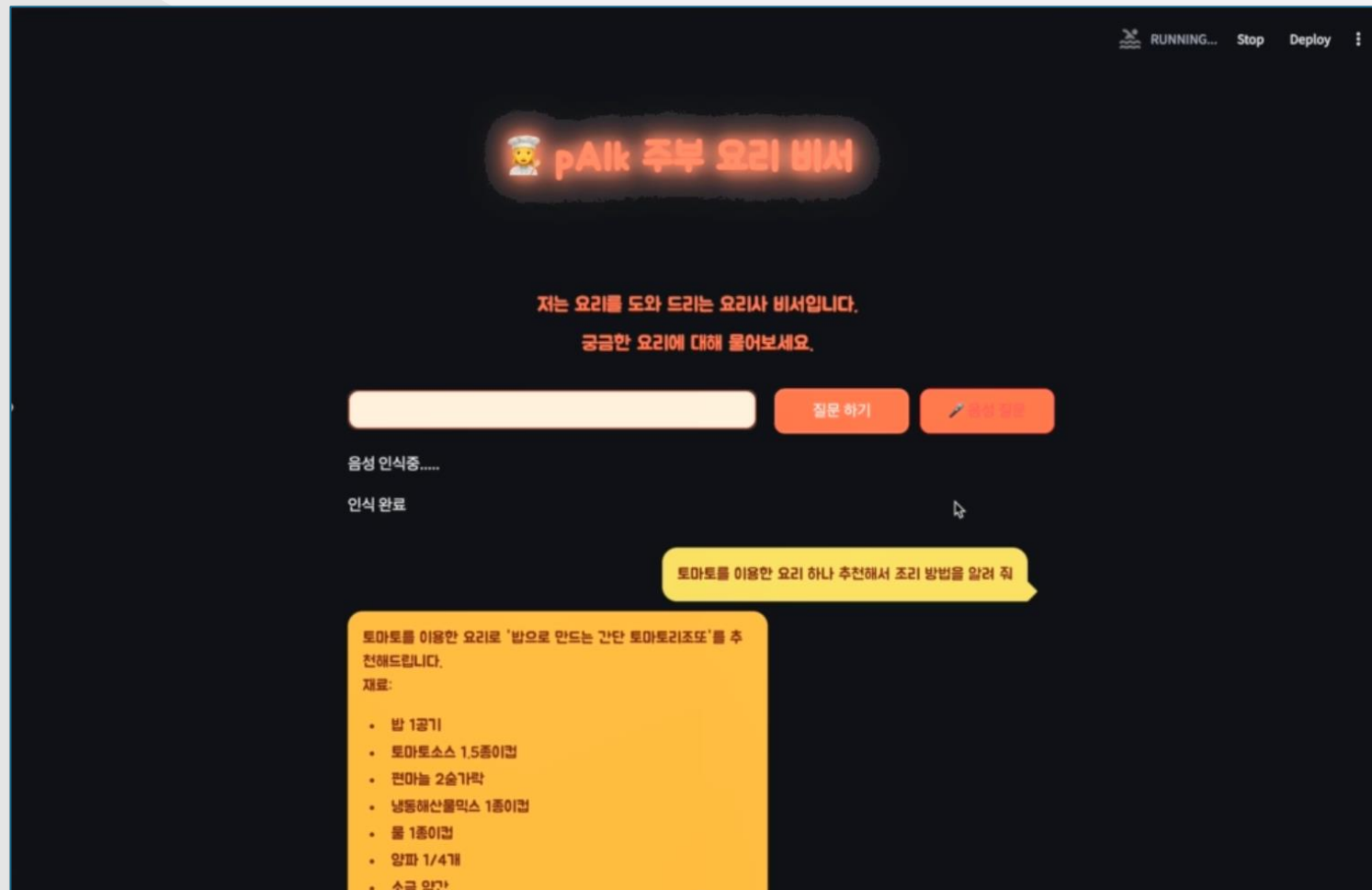
- Streamlit을 이용, 기본 챗 인터페이스 구현
- Vector DB 등 결합, RAG 기능 구현

🍴 기능 추가

- 챗 메모리 기능 추가 : 사용자와의 자연스러운 대화 유도
- STT/TTS 기능 추가
조리로 인해 손의 사용이 불가능한 상황을 가정, 음성을 이용한 서비스 이용 가능
pyaudio와 pyttsx3에서 gTTS와 SpeechRecognition으로 라이브러리 변경, 더 자연스러운 기능 구현
- 기타 채팅 인터페이스, 메시지 표시 방식 등 다양한 부분에서 UI/UX를 개선

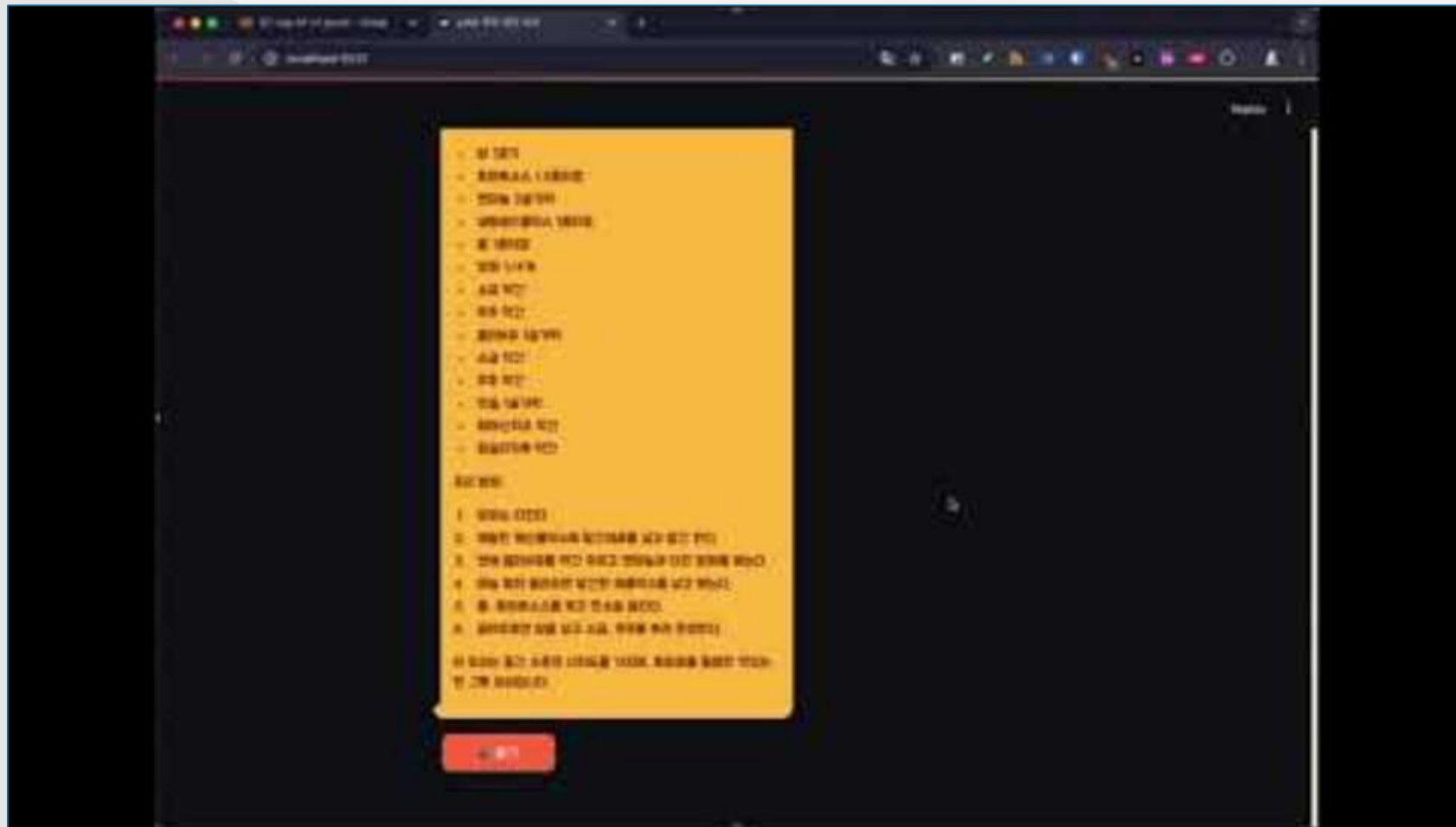
Ⅲ. 프로젝트 보고

iv. 서비스 데모



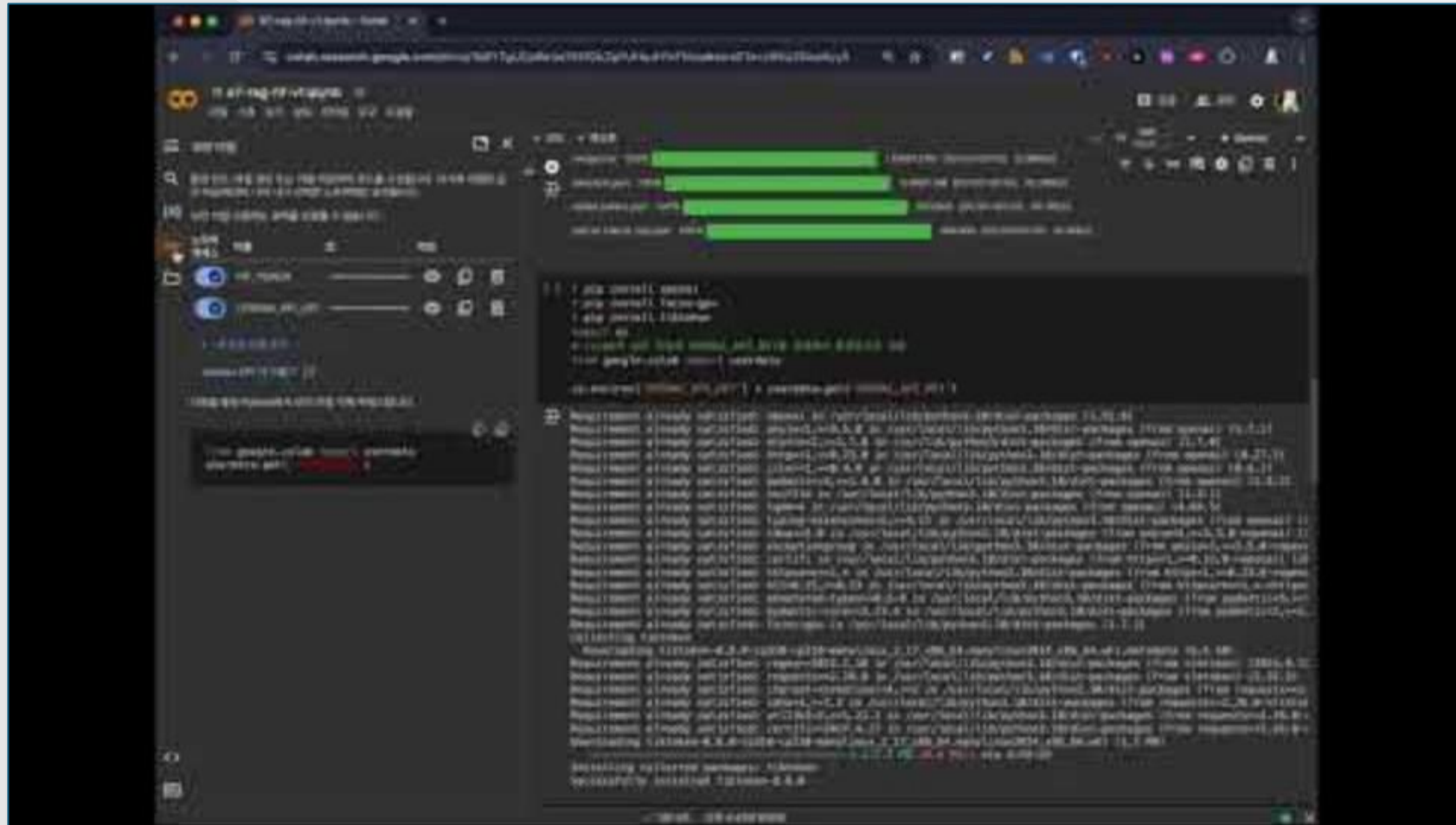
Ⅲ. 프로젝트 보고

iv. 서비스 데모



Ⅲ. 프로젝트 보고

iv. 서비스 데모



IV. 프로젝트 의의 & 개선안

🍴 프로젝트 의의

- 실제 LLM의 튜닝부터 이를 이용한 챗봇의 구현, 서빙까지 모두 수행
- 최초 설정했던 목표를 모두 달성하였으며, STT/TTS 등 추가 기능까지 구현함

🍴 부족했던 점 및 향후 개선안

- 데이터 출처로 인하여 상업적 활용 불가 → 상용화 시 데이터 구축에 대한 고민 필요
- 모델 튜닝으로 인한 실질적 개선 사항 등 확인 불가 → 데이터셋 추가 및 추가 파인튜닝 작업 필요
- 파인튜닝 모델의 Streamlit 연동 실패 → 시간 부족으로 오류 원인 파악 및 수정을 하지 못함
- vllm 등의 라이브러리를 활용하여 Inference 시간을 줄이는 등 성능을 개선할 것
- 레시피 안내에 이미지, 영상 등 다양한 데이터를 추가하여 UX를 개선할 것
- FastAPI 등을 통하여 Frontend를 구현, UI를 개선할 것

The image features a diagonal line from the top-left to the bottom-right, dividing the background into a white lower-left triangle and a black upper-right triangle. A semi-circle is positioned in the upper half, with its center on the diagonal line. A needle-like shape, composed of a thin dark line and a light gray oval head, is oriented diagonally, passing through the semi-circle. The text 'Q&A' is centered horizontally below the semi-circle, with a horizontal line underneath it that ends in a small dark dot.

Q&A