



프로젝트 소개 (2)






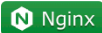




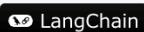












[GitHub Link](#)

프로젝트 소개

프로젝트를 진행하게 된 배경과 목표, 기획 의도 등 소개

- 프로젝트명 : 워라벨 메이트
- 프로젝트 주제: 워라벨 메이트는 조직 내부 구성원의 워크라이프 밸런스 향상을 목표로 한 LLM 기반 업무 자동화 시스템입니다. 일상 업무에서 시간을 많이 소모하는 작업들을 인공지능으로 자동화하여, 직원들이 보다 핵심 업무에 집중하고 업무 효율을 높일 수 있도록 도와줍니다
- 진행 기간 : 2022.04.22 ~ 06.18 (1개월 28일)

- 팀원: 5명
- 주요 업무 : Database 연동
- 활용 기술

분류	기술
Frontend	 Vite  React  JavaScript  Nginx
Backend	 FastAPI  Python
LLM 모델	 Ollama  Qwen2.5
RAG 프레임워크	 LangChain
STT (음성 인식)	 WhisperX
벡터DB	 Qdrant
문서 임베딩	 SBERT  all-MiniLM-L6-v2
데이터 전처리	 PyMuPDF  HWP  olefile
데이터 저장소	 MySQL  Qdrant Client
배포 환경	 Docker  Shell Script
API 연동	 REST API

- 프로젝트 목표:

반복적이고 수작업 기반의 행정 업무를 LLM 기반 인공지능 시스템으로 자동화

프로젝트 진행과정 및 나의 역할

1. 기능 기획

- 회의록 자동 작성 및 요약 (NoteMate)
- 사내 문서 기반 질의 응답 (ChatMate)
- 반복 민원 자동 응답 처리 (QueryMate)

- 전화 상담 녹취 분석 (CallMate)

2. 데이터 수집 및 전처리

✓ What I do

- 비정형 문서(사내 문서) 및 정형 문서(법령) 수집
- 텍스트 임베딩 및 벡터화
 - 임베딩 모델: BM-K/KoSimCSE-roberta
 - 임베딩 방식:
 - AutoTokenizer → 토큰화
 - AutoModel → 문장 임베딩
 - 평균 풀링 → 최종 벡터 생성
- Qdrant 기반 벡터 검색

3. 인공지능 모델 개발

- 인공지능 모델 선정: Qwen 2.5 (7B, Base)
- 용도 → 문서 요약, 질의 응답, 민원 답변, 회의록 요약 등

✓ What I do

<민원 자동 답변>

- "문장 임베딩 → 벡터 검색 → LLM 모델 답변"의 과정으로 진행
- 저장된 질문과 동일한 쿼리를 입력하여, 벡터 검색 결과의 일관성과 답변의 변동 여부를 확인
- 기존 벡터에 없는 질문의 경우, 외부 웹 검색 연동

4. 웹 사이트 구현

5. 유지보수

- 실제 기능 테스트
- 코드 리팩토링

✓ What I do

- Prompts 추출 파일 리팩토링
- RunPod 기반 모델 서빙 환경 운영 및 유지보수

✓ 프로젝트 개선사항

향후에는 다양한 형식의 **파일 지원 확대**를 통해 입력 유연성을 높이고, **다국어 대응 및 발화자 구분 기능**을 추가하여 사용자의 언어·상황에 맞는 정확한 응답이 가능하도록 개선하고자 합니다. 또한 **데이터 보강 및 모델 튜닝**을 통해 응답 품질을 지속적으로 고도화하고, 궁극적으로는 **실시간 AI 음성상담 기능**까지 확장하여 더 직관적이고 효율적인 사용 경험을 제공하는 것을 목표로 합니다.

✓ 프로젝트 성과 및 배운점

- 다양한 사내 문서를 벡터화하고 Qdrant 기반 검색 시스템을 연동하여, **정형·비정형 문서에 대한 질의 응답 기능**을 구현했습니다.
- 개인적으로는 **임베딩 파이프라인 구축, LLM-기반 질의 응답 체계 설계, RunPod를 활용한 모델 서빙 환경 운영 등 실제 AI 시스템 운영에 필요한 전반적인 역량을** 체득할 수 있었습니다.
- 특히, 유사 쿼리 검색 정확도와 답변 일관성을 확인하는 과정에서 **RAG 시스템의 작동 원리를 깊이 있게 이해**하게 되었으며, 코드 리팩토링 및 모델 배포 과정은 **지속 가능한 AI 서비스 운영의 기반을 다지는 계기**가 되었습니다.