SK네트웍스 Family AI과정 3기

LLM 활용 소프트웨어

□ 개요

- 산출물 단계 : LLM 활용 소프트웨어
- 평가 산출물 : 개발된 LLM 활용 소프트웨어
- 제출 일자 : 2024.12.24
- 깃허브 경로 : https://github.com/SKNETWORKS-FAMILY-AICAMP/SKN03-FINAL-6Team
- 작성 팀원 : 최연규

개요	● 목적: LLM을 활용하여 차량 관련 서비스를 지원하는 챗봇 소프트웨어 구현
데이터 전처리	● **추천 데이터** ■ **내용:** ◆ 차량 스펙 데이터: 차량 모델, 제원, 가격 등. ◆ 사용자 후기 데이터: 사용자의 차량 사용 경험, 만족도 등. ■ **출처:** ◆ 제네시스 공식 홈페이지 ◆ 엔카 ● **매뉴얼 데이터** ■ **내용:** ◆ 차량 매뉴얼 PDF 파일: 사용 설명서, 기능 설명 등. ■ **출처:** ◆ 제네시스 공식 홈페이지 ● **보험 데이터** ■ **내용:** ◆ 보험 약관 PDF 파일: 계약 조건, 보장 내용 등. ■ **출처:** ◆ 보험사 제공
기술적 구현	● LLM과 벡터 데이터베이스 연동 ■ 벡터 데이터베이스: ◆ Milvus를 사용하여 문서를 벡터화하고, 효율적인 검색 시스템을 구축 ◆ openAI API와 ,bge-m3모델을 활용해서 문서를 벡터로 변환 ■ 프롬프트 설계: ◆ 사용자의 질문 유형에 따라 적절한 프롬프트를 생성하여 LLM에 입력 ◆ 프롬프트와 벡터 연동 방식: 1. 사용자의 질문을 입력받아 벡터화 2. 벡터화된 질문을 가지고 벡터 데이터베이스에서 적절한 문서를 검색 3. 검색결과를 reranking 4. Reranking된 결과를 사용자 질문과 조합하여 LLM에 전달
코드 모듈화	● **데이터 처리 모듈** ■ **파일:** utils/preprocess.py ■ **역할:** 문서 전처리와 관련된 함수들로 텍스트 정제, 토큰화, 데이터 변환 등을 수행. ● **PDF 로딩 모듈** ■ **파일:** utils/pdf_loader.py ■ **역할:** PDF 파일에서 텍스트를 추출하고, 문서 형태로 변환. ● **벡터 검색 모듈** ■ **파일:** tools/milvus_search.py ■ **역할:** Milvus를 활용하여 벡터 데이터 검색 및 관련 로직을 처리. ● **프롬프트 생성 모듈**

	■ **파일:** models/prompt_templates.py ■ **역할:** 다양한 상황과 질문 유형에 적합한 프롬프트 템플릿을 생성. ● **임베딩 생성 및 저장 모듈** ■ **파일:** models/embeddings.py ■ **역할:** 문서 데이터를 임베딩 벡터로 변환하고 저장. ● **RAG 노드 모듈** ■ nodes/nodes.py ■ Retrieval-Augmented Generation(RAG)에서 사용되는 노드 로직 관리. ● **RAG 파이프라인 모듈** ■ **파일:** nodes/rag_pipeline.py ■ **역할:** RAG 시스템의 주요 데이터 흐름과 파이프라인 로직을 구현. ● **챗봇 응답 모듈** ■ **파일:** router/chatbot.py ■ **역할:** FastAPI 라우터를 활용해 사용자 요청을 처리하고 응답을 반환. ● **재랭킹(Reranking) 모듈** ■ **파일:** models/reranker.py ■ **역할:** 검색 결과를 재정렬하여 사용자가 원하는 결과를 우선적으로 제공. ● **LLM 모델 관리 모듈** ■ **파일:** models/model.py ■ **역할:** LLM 호출 및 모델 관련 로직 관리.
보완처리	● 환경 변수 활용: ■ OpenAI API Key와 같은 민감한 정보는 **AWS Parameter Store**에 저장하여 사용
도구 및 환경	● 도구: ■ OpenAI API ● 데이터베이스: ■ MySQL ■ Milvus ● 클라우드 환경: ■ AWS
결론	■ 주요 성과: ■ 제네시스 차량 매뉴얼 기반 Q&A 챗봇: 매뉴얼 기반으로 사용자 질문에 대한 정확한 응답 제공. ■ 차량 추천 챗봇: 사용자 데이터 분석을 통해 맞춤형 차량 추천 제공. ■ 보험 약관 기반 Q&A 챗봇: 복잡한 약관 내용을 간단히 요약하여 응답. ■ 향후 계획:

	■ **추가 기능 개발**: ◆ 다국어 지원으로 글로벌 사용자 타겟팅. ◆ 사용자 피드백 학습을 통한 성능 개선. ■ **데이터 확장**: ◆ 다양한 차량 브랜드의 매뉴얼 및 사용자 데이터를 추가 확보. ◆ 보험 약관 데이터의 최신화 및 추가 데이터 확보.