

실증적SW개발프로젝트 주간보고 (5주차)

작성일: 2025/04/07 팀: CivilWord

팀 활동 보고	활 동 일 시	2025-03-31 ~ 2025-04-7
	장 소	승학 도서관 스터디룸 및 비대면 진행
	참 석 자	양한나, 배준영, 임세희, 임지은, 최지은
	특 이 사 항	없음

이번주 진행사항	<p>0. 팀 공통 진행상황</p> <p>-와이어프레임 및 기능명세서 작성</p> <p>-백엔드 FastAPI 기반 구조 및 가상환경 구축, 더미 테스트</p> <p>-사하구청 민원 데이터 크롤링 및 엑셀 파일로 저장</p> <p>LLM은 Mistral 7B 모델을 기반으로 LoRA 튜닝을 진행하고 LangChain과 연동하여 테스트 프롬프트 검증</p> <p>1. 프론트엔드(임세희, 최지은)</p> <p>- 기능 명세서 작성 완료: 사용자 인터페이스(UI)에서 제공할 주요 기능(민원 접수 입력, 자동 응답 보기, 검색, 기록 확인 등)을 정리하고, 기능 흐름도 포함해 명세함.</p> <p>- 와이어프레임 작성 완료: Figma를 활용하여 주요 화면 구조를 시각화함.</p> <p>2. 백엔드(양한나)</p> <p>- API 경로 설계 완료: /request, /response, /admin, /log등 RESTful 방식의 주요 엔드포인트 정의.</p> <p>- Python venv 가상환경 및 자동화 환경 구축</p> <p>: Ubuntu(Wsl) 환경에서 venv로 가상환경 생성.</p> <p>: requirements.txt, .env, uvicorn, auto-reload등 FastAPI 서버 자동 실행 설정 완료.</p> <p>- FastAPI 기반 LLM 응답 더미 테스트: LLM 서버 연동 전, JSON 형태의 더미 응답을 미리 생성하여 프론트엔드와의 연동을 테스트함.</p> <p>3. 데이터 크롤링(임세희, 최지은)</p> <p>- 사하구청 전자민원 창구 대상 데이터 수집 테스트: BeautifulSoup, Requests 등 사용하여 민원 게시글 목록, 상세내용, 작성일시, 처리결과 등 크롤링.</p> <p>- 엑셀 저장 테스트 완료</p> <p>: Pandas를 이용해 수집한 데이터를 .xlsx형태로 저장하는 작업 수행.</p> <p>: 필드 정규화: "제목", "작성일", "민원 내용", "처리 결과" 컬럼 구성 테스트 완료.</p>
-------------	--

	<p>4. LLM 파인튜닝(배준영, 임지은)</p> <ul style="list-style-type: none"> - LLM 개념 공부 <ul style="list-style-type: none"> -> LLM 구조 & 모델, 파라미터, 양자화 & 비양자화, 토큰과 컨텍스트 등 다양한 개념 파인튜닝을 하게 될 때, overfitting이나 underfitting 조심해야 함. - LLaMA 2 7B 모델에서 Mistral 7B 모델로 변경 <ul style="list-style-type: none"> -> LLaMA보다 Mistral이 다국어 지원이 우수, 한국어 답변이 필요함. - Hugging Face에서 사용 허가 받은 후 소규모 데이터로 학습 <ul style="list-style-type: none"> -> LoRA (Low-Rank Adaption) 을 통한 경량화 테스트 - LoRA 튜닝 진행 후 LangChain 연결 <ul style="list-style-type: none"> -> LangChain은 LLM 활용한 애플리케이션 개발에 특화된 오픈소스 프레임워크 -> 외부 데이터 소스와 통합하여 설계할 수 있을 것으로 예상 - 테스트 프롬프트 작성 후 실행하여 답변 테스트
<p>다음주 계획</p>	<p>0. 팀 공통 진행계획</p> <ul style="list-style-type: none"> - 데이터 전처리 계획 수립 및 일부 실행 - 불필요한 필드 제거, 텍스트 정제 방식 등 사전 합의 - DB 테이블 설계 및 백엔드 연동 구조 논의 - 민원, 응답, 로그 관련 테이블 구조 및 키 설정 - 대량 데이터 크롤링 본격 진행 - 사하구 외 구청 또는 과거 민원 데이터까지 확장 수집 - 사하구 데이터 기반 LLM 파인튜닝 <p>1. 백엔드(양한나)</p> <ul style="list-style-type: none"> - FastAPI 서버에 DB 연결 기능 구현 - 크롤링 데이터 기반 입력 → 저장 → 조회 테스트 수행 - LLM 추론 결과를 JSON 형태로 반환하도록 구조 정비 <p>2. 데이터 크롤링 및 엔지니어링(임세희, 최지은)</p> <ul style="list-style-type: none"> - 대량 크롤링 로직 정비 및 실행 (자동화/오류처리 포함) - 수집된 민원 데이터를 기반으로 전처리 파이프라인 제작 - 모델 학습에 적합한 형식(CSV/JSONL 등)으로 변환 <p>3. LLM 담당(배준영, 임지은)</p> <ul style="list-style-type: none"> - 크롤링된 실데이터를 이용한 Mistral 7B 학습 초기 테스트 - LoRA 기반 파인튜닝 스크립트 정비 - 학습 후 LangChain과 연동하여 RAG 포함한 추론 테스트