

심화 프로젝트 기획서

팀 명	우드저지	트랙	NLP
팀 원	문상혁, 민지원, 신진수, 안성재, 우은진, 장은서, 한지훈		

주제	RAG 활용 유사 판례 기반 법률 조언 서비스
프로젝트의 목적 및 배경	1. 문제 상황 설명 - 법률 문제에 직면한 일반 사용자는 법률 지식 부족으로 인해 대응이 어렵고, 변호사 선임에는 높은 비용이 발생함. - 본 서비스는 무료로 제공되는 유사 판례 기반 법률 조언 시스템으로, 비용 부담 없이 신속하고 신뢰도 있는 법률 판단을 도와주는 것을 목표로 함. 2. 기술적 배경 및 시장 동향 - LLM 기반 법률 상담 서비스: GPT 기반 법률 상담 챗봇은 사용자의 자연어 질의에 응답할 수 있지만, 대부분 공개된 웹 문서를 기반으로 응답을 생성함. - 기존 판례 검색 서비스: 예를 들어, bigcase.ai는 방대한 판례 데이터를 보유하고 있으며, 사용자는 키워드 기반 검색이나 다양한 필터 기능을 통해 유사 사건을 찾아볼 수 있음. 다만, 자연어 질의 입력은 지원하지 않으며, 판례 결과를 기반으로 한 구체적인 행동 조언은 제공하지 않음. - RAG의 기술적 배경: Retrieval-Augmented Generation(RAG)은 LLM의 한계로 지적되는 사실성 부족(factuality issue)을 보완하기 위한 방식으로 최근 다양한 분야에서 주목받고 있음. 특히 법률, 의료, 금융 등 출처 기반 정보의 신뢰성이 중요한 영역에서 활용 사례가 증가하고 있음. RAG는 사용자 질의에 대해 외부 지식 소스로부터 관련 정보를 검색하고, 이를 기반으로 응답을 생성함으로써 출처가 명확하고, 최신성이 보장된 결과를 제공할 수 있음. - 본 프로젝트에서는 응답의 신뢰도와 설명 가능성이 핵심이므로, 단순 생성형 모델이 아닌 판례 검색과 생성의 결합이 가능한 RAG 구조를 채택함.

정량 / 정성적 평가 지표	- 정량적 목표 - 판례 검색 Top-5 Recall $\geq 80\%$ - RAG 응답의 BERTScore 기준 baseline 대비 $\geq 15\%$ 향상 - 정성적 목표 - 사용자 설문 기반 만족도 평가 ≥ 4.2점 Ex) "제안된 대응 전략이 현실적으로 도움이 된다고 느꼈습니까?" "내가 입력한 상황과 제시된 판례가 유사하다고 느꼈습니까?"
핵심 아이디어 및 차별점	- 핵심 아이디어 - 사용자가 자연어로 자신의 상황을 설명하면, 유사한 실제 판례를 찾아 해당 판례를 근거로 법적 대응 방향을 안내하는 서비스임. - 비용 부담 없이 실제 사례 기반의 법률 조언을 받을 수 있도록 하는 것을 목표로 함. - 출력 창 구성요소 - 판단 결과 요약 Ex) 유사 판례와의 유사점 및 유사도: .../ 87% 판례 요약: ... 판결 결과: 벌금 50만 원 선고 - 추천 대응 전략 Ex) 합의 제의: 상대가 사과 및 일정 금액 보상 시 실행 가능성 낮음 증거 확보: CCTV 영상은 핵심 근거가 될 수 있음 변호사 상담: 상대가 법적 대응을 예고할 경우 권장 - 벌금/처벌 강도 제시 - 피해자 입장에서 기존 판례들 기반으로 예상 벌금/처벌 강도 제시 자체가 유의미할 수 있음 Ex) 예상 처벌이 사회봉사인 경우 → 고소하지 않음 예상 벌금이 변호사 선임 비용보다 적은 경우 → 고소하지 않음 예상 벌금이 매우 높은 경우 → 고소 진행함 - 차별점 - GPT와 같은 LLM은 판례 전문 데이터에 접근하기 어려워 실제 법률 판단과 차이가 발생할 수 있음. - 기존 판례 검색 서비스는 자연어 질의가 불가능하고, 키워드 입력이나 필터 설정이 필요함. - 본 서비스는 RAG(Retrieval-Augmented Generation) 기법을 활용하여 사용자의 질의에 대해 유사 판례를 실시간으로 검색하고, 해당 내용을 기반으로 조언을 생성함.

	• 사용자는 이 서비스를 통해 자신의 상황과 유사한 판례 정보, 해당 판례를 바탕으로 한 구체적인 대응 조언, 판결의 법적 근거를 확인할 수 있음.
데이터	• AI Hub 민원 텍스트 / 법률상담 데이터셋 • 공공데이터포털 판례 파일 (형식 및 메타데이터 정제 필요) • 자체 구축 데이터셋 (판례 요약 + 법률 조언 Pair 구축)
모델링 기법 및 분석 방법	1. 사용자 입력 • 사용자는 자연어로 자신의 법적 상황을 입력함. Ex: "택배 기사가 집 앞에 택배를 던지고 가서 물건이 파손됐습니다." 2. RAG Retriever – 유사 판례 검색 • 판례를 벡터화하기 위한 사전 학습된 법률 특화 임베딩 모델을 (e.g., KLUE-BERT, KoLegalBERT 등) 활용함. • 상황에서 추출된 쟁점 키워드나 요지를 기반으로 벡터 검색함. 1) DB: 벡터화된 판례 요약문단들 (e.g. FAISS, Elasticsearch + SBERT) 2) 입력: 사용자 사건 요지 or keyword vector 3) 출력: 유사 판례 Top-K (요약 or 원문 문단) 3. RAG Augmentation – 판례 요약 및 문단 정제 • 검색된 판례들에서 핵심 문단(사건 개요, 판결 이유, 결과)를 추출함. • 너무 긴 경우 LLM 요약으로 축소하여 context length 를 최적화함. 4. LLM + Prompt Engineering – 대응 전략 생성 • 상황과 판례 문단이 주어졌을 때 1) 사건의 유사성 분석 2) 예상 벌금/처벌 강도 제시 3) 추천 행동 (ex. 합의/고소 여부, 증거 수집, 변호사 선임) 4) 판단 근거 제공

기대 효과	1. 활용 방안 또는 실현 가능성 • 일반 사용자에게 신속하고 정확한 법률 조언을 제공함. • 민원 처리, 법률 상담 사무소, 공공기관 콜센터 등에 적용 가능함. 2. 향후 발전 방향 또는 확장 아이디어 • 사건 유형별로 대응 로직을 세분화해 상황 맞춤형 법률 조언 시스템으로 고도화 가능함. • 유사 판례뿐 아니라 관련 법령 조항, 행정 지침 등으로 확장하여 다층적인 법률 정보를 제공함. • 상담 이력 데이터를 기반으로 사용자의 반복 질의와 관심 사건 유형을 분석해, 유사 사례 추천과 사전 안내 기능을 제공하는 개인화된 법률 조언 시스템으로 확장 가능함. • 실제 법률 전문가(변호사)와의 연계 기능을 추가해, 1차 AI 상담 → 2차 전문가 연결로 이어지는 통합 법률 지원 플랫폼 구축 가능함.
리스크 및 대응 방안	1. 잠재적 위험 요소 • 법률 데이터의 불균형/편향 • 개인 정보 포함 가능성 • 판례 응답 신뢰도 확보 • 판례 검색 속도 저하 2. 대응 전략 • 데이터 전처리 단계에서 클래스 균형을 조정하고 데이터를 증강함. • 판례 텍스트 정제 시 익명화 처리 알고리즘을 적용함. • 응답 내 출처를 표기하고, 판례 요약문, 인용문을 함께 제공함. • Lightweight Embedding 및 Approximate Nearest Neighbor를 사용함.
분석 타임라인	~ 06/27 : 프로토타입 설계(데이터 정제/ 벡터 임베딩/ 데이터베이스 구축/ 검색 로직) ~ 07/4 : 프로토타입 발전시켜 실제 구현 ~ 07/5 : 프로젝트 마무리 ~ 07/6 : 결과 분석 및 발표 자료 제작 ~ 07/7 : 최종 제출