# **부동산 계약 리스크 분석 챗봇 - 프로젝트 계획**

## 주요 진행 요소

1. **데이터 수집 및 전처리**
   * **주택임대차보호법:** 전체 조항 (시행령 포함)
   * **민법 선별 조항 (Whitelist): 제390조, 제615조, 제618조, 제623조, 제626조, 제627조, 제640조, 제750조**
   * **임대차보호법 관련 상담 사례, 대표 판례**

* 빈번히 분쟁이 되는 조항 성격 및 특약사항 관련해 추가 학습

1. **Chunking / Embedding / Keyword Dictionary**
   * **제N조 단위 청킹**
   * **임베딩**
   * **Pinecone Store - 청크별 메타데이터**
2. **LLM**
   * **모델 비교**
3. **Frontent - Django 웹 구현**
4. **선택 추가 요소**
   * **OCR, Vision AI, 멀티모달 (계약서 파일 입력)**
   * **상가건물임대차보호법 적용**

## 진행 방식

1. 전 팀원 전 단계 동시 진행
2. 중간 각 테스트 단계 데이터 보존 - 모델 비교군 등으로 활용
3. 각 일자 팀장 체크 및 단계별 최종본 채택

## 진행 일정

| **주차** | **일자** | **주요 활동 (Key Activities)** |
| --- | --- | --- |
|  | **Day 0** | 기획 확정 |
| **1주차** | **Day 1** | 법령/판례 데이터 수집 및 Semantic Chunking 수행 (JSON화) |
|  | **Day 2** | [AI] 임베딩 모델(OpenAI/BGE-m3) 성능 테스트 및 선정  Pinecone Free Tier 인덱스 생성 및 API 키 발급 |
|  | **Day 3** | Pinecone에 청킹된 데이터 Upsert (Vector DB 구축) |
|  | **Day 4** | LangChain을 활용하여 Django View 내 검색 로직 구현 (Retrieval)  채팅/결과 화면 UI (HTML/Bootstrap) 기초 작업 |
|  | **Day 5** | [Integration] MVP 통합: Django 폼 입력 -> Pinecone 검색 -> LLM 답변 -> 화면 출력 |
| **2주차** | **Day 6** | [Evaluation] RAGAS 활용 답변 정확도 평가 (특히 민법/특법 우선순위 확인)  프롬프트 엔지니어링 ("변호사 페르소나" 주입) |
|  | **Day 7** | UI 고도화 (로딩 스피너, 에러 메시지, 결과 하이라이팅)  [Feature] '표준 계약서 비교' 기능 로직 추가 |
|  | **Day 8** | [Opt] 검색 정확도 개선 (Hybrid Search 또는 Reranker 도입 검토)  다양한 악성 특약 케이스 테스트 |
|  | **Day 9** | 최종 버그 수정 및 배포 리허설 (로컬 시연 준비)  포트폴리오용 아키텍처 다이어그램 및 문서 작성 |
|  | **Day 10** | 포트폴리오 문서(README, 기술 블로그) 작성 및 프로젝트 종료 |

* Django 프로젝트 생성 (startproject), Git Repo 설정
* Django View - 템플릿 연동, 입력 폼(Form) 생성