# **부동산 AI 챗봇 <도와줘 홈즈냥즈> 개발 기획서**

## **1. 프로젝트 개요**

### **1.1 프로젝트명**

**도와줘 홈즈냥즈** - AI 기반 부동산 가디언 서비스

### **1.2 핵심 가치 제안**

* + **일반 사용자 대상**: 허위매물과 사기 위험을 방지하여 안전한 부동산 거래를 지원하는 **AI 기반 맞춤형 매물 추천 서비스**
  + **부동산 종사자 대상**: 단순 반복 업무(서류 분석, 정보 검색 등)을 AI가 대체함으로써 **업무 효율성 및 생산성 극대화**

### **1.3 프로젝트 배경**

**일반 사용자 대상 문제점**

* 허위매물로 인한 신뢰도 저하 (사용자 최대 불만사항)
* 전세사기 등 부동산 범죄에 대한 사회적 불안 증가
* 복잡한 거래 절차와 정보 비대칭성으로 인한 혼란
* 정보 과잉 및 파편화로 인한 탐색 시간 지연과 의사결정 어려움

**부동산 업계 대상 문제점**

* 반복적 업무로 인한 인건비 및 업무시간 증가
* 복잡한 업무 구조로 인한 비효율성

### **1.4 타겟 사용자**

**주요 타겟**:부동산을 구하는 일반인과 현업에 종사하는 부동산 중개인 모두를 아우르는 **양방향(Two-sided) 플랫폼**을 지향

* 생애 최초 임차인 (20대) - 애매해해
* 1인 여성 가구 (안전 중심)
* 생애 최초 주택 구매자 (30-40대)
* 외국인 거주자
* 공인중개사 및 부동산 종사자

### **1.5 비전**

"신뢰할 수 있는 AI 부동산 파트너로 모든 부동산 거래의 안전과 효율성을 보장한다"

* **부동산 허위매물 근절:** AI 검증 에이전트를 통해 매물 신뢰도를 상승
* **조기 사기 방지:** 거래 전 서류 분석 및 패턴 인식을 통해 사기 위험을 사전 경고
* **부동산 법률 정보 접근성 개선:** 법률 용어를 일반인이 이해하기 쉽게 풀어내어 정보 불균형 해소

## **2. 시장 분석**

### 

### **2.1 기존 부동산 앱 현황**

한국의 프롭테크 시장:

1세대(정보 포털) -> 2세대(모바일 플랫폼) -> 3세대(AI/빅데이터 기반)

* **빅4 플랫폼**: 직방, 다방, 호갱노노, 네이버부동산
* **시장 점유율**: 평균 단독 사용률 53.6% (높은 사용자 충성도)
* **주요 문제점**: 허위매물, 광고 중심 비즈니스 모델의 한계

### **2.2 부동산 AI 앱 동향**

* **부톡**: Geo-AI 기술 활용 전세사기 위험도 분석
* **부동산의 신**: GPT 기반 AI 리뷰 서비스
* **해외 사례**: Zillow의 EliseAI, Redfin의 Ask Redfin

### **2.3 사용자 PPN(Problem, Pain Point, Needs)**

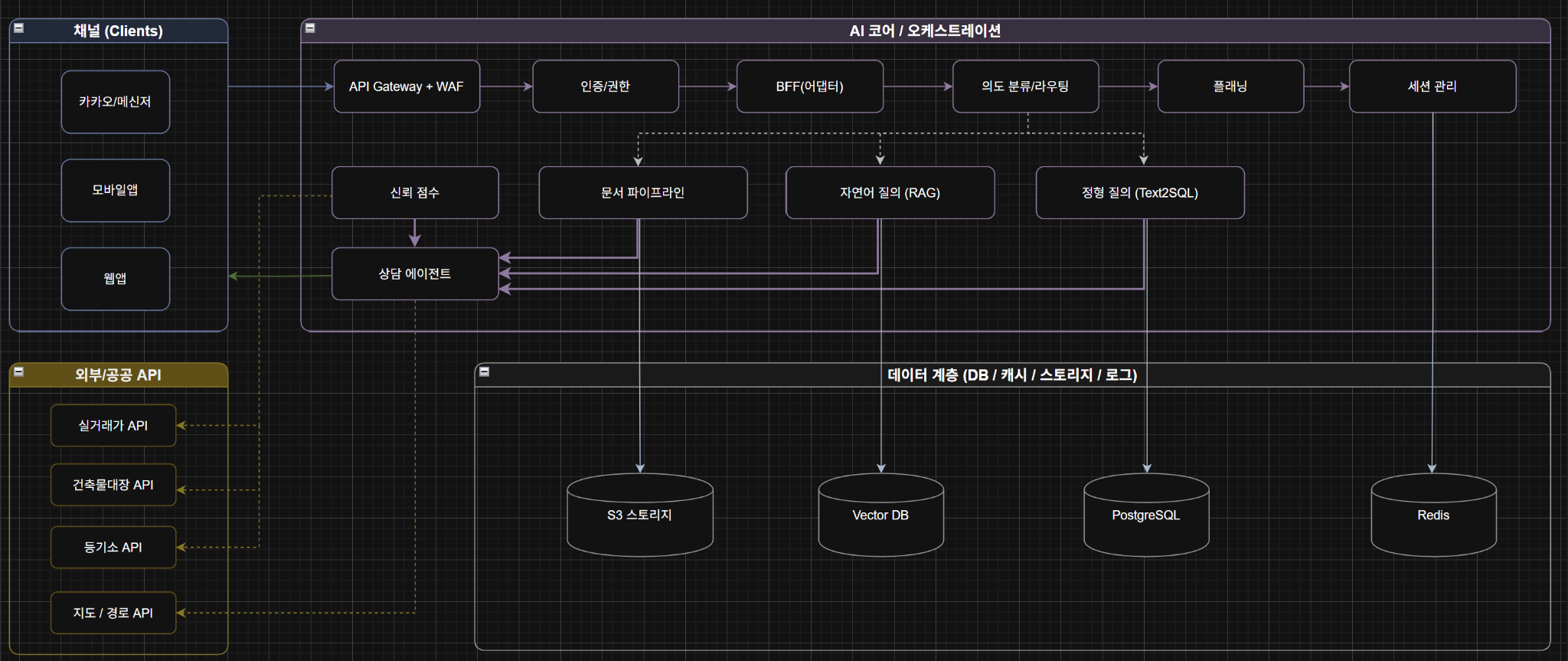
* **허위매물 문제**: 가장 심각한 사용자 불만
* **정보 과잉**: 방대한 데이터 속 통찰력 부족
* **거래 불안**: 전세사기, 계약서 이해 부족
* **접근성 문제**: 외국인, 초보자의 진입 장벽

**시장 기회**: 시장의 고질적인 문제(신뢰 부족, 정보 불균형)는 곧 **도와줘 홈즈냥즈**와 같은 신뢰 기반 AI 서비스의 중요한 기회. 특히, 허위매물에 대한 불안감, 복잡한 서류에 대한 이해 부족, 그리고 전세 사기 같은 사회적 문제들은 AI 기술로 해결

## **3. 앱 설계**

### **3.1 시스템 아키텍처**

다중 에이전트 시스템을 중심으로, 사용자의 질의가 들어오면 각 에이전트가 순차적으로 혹은 병렬적으로 데이터를 처리 → 통합하여 최종 답변을 생성



### **3.2 플로우차트**

이미지 첨부 예정

### 

### 

### **3.3 핵심 기능**

**1. 신뢰 점수(Trust Score) 시스템**

* 자동화된 권리분석 및 위험도 평가를 통한 매물 신뢰도 점수 제공

**2. 허위매물 탐지 기능**

* 공적 데이터(등기부등본, 건축물대장 등)와 매물 정보 교차 검증
* 데이터 불일치 시 허위매물 가능성 알림

**3. AI 기반 계약서 분석**

* 임대차 계약서, 등기부등본 등 비정형 서류 업로드 지원
* AI를 통한 핵심 내용 자동 요약 및 특이사항 추출
* 위험 요소 및 불리한 조항 자동 탐지

**4. 부동산 사기 예방 시스템**

* 기존 사기 패턴 학습을 통한 AI 모델 구축
* 거래 방식 및 서류에서 사기 징후 조기 탐지 및 경고

**5. 안전 중심 매물 검색 :** 1인 가구(특히 여성) 맞춤형 안전 필터링 기능

**6. 다국어 서비스:** 외국인 사용자를 위한 다국어 인터페이스 제공

### **3.4 AI 에이전트 시스템**

에이전트 단계: 검색 -> 분석/검증(병렬) -> 상담

| 구분 | 에이전트 기능 |
| --- | --- |
| 검색 에이전트 | * 부동산 매물 정보 검색 * 청약 정보 조회 * 등기부등본, 건축물대장 정보 수집 * 실거래가 데이터 조회 |
| 분석 에이전트 | * 계약서 조항 분석 및 위험요소 탐지 * 시장 동향 및 가격 적정성 분석 * 투자수익률(ROI) 계산 * 대출 한도 및 금리 시뮬레이션 * 정부 지원 정책 안내(ex. 공공주택, 신혼부부/초혼 전세금 지원 등) |
| 검증 에이전트  (머신러닝 기반) | * 허위매물 패턴 인식 * 공적 장부 교차 검증 * 전세사기 위험도 평가 * 임대인 신용도 체크 |
| 상담 에이전트 | * 거래 절차 단계별 가이드 * 계약 특약사항 추천 * 분쟁 해결 방안 제시 * 맞춤형 매물 추천 * 정부 지원 정책 안내 |

### **3.5 챗봇 구현 핵심 기술**

**3.5.1 모델 튜닝**

* LLM (Large Language Model) 튜닝 -> sLLM
* 임베딩 모델 최적화
* 리랭커(Re-ranker) 모델 학습

**3.5.2 주요 기술 구현**

* **하이브리드 서치**: 정형DB+백터DB사용 / 벡터검색시 키워드+유사도기반 검색
* **서류 자동생성**: 계약서, 신청서 등 부동산 관련 문서 템플릿 자동 생성
* **Text2SQL**: 자연어 질의를 데이터베이스 쿼리로 자동 변환
* **RLHF**: 사용자 피드백 기반 지속적 모델 성능 개선

## **4. 기술 스택**

### **4.1 개발 환경**

* **Language**: Python 3.12
* **Orchestration**: LangGraph 0.6+
* **LLM**:
  + Primary: GPT-4o
  + Secondary: sLLM(specialized Large Language Model)
* **패키지 관리**: UV

### **4.2 인프라**

* **Frontend**: React 18 + TypeScript
  + **Mobile**: React Native (우선순위 낮음)
* **Backend**: FastAPI 0.100+
* **Vector DB**: ChromaDB / Milvus 등
* **Relational DB**: PostgreSQL 15
* **Document Store**: MongoDB (비정형 문서 저장)

### **4.3 AI 모델**

* **Embedding**: KURE-v1 (한국어 특화 모델)
* **Reranker**: BGE-reranker-ko (추가 검토 필요)
* **OCR**: Tesseract + EasyOCR (등기부등본 처리, 우선순위 낮음)

## **5. 데이터베이스 구축**

### **5.1 벡터 DB (ChromaDB)**

* 부동산 관련 법령 및 판례
* 부동산 용어사전
* FAQ 및 상담 사례
* 청약 정책 정보

### **5.2 정형 DB (PostgreSQL)**

* 지역별 생활 정보 (아파트 단지, 학군, 교통, 편의시설)
* 역세권 및 교통 접근성 데이터
* 실거래가 이력
* 청약 통계 및 당첨 정보 ( 벡터DB, 하이브리드 검색 )
* 제 1 금융권 대출정보 ( 벡터DB, 하이브리드 검색 )
* 각종 혜택 ( 생애첫내집마련 등 )

### **5.3 비정형 DB (MongoDB)**

* 임대차계약서 템플릿
* 등기부등본 스캔 문서
* 건축물대장
* 사용자 상담 히스토리

## **6. 데이터 수집**

**수집 지역**: 강남권역(강남구, 서초구, 송파구) 또는 영등포권역()

### **6.1 공공 API 활용**

* 국토교통부 실거래가 API - <https://www.data.go.kr/data/15126469/openapi.do>
* 건축물대장 API (건축HUB) - <https://www.data.go.kr/data/15134735/openapi.do>
* 대법원 등기정보 API - <https://data.iros.go.kr/rp/oa/openOapiIntro.do?ris_menu_seq=0000000008>
* 한국부동산원 공시가격 API - <https://www.reb.or.kr/r-one/portal/openapi/openApiIntroPage.do>

### **6.2 크롤링 대상**

* 주요 부동산 플랫폼 매물 정보 - <https://land.naver.com/>, <https://www.zigbang.com/>, <https://hogangnono.com/>
* 부동산 관련 뉴스 및 정책 - <https://news.naver.com/breakingnews/section/101/260>, <https://www.molit.go.kr/USR/NEWS/m_71/lst.jsp?search_section=p_sec_2>
* 지역별 생활 정보 - <https://map.naver.com/p>, 지자체/구청 홈페이지

## **7. RLHF 시스템(사용자 피드백 기반 학습)**

### **7.1 허위매물 탐지 RLHF**

* 사용자 신고 데이터 수집
* 허위매물 패턴 학습 및 개선
* 탐지 정확도 지속적 향상

### **7.2 상담 품질 RLHF**

* 답변 만족도 평가 수집
* 부정확한 답변 수정 피드백 반영
* 사용자 선호 답변 스타일 학습

## **8. 구현 일정**

9/11(목) 오프라인 회의에서 논의 예정

## **9. 역할 분담**

### **9.1 결정권자 (팀장)**

* **DB 팀장**: 김승학
* **LangGraph 팀장**: 김도윤
* **sLLM 팀장**: 황차해
* **ML 팀장**: 이정민

### **9.2 개발 분담**

1. **데이터 수집**: 도메인별 분할 크롤링
2. **데이터 정제**: 수집 완료 후 팀 협의를 통한 정제 방안 및 형식 결정
3. **LangGraph**: 팀 협의를 통해 설계 후 구현 (도윤 팀장 주도)
4. **sLLM 개발**: 팀원별 개별 모델 담당