

실시간 대화형 AI 보드게임 도우미

25-2 서강 융합기술 경진대회 | 지도교수: 장두성 · 팀원: 신동준 김동건 이동주 조원빈

프로젝트 배경

문제1: 룰북을 찾아보는 번거로움

게임 중 의견이 엇갈리는 상황에 룰북을 찾아보는 데에 많은 시간 소요

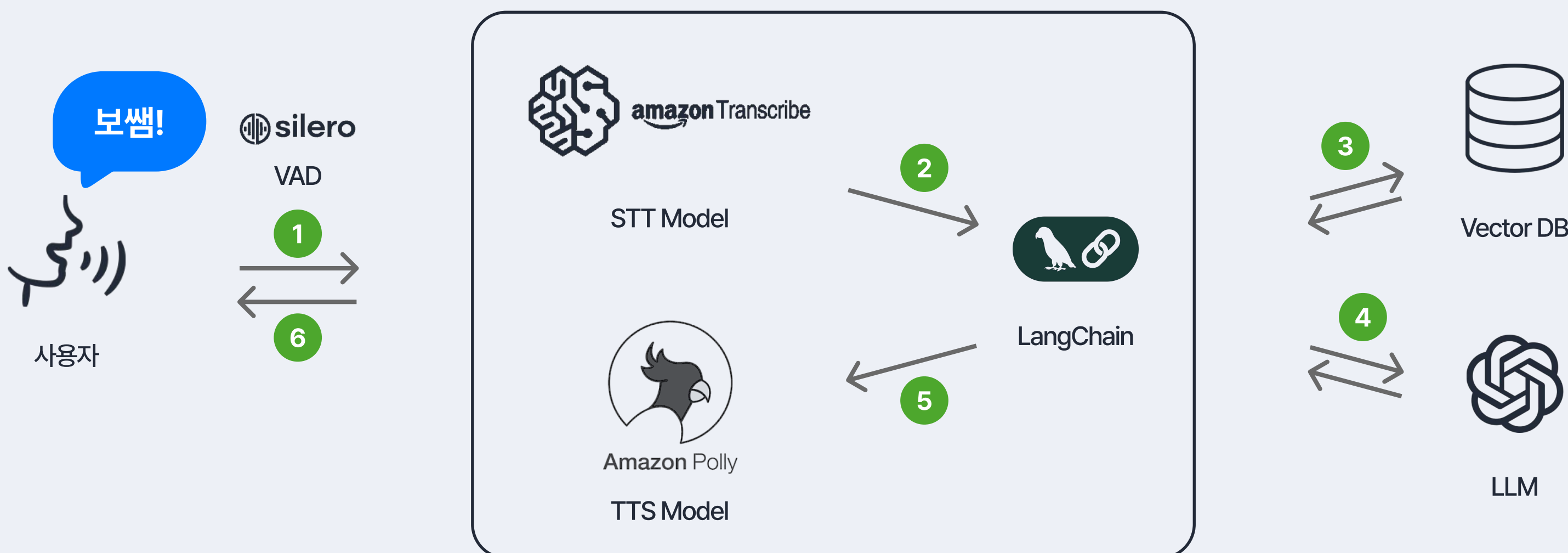
→ 음성을 통한 간편한 질문과 답변

문제2: 기존 LLM의 환각

GPT, Gemini 등 기존 LLM의 부정확한 답변 (Hallucination)

→ 룰북에 근거한 신뢰성 있는 답변

주요 기술



- 실시간 음성 활동 감지
- 음성 데이터 → 텍스트 데이터
- 임베딩된 질문으로 답변 검색
- 검색된 데이터 기반 답변 생성
- 텍스트 데이터 → 음성 데이터
- 실시간 답변 제공

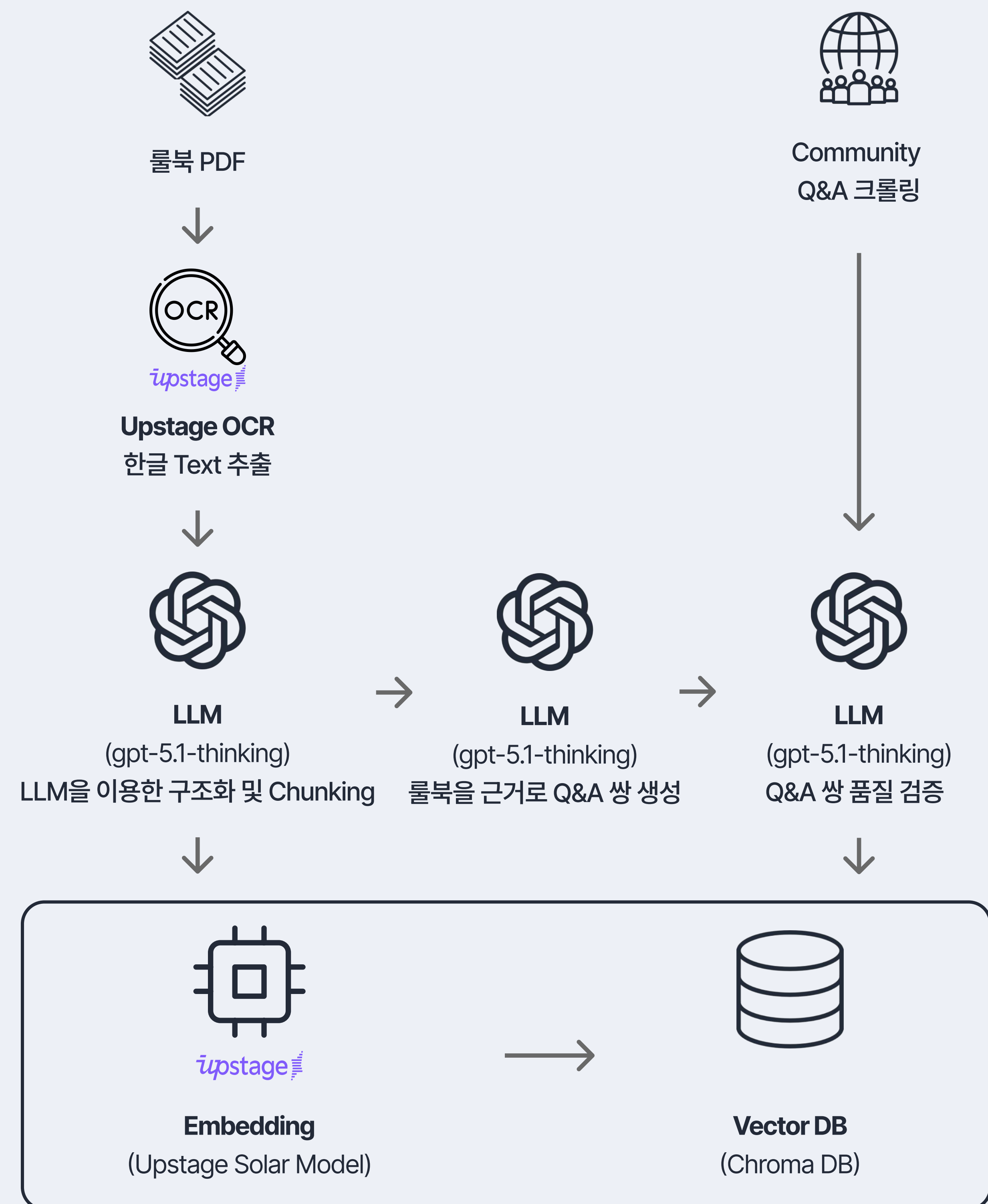
RAG(Retrieval-Augmented Generation)

기존 LLM은 학습 데이터에만 의존하기 때문에 환각이 발생할 수 있다.

이를 해결하기 위해 전처리된 룰북 데이터를 기반으로 답변을 생성하는 RAG 시스템을 구축하였다.

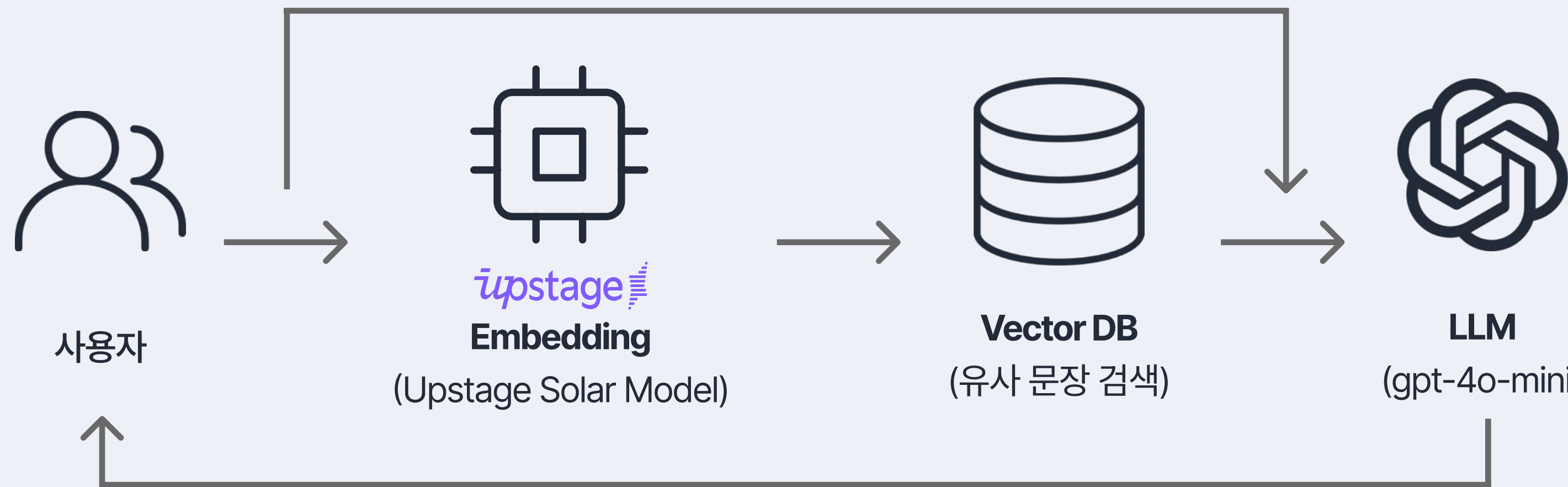
1. 전처리 및 데이터 구축

- 공식 룰북 구조화
 - 사진 기반 룰북은 OCR을 통해 텍스트로 변환
 - LLM을 이용하여 불필요한 정보를 제거하고, 섹션 단위로 재구성
 - 각 섹션은 필요시 섹션 내에서 추가 Chunking 후, Embedding하여 Vector DB에 삽입
- Semantic Gap 보정
 - 사용자 질문과 공식 룰북 간 표현의 차이로 Semantic Gap 발생
 - 보드게임 커뮤니티에서 실제 Q&A를 수집하고, 공식 룰북에 기반한 추가 Q&A 데이터(100~500쌍)를 LLM으로 생성
 - 고성능 모델을 통해 검증된 Q&A 쌍만 Vector DB에 삽입



2. 질의 처리 및 응답 생성

- 유사도 검색
 - 사용자의 질의를 임베딩하여 Vector DB에서 상위 후보를 검색
 - 검색 결과에는 공식 룰북 Chunk와 Q&A 쌍이 함께 포함됨
- Q&A 후보수에 따른 답변 구성 전략
 - 검색된 Q&A 후보가 충분한 경우, 이를 기반으로 LLM이 답변을 구성
 - 검색된 Q&A 후보가 불충분한 경우, 룰북 원문을 Context로 제공하고 LLM이 이를 기반으로 답변을 구성
- 신뢰도 기준 미달 시 답변 제한
 - 유사도 점수가 임계값에 미치지 못해 의미 있는 후보를 찾지 못한 경우, "해당 질문에 대한 정보를 찾을 수 없음"으로 답변하여 환각을 방지



STT/TTS

- VAD - Silero VAD
 - 소음이 많은 보드게임 카페에서도 정확한 발화 구간 감지
- STT - AWS Transcribe
 - 클라우드 기반 아키텍처로 모바일 클라이언트의 연산 부하 최소화
- TTS - AWS Polly
 - 동적인 텍스트에 대하여 SSML로 더욱 자연스러운 음성 구현

프로젝트 결과물



기대효과 및 활용방안

비즈니스 모델

- 앱 내 광고 서비스
- 보드게임 카페 대상 B2B 확장

추후 계획

- 고성능 추론 답변 기능 제공
- 질의 데이터를 RAG 학습 데이터로 축적하여 모델 정교화