

RAG 파이프라인 구축 프로세스 보고서 양식

<p>서비스 명 및 개요</p>	<p>서비스명: 복약 및 마음 상담 챗봇</p> <p>서비스 개요:</p> <p>일반 사용자들은 약물을 복용할 때 복용법, 식사와의 관계, 병용 금기 여부 등을 알기 어려워 부작용을 겪거나, 잘못된 복용으로 건강에 위험을 초래할 수 있습니다. 기존에는 이를 해결하기 위해서 의료기관을 방문하거나 인터넷에 전문적인 검색을 해야 했으나, 이 과정은 번거롭거나, 신뢰성의 문제가 발생하기도 합니다.</p> <p>이러한 문제를 해결하기 위해 공공 데이터를 활용해 사용자의 질문에 맞는 복약 정보를 정확하게 제공하는 복약 도우미 챗봇 서비스를 개발합니다. 사용자가 약품명, 성분명, 제조사 등을 입력하면, 해당 약물에 대한 복용 방법, 다른 약물과의 병용 가능 여부 등 필수적인 복약 정보를 손쉽게 확인할 수 있습니다. 챗봇은 공공 데이터를 기반으로 신뢰성 높은 정보를 제공하여, 사용자가 빠르고 손쉽게 정확한 정보를 얻을 수 있도록 돕습니다.</p> <p>더해 마음 상담 챗봇은 기본 LLM 모델에 시스템 프롬프팅을 더해 사용자의 마음을 위로해주기 위한 용도로 사용됩니다.</p> <p>이 서비스를 통해 사용자는 복약에 관한 다양한 질문을 빠르게 해결하고, 건강을 지킬 수 있는 올바른 복약 습관을 유지할 수 있습니다. 또한, 심리적 안정이 필요한 사용자에게는 마음 상담 챗봇을 통해, 신체적 건강과 함께 정신적 건강도 챙길 수 있습니다.</p>
<p>타겟 사용자 및 시장 분석</p>	<p>타겟 사용자:</p> <ul style="list-style-type: none"> - 일반 사용자: 만성질환자, 노년층, 임산부 등 - 병원 방문이 어려운 상태에서 복약 정보를 급히 확인해야하는 사용자 - 의약품 취급 주의사항을 꼼꼼하게 확인하고 복약을 희망하는 합리적 소비자 <p>주요 요구사항 및 사용 목적:</p> <ul style="list-style-type: none"> - 사용자가 입력한 약물에 대한 복용법, 병용 금기 정보 제공 - 약물의 이름 또는 영양제 정보로 정확한 복약 가이드 제공 - 공공 데이터를 활용하여 신뢰성 높은 정보 제공 - 의료사고 및 부작용 예방을 위한 빠르고 정확한 정보 제공 - 홀로 생활하며 충분한 케어를 받지 못한 이들에게 마음의 위로와 안정 제공 <p>시장 분석:</p> <p>이 서비스는 단순 검색 챗봇이 아닌, 공공 API 기반의 정확한 데이터를 LangChain 과 LLM 을 통해 실시간으로 제공하는 구조입니다. 사용자는 챗봇을 통해 병용 가능 여부와 복약 정보를 직관적으로 얻을 수 있으며, 실시간 질의응답과 병용 정보 제공 기능은 기존 복약 안내 앱과 차별화된 경쟁력을 가집니다.</p>

<p>목표 및 기대효과</p>	<p>서비스 목표:</p> <ul style="list-style-type: none"> - 사용자 질문에 신속하고 정확한 복약 정보를 제공 - 약물 상호작용을 예방하여 사용자 건강을 보호 - 대화를 통해 심리적 지지를 제공하여 사용자의 마음 안정을 도모 <p>기대효과:</p> <ul style="list-style-type: none"> - 사용자는 복약 시 주의사항을 이해하고 적절한 복용을 실천할 수 있습니다. - 사용자는 투병으로 지친 마음을 털어놓으며 심리적 안정감을 얻을 수 있습니다. - 약물 부작용 예방을 통해 국민 건강 수준을 향상시킬 수 있습니다. - 고객 서비스팀이나 약사들은 반복 문의를 줄이고 효율적인 대응이 가능합니다.
<p>데이터 구성 및 활용</p>	<p>원천데이터 소스:</p> <ul style="list-style-type: none"> - 건강보험심사평가원 DUR 데이터 (공공데이터포털) <ul style="list-style-type: none"> ↳ 의약품안전사용서비스(DUR) 의약품 목록 - 사용 파일 목록: <ul style="list-style-type: none"> ↳ 병용금기 / 임부금기 / 연령금기 / 노인주의 등 주의사항 파일 <p>원천데이터 형식 :</p> <ul style="list-style-type: none"> - CSV (공공데이터포털 제공) <p>데이터 처리 방법:</p> <ul style="list-style-type: none"> - DUR CSV 파일별 필요한 컬럼만 추출 - 성분명, 제품명, 약품상세정보, 급여여부, 금기사항 등 추출 - 카테고리별로 분리하여 문장화 후 새로운 csv 파일 생성 - 예시 문장: <p>“임신 중 명문제약(주)에서 제조한 날페인주사 10 밀리그램(날부핀염산염)_(10mg/1mL)(nalbuphine hydrochloride 10mg(10mg/mL), 급여구분: 급여)는 복용 금지입니다. 상세: 임부에 대한 안전성 미확립. 분만시 신생아의 호흡억제 가능성.”</p>

<p>RAG 파이프라인 설계</p>	<p>데이터 최적화:</p> <ul style="list-style-type: none">• Chunk Size: 1000• Overlap: 100 <p>벡터 데이터베이스 구축 및 임베딩:</p> <ul style="list-style-type: none">• 벡터 DB : Pinecone• 임베딩 모델 : Upstage Solar Embedding <p>Retriever 및 Reranker 구현:</p> <ul style="list-style-type: none">• VectorStore Retriever 이용-하이퍼파라미터 튜닝 : -search_type="mmr"(default : similarity(유사도) / mmr 알고리즘), k=3 <p>LLM 프롬프트 설계 및 답변 생성, 평가 :</p> <p>1/ Task : QA / 챗봇</p> <p>2/ 프롬프트 예시</p> <p>[답변 지침]</p> <p>당신은 전문 복약도우미입니다. 사용자 질문에 대해 기존 데이터를 우선 참조하여 안전한 복약 정보를 제공하세요.</p> <p>규칙:</p> <ol style="list-style-type: none">1. 약물 상호작용 등 위험이 있으면 반드시 '경고'라고 명확히 표기하세요.2. 불확실한 경우가 있으면 '가' 사실을 반드시 명시하고, 전문의 상담을 권고하세요.3. 응급상황이 의심될 경우 즉시 병원 방문을 안내하세요.4. 모든 약물 반응에는 개인차가 있음을 항상 언급하세요. <p>답변 스타일:</p> <ul style="list-style-type: none">- 객관적이고 신뢰성 있는 의료 정보 제공- 제품명, 성분, 복용 주의사항 등은 구체적으로 명시- 사용자 질문에 대한 핵심 정보 제공 <p>[질문]</p> <p>{question}</p> <p>[참고 정보]</p> <p>{context}</p> <p>위 지침에 따라, 위 참고 정보를 바탕으로 사용자에게 정확하고 신뢰도 높은 답변을 2~4문장으로 설명해 주세요.</p> <p>3/ 답변 생성 : Upstage Solar Pro</p>
<p>RAG 파이프라인 평가 및 결과</p>	<p>평가방법</p> <ul style="list-style-type: none">• 정량 평가 : RAGAS 평가 지표<ul style="list-style-type: none">- 키워드 매칭 점수: 질문에 포함된 주요 키워드와 문서 내 일치 정도- 응답 품질 점수: 문법, 구조, 맥락에 기반한 응답 일관성 평가- 메타데이터 일관성: 질문과 응답이 동일한 문서에서 출처되었는지 여부• 정성 평가<ul style="list-style-type: none">- 산출물 방식: 실제 DUR 의약품 문서에서 추출한 금기/주의 사항을 기준으로 평가 <p>기준:</p> <p>정확성: 생성된 응답이 실제 DUR 정보와 얼마나 일치하는가?</p> <p>신뢰성: 출처 문서를 기반으로 명확한 근거를 제시하는가?</p> <p>일관성: 질문-응답 간 논리적 흐름과 의료적 표현의 정확성</p> <p>평가 결과</p> <ul style="list-style-type: none">• 정량 평가 (값은 소수점 셋째자리에서 반올림)<ul style="list-style-type: none">- 맥락 정밀도(기본) : 0.73

	<ul style="list-style-type: none"> - 맥락 재현율(기본) : 0.76 - Dense Retriever : 위와 같음 • 정성 평가 <ul style="list-style-type: none"> -5 개 질문 중 5 개 모두 DUR 문서를 기반으로 정확하게 응답 -약물 추천 질문에 대해서는 약물 추천을 지양하고 의사와의 상담을 권장 -사용자 제공 정보를 바탕으로 조건에 맞는 맞춤형 답변을 제공, 의약품 외에도 음식 등 건강에 유익한 부가 정보를 잘 전달 -따뜻한 말투로 사용자의 고민을 경청하며 응답 -전반적으로 구조화된 답변을 제공하며, 주요 복약 정보를 명확하고 간결하게 요약
결론 및 향후 발전 방향	<p>결론</p> <ul style="list-style-type: none"> - DUR 기반 데이터를 벡터 DB 에 저장하고, 이를 기반으로 RAG 파이프라인을 구성하여 약물 복용 주의사항, 병용금기, 임부금기 여부 등을 빠르게 질의-응답할 수 있는 시스템을 구현함 - 질문 의도와 복약정보가 정확히 매칭되도록 임베딩과 프롬프트 설계가 중요했으며, 대부분의 응답이 문서 기반으로 신뢰할 수 있는 결과를 도출함 <p>향후 발전 방향</p> <ul style="list-style-type: none"> -의약품 전체 리스트 DB 제공으로 사용자 선택 편의성 향상 -의약품 사진 검색 기능 추가 -멀티턴 대화 기능으로 사용자 만족도 증대 -맞춤형 복약 캘린더 서비스 제공