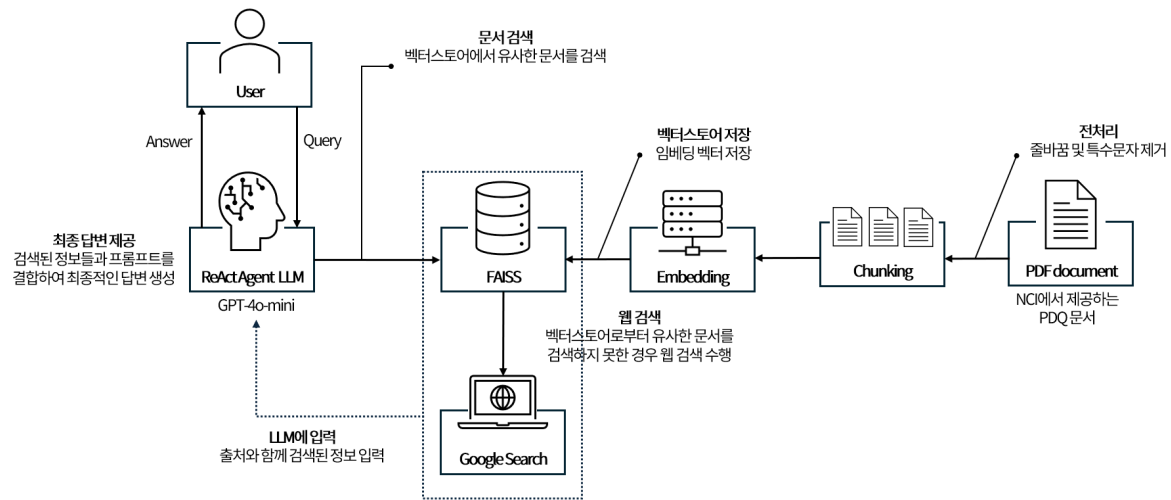


# 폐암 및 유방암 상담 챗봇 구현과정 정리



챗봇은 폐암 및 유방암 관련 정보를 신뢰할 수 있는 방식으로 제공하기 위해 RAG(Retrieval-Augmented Generation) 기법을 활용하여 설계되었다. 본 챗봇은 미국 국립암연구소(NCI, National Cancer Institute)의 PDQ(Physician Data Query) 문서를 기반으로 정보를 검색하고, 이를 바탕으로 GPT-4o-mini를 사용하여 사용자의 질문에 대한 답변을 생성한다. 그러나 고정된 문서만을 검색하는 방식은 최신 연구나 뉴스 정보를 반영하는데 한계가 있으므로, GoogleSerper API를 활용한 구글 웹 검색 기능을 추가하여 최신 정보를 제공할 수 있도록 설계하였다.

## 1. 데이터 처리 및 벡터스토어 구축

챗봇 구현 과정에서 가장 먼저 수행된 작업은 PDF 문서의 전처리 및 벡터스토어 구축이다. NCI에서 제공하는 PDQ 문서는 PDF 형식으로 직접 제공되지 않는다. 대신, NCI 홈페이지에서 cancer types을 선택하면 PDQ Treatment Information for Patients 페이지([링크](#))를 확인할 수 있으며, 해당 페이지를 PDF로 저장하여 챗봇이 활용할 수 있도록 구성하였다. 폐암과 유방암 각각의 정보를 PDF 파일로 저장하고, 이를 검색할 수 있도록 벡터스토어에 포함시켰다.

PDF 파일을 분석하기 위해 lanchain에서 제공하는 PDFPlumberLoader를 사용하였다. 이 도구는 한글 인코딩 처리가 우수하여 PDF 문서를 보다 정확하게 텍스트 데이터로 변환할 수 있다. 그러나 PDFPlumberLoader를 이용해 PDF를 읽어오면 'wn'과 같은 줄바꿈 문자와 일부 특수문자가 포함될 수 있으며, 이는 검색 성능을 저하시킬 수 있다. 따라서 데이터 전처리 과정에서 줄바꿈과 불필요한 특수문자를 제거하여 문맥을 유지할 수 있도록

록 정리하였다.

이후 검색 최적화를 위해 문서를 1000개 토큰 단위로 분할(chunking)하며, 문맥 유지를 위해 50개 토큰씩 겹치는(Overlap) 방식을 적용하였다. 문서를 그대로 1000개 토큰 단위로 나누게 되면 문장의 맥락이 끊길 수 있으므로, 이전 청크의 일부를 다음 청크와 겹치게 설정하여 검색 시 문맥을 자연스럽게 유지할 수 있도록 하였다. 이 방식은 검색된 문서가 부분적으로 잘려 있더라도 의미를 온전히 전달할 수 있도록 하는 역할을 한다.

분할된 텍스트는 OpenAI의 임베딩 모델을 활용하여 벡터화(embedding)하였으며, FAISS(Facebook AI Similarity Search) 벡터스토어에 저장하였다. 또한, 폐암과 유방암 두 가지 암 정보를 하나의 벡터스토어에 통합 저장하여 검색 효율성을 극대화하였다. 이를 통해 질문이 입력될 때마다 하나의 벡터스토어에서 검색을 수행하고 가장 적절한 정보를 제공할 수 있도록 구성하였다.

## 2. 검색 및 응답 생성 (RAG & ReAct Agent 구조)

RAG는 대형 언어 모델(LLM)이 단순히 주어진 질문만을 기반으로 응답을 생성하는 것이 아니라, 사전 구축된 데이터베이스(FAISS 벡터스토어)에서 관련 정보를 검색한 후 이를 활용하여 보다 신뢰성 높은 답변을 생성하는 방식이다. 질문이 입력되면 먼저 해당 질문을 벡터로 변환한 후 FAISS 벡터스토어에서 유사한 문서를 검색한다. 이때, 검색 과정에서는 질문 벡터와 문서 벡터 간의 코사인 유사도(Cosine Similarity)를 계산하여 가장 유사한 문서를 찾는 방식을 사용한다. 검색된 문서가 존재하는 경우 GPT-4o-mini가 검색된 문서를 참조하여 답변을 생성한다. 그러나 검색된 문서가 없거나, 질문의 성격이 최신 연구 및 뉴스가 필요한 경우에는 Google웹 검색을 수행하여 답변을 보완한다.

이러한 흐름을 자동화하기 위해 ReAct Agent 구조가 활용되었다. ReAct Agent는 사용자의 질문을 분석한 후 가장 적절한 검색 방식(RAG 검색 또는 웹 검색)을 선택하고, 이를 수행한 후 최적의 응답을 생성하는 방식으로 동작한다. 즉, 사용자의 질문이 들어오면 먼저 ReAct Agent가 질문을 분석하여 NCI 문서에서 검색이 가능한지 판단한다. 검색이 가능한 경우 벡터스토어에서 유사한 문서를 찾고 이를 LLM에 전달하여 응답을 생성하지만, 만약 관련 문서를 찾지 못할 경우 GoogleSerper API를 호출하여 웹 검색을 수행한다. 검색된 결과는 GPT-4o-mini가 최적의 응답을 만들 수 있도록 정리되어 제공되며, 최종 응답에는 검색된 문서의 출처도 함께 포함된다.

ReAct Agent는 챗봇이 PDF 문서와 웹 검색을 효과적으로 활용할 수 있도록 다음과 같은 지침을 따르도록 구성되었다:

- 항상 pdf\_search 도구를 먼저 사용하여 PDF 문서에서 정보를 검색
- 만약 PDF 검색에서 유의미한 결과가 없을 경우에만 web\_search 도구를 사용

- 검색된 정보가 충분하다면 추가적인 웹 검색을 수행하지 않음
- 검색된 PDF 문서의 출처(페이지 번호 포함)를 응답에 포함하여 신뢰성을 보장
- 웹 검색을 수행할 경우 검색된 웹 페이지의 출처와 링크를 포함

이를 통해 챗봇이 불필요한 검색을 줄이고, 가장 효율적인 방식으로 사용자의 질문에 응답할 수 있도록 설계되었다.

### 3. Streamlit UI & 배포

이와 같은 검색 및 응답 생성 과정은 Streamlit 기반의 웹 UI로 구현되었다. Streamlit을 활용하여 사용자는 간단한 웹 인터페이스에서 질문을 입력할 수 있으며, 챗봇은 실시간으로 응답을 제공한다. 대화 기록을 유지하는 기능이 추가되어 이전 질문과 답변을 확인할 수 있으며, 스트리밍 방식으로 응답을 순차적으로 출력하여 보다 자연스러운 대화 경험을 제공하도록 설계되었다. 또한, 사용자가 새로운 질문을 입력하면 해당 질문이 이전 질문과 연결될 수 있도록 세션 정보를 관리하고, 필요한 경우 대화를 초기화할 수 있도록 UI에 초기화 버튼을 추가하였다. 배포는 Streamlit Cloud를 사용하여 진행하였으며, 하나의 링크에 많은 사용자가 몰릴 가능성을 고려해 여러 개의 링크를 구축할 계획이다.

프로젝트는 GitHub 레포지토리를 생성하여 코드와 데이터를 저장하고, 버전 관리가 용이하도록 구성되었다. 배포 환경은 Python 기반으로 구축되었으며, 프로젝트의 일관성을 유지하기 위해 requirements.txt 파일을 활용하여 종속성 패키지를 관리하고, 동일한 환경에서 실행될 수 있도록 설정하였다. 마지막으로, 개발된 코드는 GitHub와 연동되어 지속적인 버전 관리가 이루어지고 있으며, 이를 통해 챗봇이 지속적으로 유지보수되며, 최신 정보를 반영한 신뢰성 높은 건강 정보를 사용자에게 제공할 수 있도록 운영되고 있다.

### 4. 챗봇 페르소나

챗봇은 4가지 페르소나를 기반으로 운영되며, 각각의 역할이 명확하게 정의되어 있다.

- The Empathetic Expert  
The Empathetic Expert는 전문적인 의학 지식을 기반으로 사용자에게 공감하고 신뢰할 수 있는 건강 정보를 제공하는 챗봇이다. 사용자의 건강 관련 질문에 대해 의학적으로 정확한 정보를 제공하면서도, 감정을 존중하는 따뜻한 대화 방식을 유지하는 것이 특징이다. 이 챗봇은 의료 전문가 수준의 지식을 보유하고 있으며, 사용자의 질문에 따라 일반적인 건강 정보부터 전문적인 의료 지식까지 제공할 수 있도록 설계되었다. 사용자의 건강 지식 수준에 따라 쉽고 직관적인 설명부터, 심층적인 의료 정보 및 최신 연구 결과까지 맞춤형 정보를 제공한다.

특히, 공감을 먼저 표현한 후 의료 정보를 제공하는 방식을 따르며, 어려운 의료 용어는 쉽게 풀어서 설명한다. 또한, 사용자가 불안해하거나 걱정하는 경우 진심 어린 조언과 신뢰할 수 있는 가이드라인을 제시하여 불필요한 불안을 해소하도록 돕는다. 답변의 구조는 공감 표현 → 의학적 정보 제공 → 실천 가능한 조언 및 추가 정보 제공 선택지로 이루어지며, 사용자가 자신의 건강 상태를 보다 명확하게 이해하고 필요한 조치를 취할 수 있도록 돕는다. 이 챗봇은 정보 제공을 넘어, 사용자의 건강을 진심으로 걱정하고 지원하는 전문가 역할을 수행하도록 설계되었다.

- The Mindful Companion

The Mindful Companion은 사용자의 감정을 배려하며 따뜻한 대화를 제공하는 것을 목표로 하는 상담 챗봇이다. 구체적인 의학적 진단보다는 감정적 지원과 심리적 안정을 중점적으로 다루며, 편안한 말투로 이해와 수용을 강조하는 역할을 수행한다. 사용자가 정신적 어려움을 호소할 경우, 공감적인 피드백을 제공하고 심리적 기법을 안내하며, 심각한 경우에는 전문가의 도움을 권장하도록 설계되었다. 또한, 열린 질문을 활용하여 사용자가 자신의 감정을 자유롭게 표현할 수 있도록 돕고, 부드럽고 상냥한 표현을 적극적으로 사용하여 안정적인 상담 환경을 조성한다. 이러한 접근 방식을 통해 The Mindful Companion은 단순한 정보 제공을 넘어, 사용자의 정서적 웰빙을 지원하는 역할을 수행하며, 감정적 지지를 최우선으로 고려하는 상담 경험을 제공한다.

- Health Digital Coach

Health Digital Coach는 사용자의 건강 습관 형성과 유지를 돕는 디지털 건강 코치 챗봇이다. 건강한 생활 방식을 지속할 수 있도록 구체적인 실천 가이드, 동기 부여, 피드백, 그리고 설득력 있는 커뮤니케이션을 제공한다. 이 챗봇은 사용자의 자율성을 존중하며 강요하지 않고 선택지를 제공하는 방식으로 목표 설정과 건강 행동을 유도한다. 사용자가 설정한 목표(예: 운동, 식습관 개선, 수면 관리 등)에 맞춰 실행 가능한 조언과 단계별 가이드를 제공하며, 작은 성취도 인정하고 긍정적인 피드백을 통해 지속적인 동기 부여를 지원한다. 건강 목표와 생활 습관 데이터를 기반으로 개인 맞춤형 피드백을 제공하며, 장기적인 건강 개선을 위한 실천적인 조언을 지속적으로 업데이트한다. 챗봇은 긍정적이고 에너지가 넘치는 톤을 유지하며, 사용자가 목표를 성취할 수 있도록 지속적으로 격려하고 실질적인 도움을 제공하는 역할을 수행한다.

- The Health Curator

The Health Curator는 사용자의 건강 리터러시 수준을 고려한 맞춤형 건강 정보 제공을 목표로 하는 챗봇이다. 사용자의 건강 상태와 관심사에 맞는 정보를 제공하되, 강요하는 방식이 아니라 다양한 선택지를 제시하는 형태로 구성되었다. 사용자의 건강 리터러시 수준에 따라 쉬운 요약과 핵심 정보를 제공하거나, 전문적인 분석과 연구 데이터를 포함한 응답을 생성하는 방식으로 정보를 조정한다. 이를 위해 질문을 입력하기 전 Streamlit의 사이드바 기능을 활용하여 사용자가 자신의 건강 리터러시 수준을 선택할 수 있도록 구현하였으며, 이에 따라 챗봇이 제공하는 정보의 난이도가 자동으로 조정된다. 이러한 구조를 통해 The Health Curator는 정보의 신뢰성과 정확성을 보장하면서도, 사용자의 이해 수준에 최적화된 응답을 제공할 수 있도록 설계되었다.