

# 오픈소스프로젝트 중간보고서

학생 팀별 작성용

과제 수행원 현황						
수행 학기	■ 2025년 9월~2025년 12월					
프로젝트명	■ 동국대학교 재학생 맞춤형 정보 제공 챗봇 시스템 '동독이'					
팀명	■ Renux					
	학과	학번	성명	성별	연락처	E-mail
팀장	경찰행정학부	2020111242	조준용	남	010-4510-9921	junyong4510@gmail.com
팀원	통계학과	2021110445	신원철	남	010-8885-0172	swc3466@naver.com
	통계학과	2021110473	육심호	남	010-5041-7743	6aimho@gmail.com
지도교수	교과목명	■ 오픈소스소프트웨어프로젝트				
	소속	■ 융합소프트웨어 연계전공				
	성명	■ 이길섭 교수				

프로젝트	
프로젝트 개요	<p><b>1. 프로젝트 개요</b></p> <p>본 프로젝트 "동독이"는 "동국대학교 똑똑이"의 줄임말로, 동국대학교의 내규, 학과 별 정보, 부서별 정보 등 학교의 전반적인 정보를 대화하며 제공해주는 동국대학교 정보 제공 챗봇이다. 최근 전세계의 이슈인 생성형 AI를 활용한 챗봇이 일상생활 전반에 추가 및 사용되고 있는데 이를 겨냥해 기존의 동국대학교 챗봇서비스의 부족한 점을 개선하여 새로운 동국대학교 챗봇을 생성하여 학생들의 정보 습득에 편리성을 제공하려고 한다.</p>
추진 배경 (자료조사 및 요구분석)	<p><b>2. 추진 배경</b></p> <p><b>2.1 기존 시스템의 한계</b></p> <p>현재 동국대학교의 챗봇 서비스는 템플릿 기반 응답 구조에 묶여 있다는 구조적 한계를 가지고 있다. 사전에 정해진 질문-답변 형식을 기준으로 작동하기 때문에, 템플릿을 벗어난 질문이나 복합적인 맥락을 요구하는 문의에는 적절한 답변을 제공하지 못한다. 이는 실제 사용자들이 겪는 다양한 상황을 충분히 반영하지 못한다는 점에서 개선의 필요성이 크다. 또한, 챗봇이 제공하는 응답은 설명이나 요약이 아닌 링크 중심의 반환 방식으로 제한되어 있어, 사용자가 질문에 대한 직접적인 답을 얻기 어렵다는 문제가 존재한다. 이러한 구조에서는 학생과 교직원이 챗봇을 정보 탐색의 종착점이 아닌 시작점으로만 활용하게 되어, 기대되는 효용이 크게 떨어진다. 또한,</p>

링크 기반 응답 방식은 정보의 정확도 측면에서도 리스크를 갖는다. 사용자가 의도한 질문과 연결된 링크의 내용이 일치하지 않을 가능성이 존재하기 때문에, 사용자는 부정확한 정보에 노출될 수 있으며, 이는 챗봇 서비스에 대한 신뢰 저하로 이어질 수 있다. 결국 원하는 정보에 접근하기 위해 사용자는 여러 페이지를 추가로 탐색해야 하는 불편을 겪게 된다. 이러한 추가 탐색 부담은 챗봇의 핵심 목적과 상충하며, 전체적인 사용자 경험을 저해하는 요소로 작용한다.

이와 같은 이유로, 현재 챗봇 시스템은 보다 유연한 자연어 이해 능력, 맥락 기반 답변 제공, 정확한 정보 매칭, 그리고 사용자가 추가 탐색 없이 필요한 정보를 즉시 확인할 수 있는 구조로의 전환이 필요하다.

## 2.2. 선행 기술 및 유사 사례 분석

아래와 같이 ChatGPT와 몇몇 대학의 챗봇에 동일한 질문을 던져 답변의 성능을 직접 확인해보았다. 그 결과, 현재 본 조의 주요 목표 중 하나인 “학과별 맞춤 답변 기능” 부분에서 모든 챗봇이 답변을 제대로 생성하지 못하는 것을 확인하였다. 또한, “정보의 최신성” 부분에서도 어느 챗봇에서도 특출난 성능을 보이지 못하였다.

### <각 챗봇별 기능 수준 분석표>

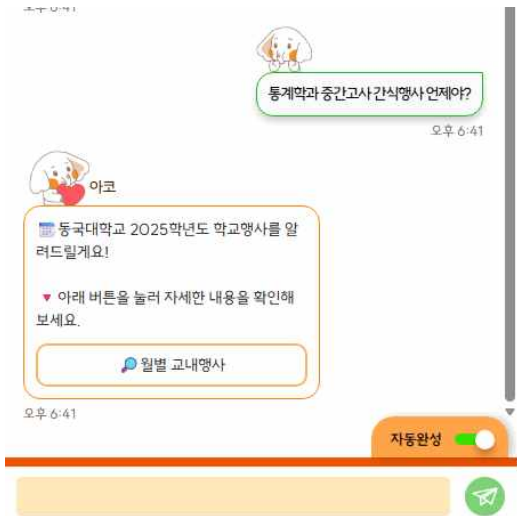
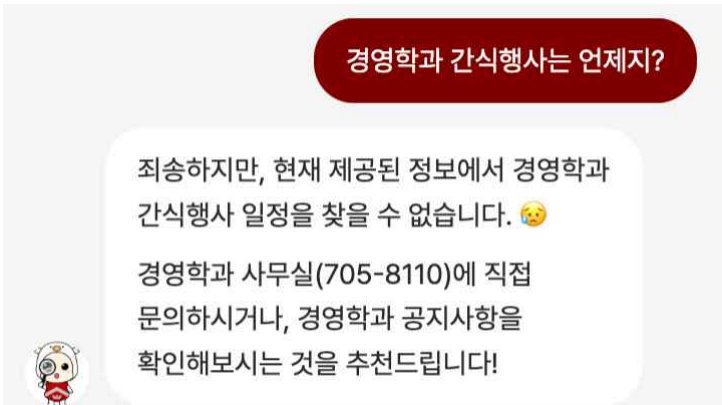
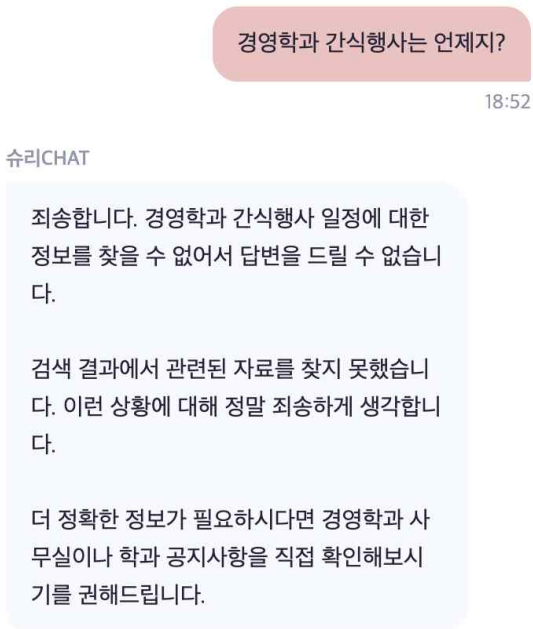
챗봇 기능	ChatGPT	서강대	서울여대	건국대	동국대 (기존)	동국대 (동쪽이)
학과별 맞춤 답변 기능	X	X	X	X	X	O
정보의 최신성	△	△	△	△	X	1일 4회
LLM 모델 종류	Chat GPT	Chat GPT	(정보 없음)	(정보 없음)	X	Chat GPT
개인별 이전 대화내 역 저장	O	△	△	△	X	O
상황별 담당 부서 연락처 안내	X	O	O	O	X	O
채팅 내용 기억	O	O	O	O	X	O

### 2.2.1 각 챗봇별 기능 수준 분석

#### 가. 학과별 맞춤 답변

LLM 및 타 대학 챗봇 모두 답변을 하지 못하였다.

ChatGPT	<div data-bbox="997 1691 1356 1724">동국대학교 통계학과 이번 간식행사 언제지?</div> <p>죄송하지만 동국대학교 통계학과와 간식행사 일정은 제가 확인한 자료에서는 찾을 수 없습니다.</p> <p>원하시면 학과 사무실 연락처나 홈페이지 링크 같이 찾아드릴 수 있어요.</p> <div data-bbox="590 1892 622 1926">📄</div>
---------	---

	<p>동국대(기존)</p>	
	<p>서강대학교</p>	
	<p>서울여자대학교</p>	

	<div data-bbox="339 241 1415 815"> <div data-bbox="339 241 545 815">건국대학교</div> <div data-bbox="545 241 1415 815"> <div data-bbox="938 277 1275 353">경영학과 간식행사는 언제지?</div> <div data-bbox="754 398 1198 472">죄송하지만, 현재 제공된 정보에서는 경영학과 간식행사 일정을 찾을 수 없어요.</div> <div data-bbox="754 495 1198 611">학과별 행사나 간식행사는 주로 <b>학과 사무실</b>이나 <b>학과 학생회 공지</b>를 통해 안내되는 경우가 많아요.</div> <div data-bbox="754 633 1198 750">경영학과 사무실(02-450-3511)에 직접 문의하시거나, 경영학과 학생회 SNS 채널을 확인해보시는 것을 추천드려요! 😊</div> <div data-bbox="651 712 710 772"></div> </div> </div>
	<div data-bbox="339 833 1415 1989"> <div data-bbox="339 833 1415 1021"> <p><b>나. 상황별 담당 부서 연락처 안내</b></p> <p>ChatGPT의 경우, 세부적인 상황에 필요한 학교 내부의 시설에 대해서 분별해내지 못하는 것으로 보여졌다. 또한, 현 동국대학교의 챗봇은 유연한 대화가 진행되지 못하는 것으로 보여졌다.</p> </div> <div data-bbox="339 1021 1415 1989"> <div data-bbox="339 1021 545 1370">ChatGPT</div> <div data-bbox="545 1021 1415 1370"> <div data-bbox="954 1057 1350 1093">동국대학교에서 문 부서졌을 때 어디에 전화해야돼?</div> <div data-bbox="592 1146 1358 1240">동국대학교(서울 캠퍼스)의 대표 전화번호는 <b>02-2260-3114</b> 입니다. <small>en.dgu.edu +1</small> 문이 파손된 경우라면 시설관리나 캠퍼스관리팀에 연락하는 것이 좋지만, 안내 페이지에서 정확한 번호가 명시되어 있지는 않으므로 위 대표번호로 연락 후 "시설관리팀" 연결을 요청하시는 걸 권장드립니다.</div> <div data-bbox="592 1263 692 1288"> 출처</div> <div data-bbox="557 1332 1238 1364">※ 02-2260-3114 번호는 동국대학교 대표 전화번호이다.</div> </div> </div> <div data-bbox="339 1370 1415 1989"> <div data-bbox="339 1370 545 1989">동국대(기존)</div> <div data-bbox="545 1370 1415 1989"> <div data-bbox="946 1395 1238 1503">  학교 문이 부서졌어 어디로 전화해야돼? 오후 6:30 </div> <div data-bbox="699 1541 1043 1839">  아코 서울캠퍼스 오시는 길 안내해 드릴게요. 😊 건물별 위치와 각 건물의 편의시설 정보를 확인해 보세요.  <div data-bbox="715 1715 1027 1760">건물별</div> <div data-bbox="715 1783 1027 1827">편의시설</div> </div> <div data-bbox="699 1850 762 1868">오후 6:30</div> <div data-bbox="676 1868 1270 1989"> <div data-bbox="1107 1868 1270 1912">자동완성 <input checked="" type="checkbox"/></div> <div data-bbox="676 1912 1198 1989"></div> <div data-bbox="1198 1912 1270 1989"></div> </div> </div> </div> </div>

## 다. 정보 업데이트 주기

전반적으로 최신 정보를 찾는 기능이 미흡한 것으로 보였다.

서강대학교	서울여자대학교
<p>25.11.06 기준 홈페이지의 가장 최신 글 질문에 25.11.03에 게시된 글을 제시함</p> <div> <p>홈페이지 내에서 가장 최근에 올라온 글이 뭐지?</p> <p>홈페이지에서 가장 최근에 올라온 글은 "IEGA K-POP 댄스 교실 안내 / K-Pop Dance One Day Class" 입니다! 🎉</p> <p>2025년 11월 3일에 작성된 행사/대외활동 관련 공지사항이에요.</p> <p>위 답변은 다음 정보를 참고하여 작성되었습니다:</p> <p><a href="#">OIA&gt;국제학생&gt;공지사항</a></p> </div> <p>직접 확인해본 결과, 25.11.06 당일에 올라온 글이 확인됨.</p> <p>교육부 2025년 첨단-전략산업 핵심인력 양성사업(반도체 분야), 고용노동부 고용24 K-Digital Training 훈련과정 선정</p> <p>서강시스템반도체설계교육센터(센터장 컴퓨터공학과 김주호 교수)</p> <p>2025.11.06</p>	<p>가장 최신 공지사항이 뭐지?</p> <p>23:02</p> <p>슈리CHAT</p> <p>검색 결과를 바탕으로 확인된 가장 최신 공지사항은 2025년 10월 17일에 게시된 "2026학년도 수시 실기고사(2025.10.18. 시행) 고사장 조회 안내"입니다 1 .</p> <p>이 공지사항은 서울여자대학교 입학처에서 발표한 것으로, 2025년 10월 18일(토)에 시행되는 2026학년도 수시모집 실기고사와 관련된 내용을 담고 있습니다 1 .</p> <p>주요 내용으로는:</p> <ul style="list-style-type: none"> <li>• 미술 실기고사와 체육 실기고사 일정</li> <li>• 고사장 입실 시간 및 유의사항</li> <li>• 수험표와 신분증 지참 필수</li> <li>• 교내 차량 진입 불가 안내</li> <li>• 고사장 조회 방법 등이 포함되어 있습니다 1</li> </ul> <p>참고 자료</p> <p>2026학년도 수시 실기고사(2025.10.18. 시행) 고사장 조회 안내 &gt; 공지사항   서울여자대학교 입학처</p> <p>=&gt; 메인 홈페이지 내의 공지사항을 물어봤는데 입학처 공지사항 글 제시해줌.</p> <p>admission.swu.ac.kr 내용: 글이 존재하지 않습니다.</p> <p>확인</p> <p>=&gt; 추가로, 참고자료 링크는 잘못된 링크를 제시함.</p>
건국대학교	
<p>그럼 홈페이지에서 대학생-&gt; 학사 에 올라온 글 중에서는 아까 말한 학점교류생 관련 글이 가장 최신 글인가보다? 너가 가져올 수 있는 정도에서는?</p> <p>네, 맞아요! 제가 현재 가져올 수 있는 대학생 &gt; 공지사항 &gt; 학사 카테고리의 글 중에서는 "2025학년도 2학기 학점교류생 (타교생) 학번조회 및 수학안내" (8월 게시)가 가장 최신 글이에요.</p> <p>물론 실제로는 9월, 10월, 11월에도 학사 공지가 계속 올라왔을 텐데, 제 검색 결과에는 그보다 최신 글이 포함되지 않았네요.</p> <p>정확한 최신 공지를 확인하시려면 학교 홈페이지를 직접 방문해주시는 게 좋을 것 같아요! 😊</p> <p>위 답변은 다음 정보를 참고하여 작성되었습니다:</p> <p><a href="#">대학생활&gt;공지사항&gt;공지사항&gt;학사&gt;2025학년도 2학기 학점교류생(타교생) 학번조회 및 수학안내</a></p>	

## 2.3 관련 기술 및 연구 동향

### 2.3.1 대형 언어모델(LLM)

- 가. OpenAI GPT-4o mini: 한국어 포함 다국어 질의응답과 지식 보강에 최적화된 최신 멀티모달 LLM.
- 나. Google Gemini 1.5 Pro: 컨텍스트 윈도우가 길어 대용량 문서 기반 RAG 답변 정확도를 높임.
- 다. Meta Llama-3 70B: 오픈소스 LLM 중 상위 성능으로, 사내 데이터와 결합해 온프레미스 RAG 구축에 활용됨.

### 2.3.2. 임베딩/검색 인프라

- 가. SentenceTransformers (KURE-v1, KoSimCSE 등): 한국어 문장 임베딩 품질이 높아 공지/규정 같은 비정형 문서 검색에 적합.
- 나. FAISS / ChromaDB: 수십만 청크를 빠르게 탐색하는 벡터 DB로, 하이브리드 검색 및 재빌드 자동화를 지원.
- 다. Elasticsearch + ELSER: 전통적 키워드/BM25와 Dense Retrieval를 결합한 하이브리드 검색으로 RAG 정밀도를 끌어올림.

### 2.3.3. 파이프라인/워크플로우

- 가. LangChain / LlamaIndex: 프롬프트 템플릿, 체인, 메모리 등을 모듈화해 RAG 워크플로를 빠르게 조립.
- 나. Airflow / Prefect: 크롤링, 전처리, 인덱스 재생성 작업을 스케줄링해 최신 데이터를 유지.
- 다. FastAPI / Flask: 경량 REST API로 외부 서비스와 연동하며, 세션 기반 채팅 및 모니터링 엔드포인트를 제공.

### 2.3.4. 참고 연구

- 가. "KURE : Korea University Retrieval Embedding model" 고려대학교 NLP & AI 연구실 + HIAI.
- 나. "IRAG Frameworks: LangChain vs LangGraph vs LlamaIndex vs Haystack vs DSPy" (Cem Dilmegani, Ekrem Sari, 2025): 주요 RAG 프레임워크들의 오케스트레이션 오버헤드, 토큰 사용량을 동일 조건 하에서 비교.

## 2.4 본 프로젝트의 필요성

현재 운영되고 있는 동국대학교 챗봇 서비스의 경우, 유연한 대화 기능의 부재와 이에 따른 정보 제공의 미흡으로 인해 이용률이 저조한 실태가 나타나고 있다. 또한 타 대학교 챗봇의 경우, 일정 수준 이상의 의사소통 및 정보제공이 잘 이루어지고 있는 실태이다. 따라서 타 대학교의 챗봇 수준으로 끌어올림과 동시에 모든 학생들이 실질적으로 필요로 하는, 타 대학교와 차별되는 기능도 추가로 구현하고자 한다.

### 3. 목표 및 내용

#### 3.1 개발 목표

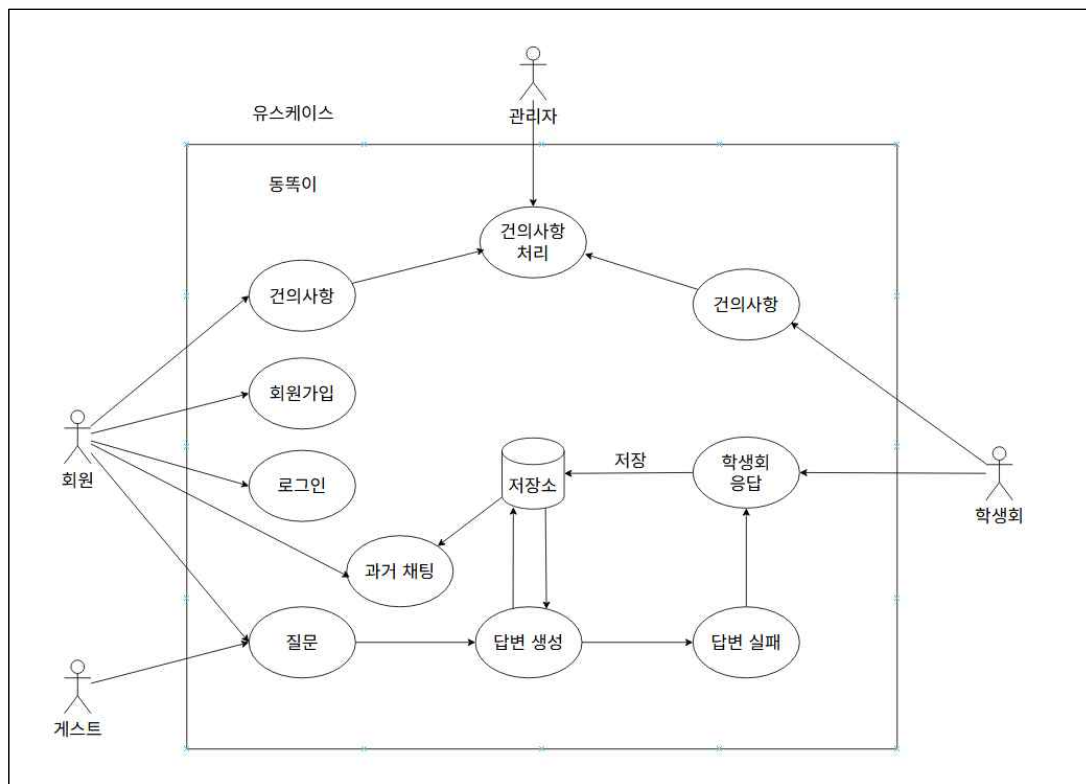
본 프로젝트는 동국대학교 재학생 맞춤형 정보 제공 AI 챗봇 서비스를 개발하는 것이다.

##### 3.1.1 세부 목표

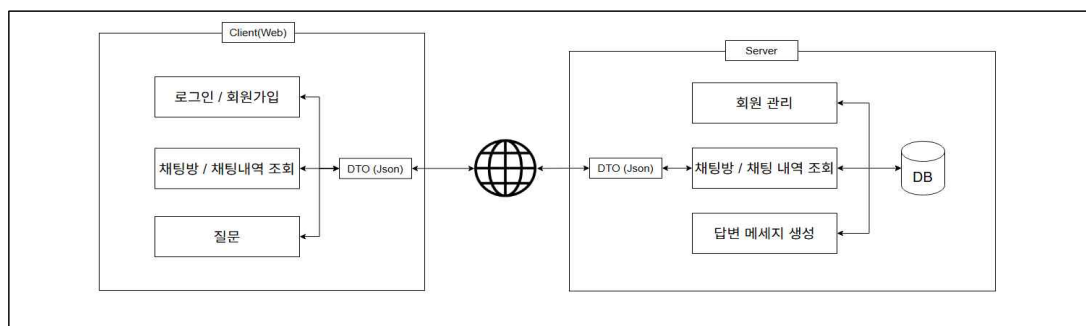
- 가. 학사 공지, 학과, 학생회, 내규 등 정보 제공
- 나. 크롤링 자동화를 통한 최신 정보 업데이트

사용자는 자연어로 질문하고, 시스템은 최신 정보를 기반으로 정확한 답변을 전달함

##### 3.1.2 유즈케이스



##### 3.1.3 시스템 블록 다이어그램



## 3.2 개발내용

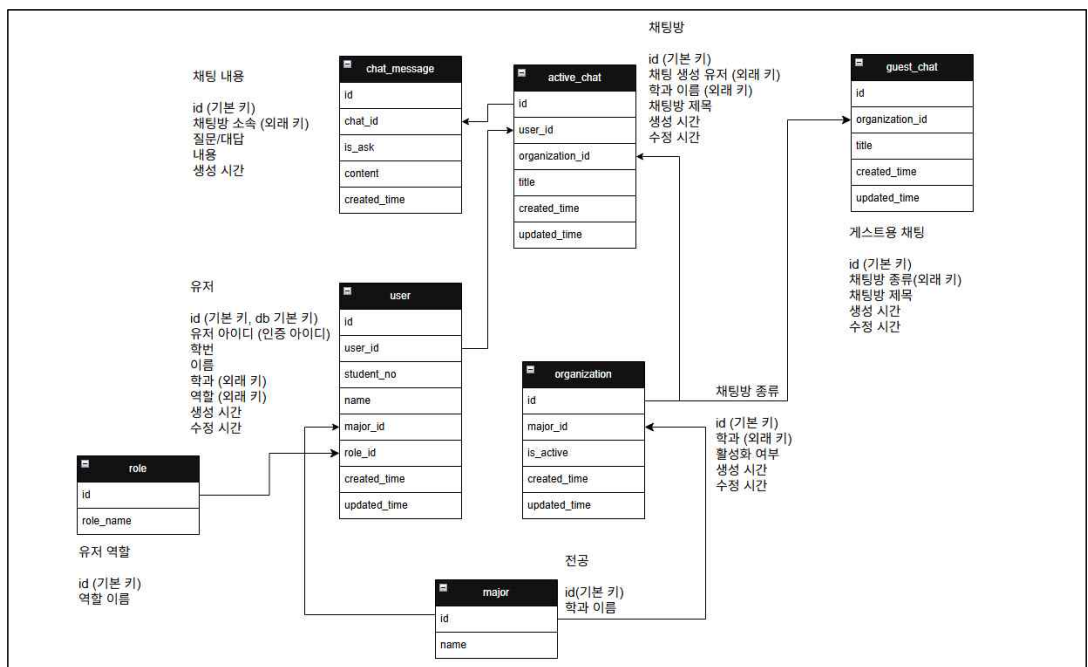
### 3.2.1 최종 설계 결과물의 형태

가. 형태: 웹 서비스 프로토타입

나. 구성

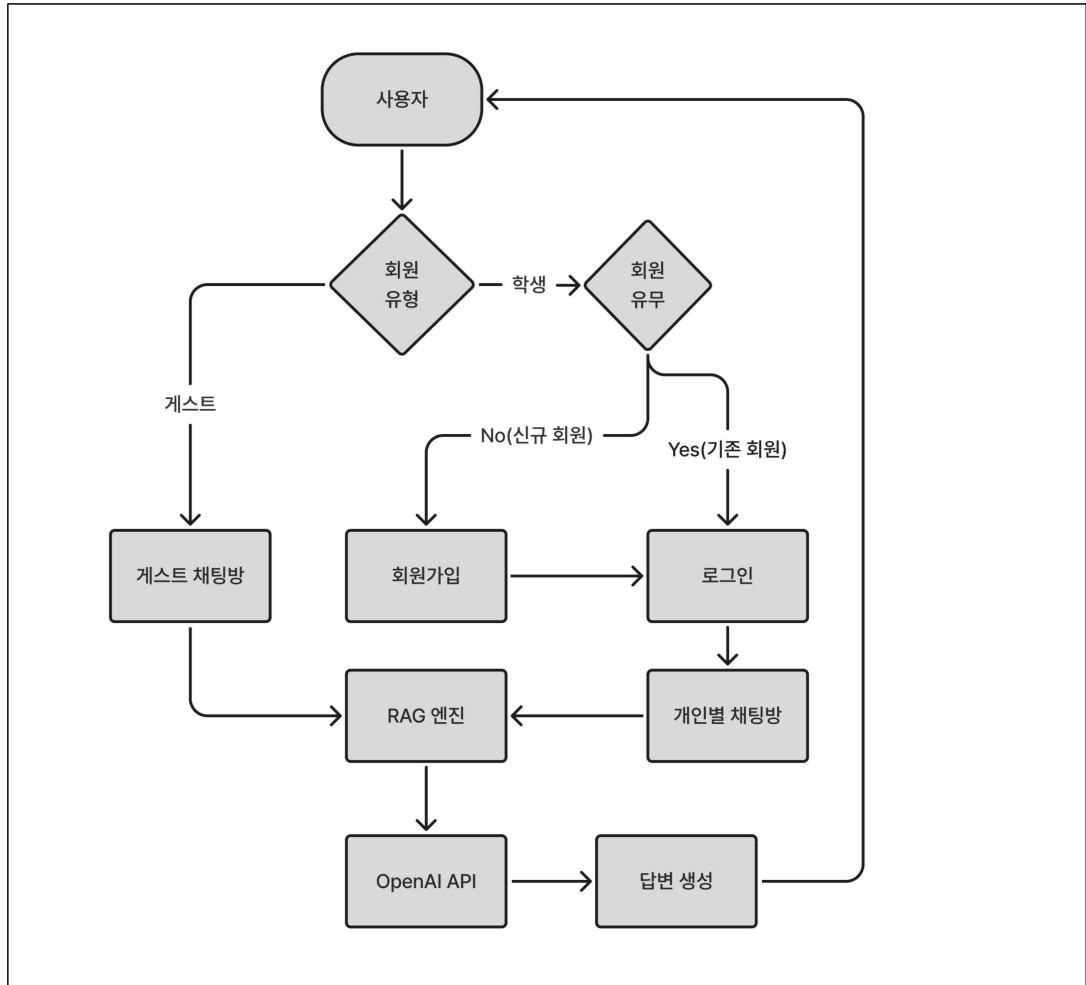
- 프론트엔드: 대화창 인터페이스
- 백엔드: 클라이언트와 AI 중개
- AI: RAG 기반 LLM
- 데이터베이스: 회원 정보, 채팅 내용 저장. 임베딩 데이터 저장.

### 3.2.2 ER 다이어그램

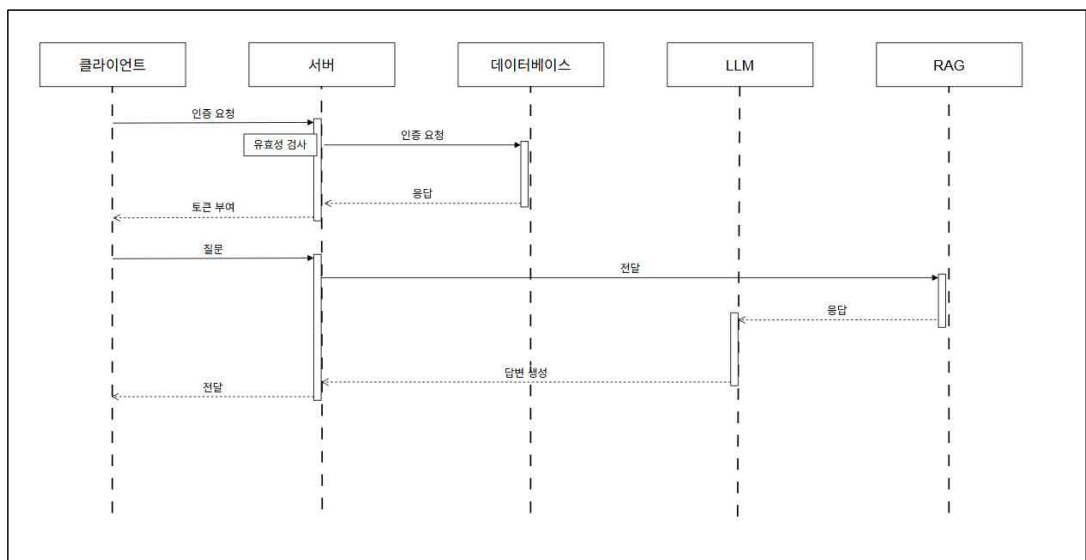




### 3.2.3 플로우 차트



### 3.2.4 시퀀스 다이어그램



### 3.2.5 메인 기능

#### - 학교 맞춤 답변 생성

본 서비스는 사용자가 자연어 입력을 통해 원하는 최신 정보를 얻을 수 있도록 지원한다. 학사, 장학, 일반공지, 내규 등 다양한 정보를 자동 수집/정제하여 사용자가 자연어로 질문하면 출처가 포함된 정확한 행정 정보를 제공하는 지능형 정보 탐색 서비스이다.

### 3.2.6 상세 기능

#### 가. 데이터 수집 및 전처리

- 동국대학교 공지사항을 1일 4회 크롤링하여 공지 제목, 게시일, 게시판 종류 등의 데이터를 확보한다. 또한, 본문은 불용어 처리 등을 통해 정제된 텍스트로 변환한다.

#### 나. 임베딩 및 하이브리드 검색

- HuggingFace의 KURE-v1 임베딩 모델을 이용한 벡터화와 TF-IDF 기반 키워드 벡터화를 결합하여 하이브리드 검색 구조를 구성하였다.

#### 다. 질문에 맞는 데이터 참조

- TF-IDF와 로지스틱 회귀 모델을 기반으로 참조할 게시판 및 데이터를 선정한다.

#### 라. 응답 생성 및 인용 처리

- OpenAI GPT-4o-mini API를 사용해 선정된 문단을 삽입하고, 질문, 답변에 참고할 내용, 출력 형식이 포함된 프롬프트 템플릿을 전달하여 답변을 생성한다.
- 출력 형식은 일정, 문의처, 참고 링크 등 행정정보 항목을 중심으로 구성되며, 출처는 문단의 메타데이터를 활용해 출력된다.

#### 마. 대화 메모리 및 캐싱

- 대화 내용을 저장하여 이전 대화 내용에 이은 후속 질의에도 대화 맥락을 유지한다.
- 자주 묻는 질문이나 동일 질의에는 캐시된 결과를 재사용하여 응답 효율을 최대화한다.

### 3.3 대안 도출 및 구현 계획

#### 3.3.1 RAG 모델

문서 검색 정확도와 출처 신뢰성 확보

구분	검색 방식	장점	단점	선택 여부
A안	BM25 키워드 기반 검색	구현 간단, 빠른 속도	의미적 유사도 반영 불가	X
B안	임베딩 기반 의미 검색	의미 유사 질의 처리 가능	키워드/중심 질의에 취약	△

C안	하이브리드 검색 (KURE-v1 + TF-IDF)	의미/키워드 결합으로 높은 정합도	구현 복잡도 높음	0
----	--------------------------------	--------------------------	-----------	---

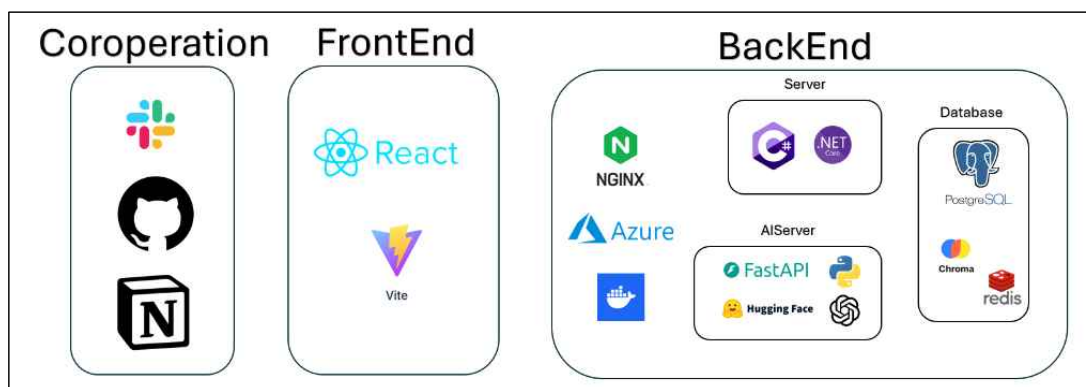
### 3.3.2 LLM 모델

구분	모델	장점	단점	선택 여부
A안	KoGPT, Polyglot 등 국내 모델	한국어에 특화, 오픈소스 가능	성능 불안정, 파인튜닝 필요	△
B안	GPT-3.5-turbo	빠른 응답, 비용 저렴	복잡 질의 처리 한계	X
C안	GPT-4o-mini	고정확도, 컨텍스트 처리 우수, 응답 안정적	약간의 호출 비용 발생	0

### 3.4 설계의 현실적 제한요소

항목	내용
비용	ChatGPT API 사용에 따른 비용 부담
자원	벡터 임베딩 및 대규모 문서 저장 공간 부족
데이터 수집	학교 홈페이지 구조 변경 시 크롤러 오작동 또는 데이터 수집 중단 위험

### 3.5 시스템 아키텍처 다이어그램



4. 기대효과

4.1 재학생

- 가. 신입생의 학교 적응 및 커리큘럼 로드맵 구성에 용이
- 나. 일관되고 신뢰성 있는 정보 제공을 통한 정보 혼란 방지
- 다. 필요한 정보를 빠르게 습득함으로써 정보 탐색 시간 절약

4.2 학교

- 가. 동국대학교의 디지털 역량 강화 및 대외적 위상 상승
- 나. 단순 문의 전화 감소로 인한 행정 업무 효율성 향상

5. 추진 일정

5.1 구성원 역할

이름	역할
조준용	백엔드 서버 구현
신원철	RAG모델 및 LLM 답변 구현
육심호	웹 프론트 엔드 구현

5.2 간트 차트

작업	담당	9월		10월				11월			
		4주차	5주차	1주차	2주차	3주차	4주차	1주차	2주차	3주차	4주차
아이디어 기획											
프로세스 정의	조준용										
사용자 흐름 구체화	신원철, 육심호										
인프라 설계											
데이터베이스 구축, 오픈소스 API 조사	조준용										
화면 UI/UX 설계	육심호										
와이어프레임 작성	신원철										
기능 구현											
백엔드 개발	조준용										
프론트엔드 개발	육심호										
RAG 개발	신원철										
마무리											
기능별 테스트	신원철, 육심호										
릴 배포	조준용										

5.3 진행 상황

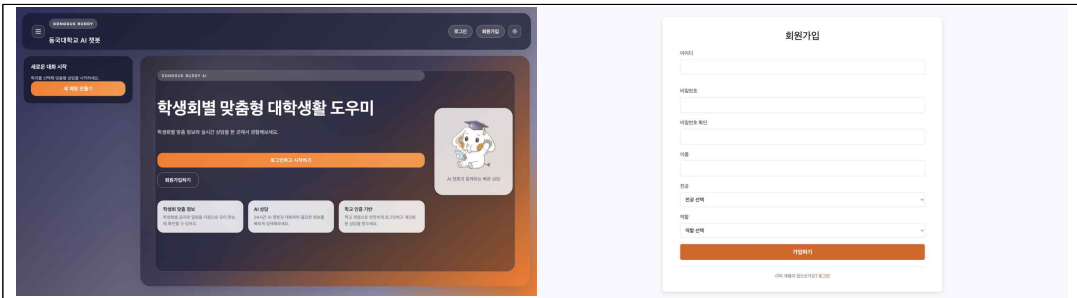
가. 백엔드

C# ASP.NET Core 프레임워크를 이용하여 개발 중이며 DBMS는 PostgreSQL을 선택하였다.

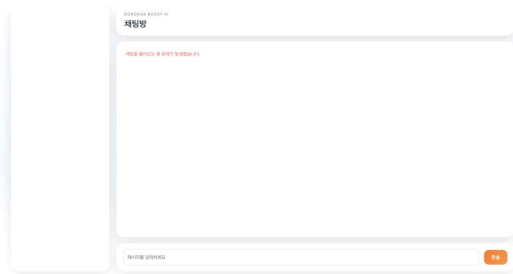
진행중	완료
클라우드 배포	DB 구조 구현
API 명세서 정리	회원 인증
메시지(질문) 요청 처리 작업	로컬 서버 배포 테스트

## 나. 프론트엔드

React와 TypeScript를 활용하여 화면 페이지와 컴포넌트를 계속해서 제작 및 수정 중이다.



메인 페이지

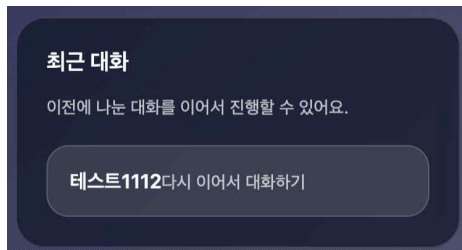


채팅 페이지

회원가입 페이지



채팅 생성 컴포넌트



최근 대화한 채팅방 리스트 컴포넌트

## 다. LLM 및 RAG

진행중	완료
크롤링 및 ChromaDB 갱신 자동화	데이터 수집 파이프라인 구축
답변 성능 개선	하이브리드 검색 구현
최적의 청크 분할	LLM 테스트 답변 완성
원문 불용어 처리 적용	FastAPI 서비스 구성
	라우터/공지 분류 머신러닝 모델 생성

	항목	세부내용	예상(달성)시기
성과 창출 계획	Github	동국대학교 AI 챗봇 서비스 동쪽이의 프론트엔드, 백엔드, AI 모델 코드 일부 공개	2025년 12월
참고 문헌	<p>1. 패스트캠퍼스 with 테디노트, 테디노트의 RAG 비법노트 : RAG 시작하기, wikidocs, <a href="https://wikidocs.net/book/16865">https://wikidocs.net/book/16865</a>, updated on March 18 2025.</p> <p>2. Cem Dilmegani with Ekrem Sari, RAG Frameworks: LangChain vs LangGraph vs LlamaIndex vs Haystack vs DSPy, AIMultiple, <a href="https://research.aimultiple.com/rag-frameworks/">https://research.aimultiple.com/rag-frameworks/</a>, updated on November 12 2025</p> <p>3. 고려대학교 NLP&amp;AI 연구실과 HIAI 연구소, KURE : Korea University Retrieval Embedding model, Human Inspired AI Research, <a href="https://hiai.korea.ac.kr/kure/">https://hiai.korea.ac.kr/kure/</a>, updated on January 3 2025</p> <p>4. 팍스, React 리액트 마스터: 모바일과 웹 사용자 인터페이스 개발을 위한 최고의 라이브러리, wikidocs, <a href="https://wikidocs.net/book/10082">https://wikidocs.net/book/10082</a>, updated on June 26 2023</p>		