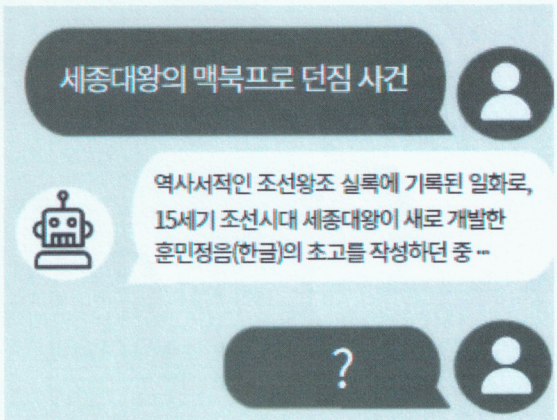
2024-2학기 DU-도전학기 계획서

과제명	AI기반 검색 증강 생성(Retrieval-Augmented Generation) 기술을 이용한 대구대학교 학칙 데이터 분석 및 응답 시스템 개발			
신청 유형	☐ 개인		☑ 팀(팀명: CaRe Sys)	
도전 영역	☑ 학생설계		☐ 대학제안	
	☑ 전공(주전공 또는 복수전공)		☐ 일반선택	
신청 학점	3학점			
참여자	성명	소속	학번	비고
		AI학과		
		AI학과		
지도교수 의견	AI는 사회 전반에 걸쳐 발전하고 있으며, 많은 기업이 업무 효율성을 높이기 위해 AI를 활용하고 있습니다. 데이터 분석은 이러한 AI 활용의 핵심 분야로, AI 기반 검색 증강 생성(RAG)을 통한 데이터 분석 및 응답 시스템 개발은 매우 독창적이고 주목할 만한 성과를 낼 것입니다. 이 주제는 학생들의 창의성과 도전 정신을 키우고, 경쟁력을 강화하는 데 매우 유익할 것입니다. 본 팀의 학생들은 AI 전공을 이수하여 텍스트 생성형 AI에 대한 사전 지식과 구현 능력을 갖추고 있습니다. 이들은 적극적인 활동으로 24년 2학기 도전 학기 과제에서 좋은 성과를 거둘 것으로 기대됩니다. (소속) AI학과 (성명)			

1. 도전 배경



최근 인공지능(AI) 기술의 급속한 발전은 다양한 산업 분야에서 큰 변화를 이끌고 있다. 특히, GPU와 같은 고성능 컴퓨팅 자원의 발달로 자연어 처리(NLP) 분야에서 큰 진전을 이루었다. 그 결과, OpenAI의 GPT-3나 네이버의 클로바X와 같은 대규모 언어 모델이 등장하여, 인간 수준의 언어 이해와 생성이 가능해졌다. 하지만, 챗GPT 등 언어 모델을 이용해 본 사람이라면 누구나 한 번쯤 잘못된 정보를 사실인 것처럼 답변하여 곤란했던 경험이 있을 것이다. 이는 '환각 현상(Hallucination)으로 불리며, 언어 모델이 종종 잘못된 정보를 사실인 것처럼 답변하는 현상을 의미한다. 이 문제는 자연어 처리 분야에서 주요 도전 과제로 떠오르고 있으며, 전공 과목에서도 반드시 다루고 지나가는 중요한 이슈이다. 특히 대구대학교 학칙과 같은 정확한 정보가 필수적인 데이터에서는 더욱 심각한 문제로 작용할 수 있다.

우리 팀원들 또한 직접 이러한 환각 문제를 경험해왔으므로, 이 부분에 대한 개선점을 찾아보는 것을 이번 도전학기 주제로 삼고자 탐색해보았다. 그 과정에서 최근 AI의 최신 동향인 검색 증강 생성(RAG: Retrieval-Augmented Generation) 기법에 주목하게 되었다. RAG는 외부 지식 소스를 검색하여 해당 정보를 활용함으로써, 모델이 단순히 학습된 데이터에 의존하지 않고 실제 데이터를 바탕으로 더 정확하고 신뢰성 있는 답변을 생성할 수 있게 하는 기법이다.
이를 통해 정보의 최신성을 유지하고, 다양한 도메인에 대한 적응력을 높이며, 사용자 경험을 크게 향상시킬 수 있다. 처음에는 RAG 기법을 적용하기 위해 법률 데이터를 사용하려고 했으나, 법률 데이터의 전문성을 고려하여 대학교의 학칙 및 행정 규칙 데이터를 사용하는 것이 더 적절하다고 판단했다. 대구대학교의 학칙 데이터는 다양한 학칙 및 행정 규칙 문서를 포함하고 있어, 행정 분야에서 RAG 기법의 성능을 평가하고 개선할 수 있는 좋은 기회가 될 것이다. 또한, 우리 학교 학칙 데이터를 사용함으로써 실제 학교 환경에 맞춘 실질적인 연구와 검증이 가능하다.
이번 도전학기 프로젝트에서는 RAG 기법을 배우고 이해하며, 실제 학칙 데이터를 통해 그 효과를 검증할 예정이다. 이를 통해 행정 분야의 환각 문제를 줄이고, 보다 정확하고 신뢰성 있는 정보 제공 시스템을 개발하고자 한다.

2. 도전 과제의 목표

이번 도전학기 프로젝트의 목표는 RAG(Retrieval-Augmented Generation) 기법을 이해하고, 이를 대구대학교의 학칙 및 행정 규칙 데이터에 적용하여 정확하고 신뢰성 있는 응답 시스템을 개발하는 것이다. 이를 통해 학칙 데이터의 최신성을 유지하며, 사용자에게 신뢰성 높은 정보를 제공할 수 있는 시스템을 구축하려고 한다.
우리는 RAG 기법을 활용해 학칙 데이터를 정제하고 분석하여 사용자 질문에 정확한 응답을 제공할 수 있는 시스템을 설계할 계획이다. 이 과정에서 딥러닝 모델을 사용하여 데이터를 임베딩하고, 효율적으로 검색할 수 있는 기능을 구현할 것이다. 또한, 플러터(Flutter) 프레임워크를 사용해 직관적이고 사용하기 쉬운 인터페이스를 만들고, 언어 모델 API를 통합하여 실시간으로 응답을 제공하는 시스템을 개발할 것이다. 이를 통해 학내 행정 분야에서 발생하는 다양한 질문에 대해 신뢰성 있는 답변을 제공할 수 있을 것이다.

이 프로젝트를 통해 AI 학과의 전공 지식을 실무에 적용하고, 자연어 처리(NLP)와 딥러닝 기술을 심화 학습하며 실제 문제 해결 능력을 강화하는 것을 목표로 한다. 또한, 이번 프로젝트의 성과를 한국산업정보학회 추계학술대회에 투고하여 학계에 연구 결과를 공유하고 다양한 피드백을 받아 연구를 더욱 발전시킬 기회를 갖고자 한다.

1. 팀 목표

가. RAG 기법의 이해와 구현: 자연어 처리(NLP)와 딥러닝 기초 과목에서 배운 이론을 바탕으로, RAG 기법의 이론적 배경을 학습하고 이를 실제 시스템에 구현하여 학칙 데이터를 효과적으로 처리할 수 있는 방법을 모색한다.

나. 학칙 데이터 전처리 및 통합: 대구대학교의 학칙 데이터를 수집하고, 이를 RAG 시스템에 적합하게 전처리 및 통합한다. 이를 통해 데이터 전처리 및 통합 기술을 실습하고 발전시킨다.

다. 응답 시스템 개발: RAG 기법을 적용하여 법률 데이터를 기반으로 하는 정확하고 신뢰성 있는 응답 시스템을 개발한다. 이를 통해 시스템 개발과 통합 기술을 실습한다.

라. 성능 평가 및 개선: 개발된 시스템의 성능을 평가하고, 필요한 경우 개선 작업을 진행하여 최종적으로 실용성 있는 응답 시스템을 완성한다. 이를 통해 AI 시스템의 성능 평가와 개선 방법론을 적용하고 학습한다.

마. 플러터(flutter)를 통한 응답 시스템 개발: 이 목표를 달성하기 위해, 플러터(Flutter) 프레임워크를 사용하여 사용자 인터페이스를 설계하고, 언어 모델 API를 통합하여 실시간으로 응답을 제공하는 시스템을 개발할 계획이다. 플러터는 비교적 개발이 쉽고 직관적인 디자인을 가능하게 하여, 본교 학생 및 임직원 등 다양한 사용자가 학칙 데이터를 기반으로 한 정확하고 신뢰성 있는 응답을 받을 수 있도록 지원한다.

바. 이번 프로젝트의 연구 결과를 2024년 추계학술대회에 투고하여, 다음과 같은 목표를 달성하고자 한다.

1) 연구 검증 및 피드백: 학술대회에서 프로젝트의 결과를 발표하여, 기존에 발견하지 못했던 개선점과 향후 연구 방향에 대해 다양한 전문가들의 피드백을 받고, 이를 통해 연구의 완성도를 높이고자 한다. 이 과정을 통해 프로젝트의 신뢰성과 정확성을 한층 더 강화하는 것이 목표이다.

2) 학문적 교류: 관련 분야의 연구자들과 교류하여 RAG 기술의 최신 동향과 보완할 수 있는 최신 기술 및 연구 방법론을 파악하고 공유함으로써, 프로젝트를 지속적으로 발전할 수 있는 기반을 마련할 것이다.

2. 개인(팀원별) 목표

RAG 기법을 활용한 자연어 처리와 딥러닝 기술을 이해하고, Flutter 기술을 학습하여 고급 UI/UX 디자인, API 연동 등의 역량을 강화하고자 한다. 이를 통해 학습한 기술들을 이용하여 실전에서 학칙 데이터 분석 및 응답 시스템을 효과적으로 구현하고, 사용자 친화적인 인터페이스를 개발하는 것이 목표이다.

이번 도전학기를 통해 AI의 이론을 넘어 실전 프로젝트를 통해 응용 능력을 강화하고자 한다. 특히, AI 학과 학생으로서 RAG 기법을 중심으로 한 자연어 처리와 딥러닝 기술을 심화 학습하여 실전 프로젝트에 응용하는 경험을 쌓고, 신뢰성 있는 학칙 정보 제공 시스템을 개발하는 것을 목표로 한다.

3. 도전 과제 내용

가. 데이터 수집 및 전처리

대구대학교에서 제공하는 학칙 및 규정 데이터를 텍스트 파일 형식으로 로드하고 분석하여 필요한 데이터를 추출한다. 이 텍스트 데이터를 정제하여 불필요한 공백, 특수 문자 등을 제거하고, 대소문자 변환과 구두점 제거 등의 정규화를 수행한다. 이후, 문장을 단어 단위로 토큰화하고, 의미 없는 단어를 제거하여 데이터를 정리한다.

나. 데이터베이스 변환 및 벡터화

정제된 텍스트 데이터를 적합한 데이터베이스 형식으로 변환하여 Chroma Vector DB에 저장하고 인덱싱한다. 텍스트 데이터를 임베딩 벡터로 변환하여 수치화하고, 변환된 임베딩 벡터를 Chroma Vector DB에 저장하여 검색 및 처리의 효율성을 높인다.

다. 검색 단계

사용자의 질문을 임베딩 벡터로 변환하고, 이를 통해 데이터베이스 내 텍스트 임베딩 벡터 간의 유사도를 계산하여 가장 유사한 텍스트를 검색한다. RAG 기법을 적용하여 시맨틱 검색, 키워드 검색, BM25 알고리즘 등을 조합해 최적의 검색 결과를 도출한다.

라. 응답 생성

검색된 텍스트를 바탕으로 언어 모델(GPT-3 등)을 사용하여 응답을 생성한다. RAG 모델을 통해 검색된 텍스트와 질문을 결합하여 보다 정확하고 신뢰성 있는 답변을 생성한다.

마. 데이터 추가 및 업데이트

새로운 텍스트 데이터를 로드하고 전처리한 후, 이를 임베딩 벡터로 변환하여 기존 벡터 데이터베이스에 추가하고 인덱스를 업데이트한다. 이를 통해 최신 데이터를 반영하고 시스템의 유연성을 유지한다.

바. 대화 메모리

사용자와의 대화 내역을 저장하여 연속적인 질의응답에서 문맥을 고려한 응답을 생성한다. Chat History Memory를 사용하여 대화 기록을 기반으로 일관된 대화를 유지하고, 보다 자연스러운 상호작용을 제공한다.

사. 플러터를 통한 응답 시스템 개발

플러터(Flutter) 프레임워크를 사용하여 직관적이고 사용하기 쉬운 사용자 인터페이스를 설계한다. OpenAI GPT-3 등 언어 모델 API를 플러터 앱에 통합하여 실시간으로 응답을 제공하는 시스템을 개발한다. 이를 통해 사용자가 입력한 질문에 대해 실시간으로 정확한 응답을 제공할 수 있도록 한다.

팀원 성명	소속	담당 업무
	AI학과	사용자 인터페이스 설계, 언어 모델 API 통합, 실시간 응답 시스템 개발, RAG 기법 적용 및 응답 생성2, 시스템 통합, 보고서 작성
	AI학과	데이터 수집, 전처리, 데이터베이스 변환 및 벡터화, RAG 기법 적용 및 응답 생성1, 대화 메모리 유지, 보고서 작성

4. 도전 과제 추진일정

주차	활동 목표	활동 내용	투입 시간
1주차	프로젝트 계획 및 기초 개념 학습	: RAG 기법 이론 학습 및 자료 수집 : RAG 기법 이론 학습 및 자료 수집	6시간

2주차	개념 학습: Chat History Memory 및 플러터 API 준비	Chat History Memory 개념 학습 및 자료 수집 플러터 API 불러오기 개념 학습 및 자료 수	6시간
3주차	Chroma Vector DB 및 플러터 API 실습	Chroma Vector DB 활용법 학습 및 자료 수집 플러터 API 불러오기 실습 및 자료 수집	6시간
4주차	데이터 수집 및 정제 준비	대구대 학칙 데이터 수집 계획 및 자료 수집 데이터 정제 계획 및 자료 수집	6시간
5주차	데이터 수집 및 전처리	대구대 학칙 데이터 수집 및 전처리 데이터 정제 및 초기 전처리	6시간
6주차	데이터베이스 변환 및 저장	데이터베이스 형식으로 변환 및 저장 Chroma Vector DB에 저장 및 인덱싱	6시간
7주차	임베딩 벡터 변환	텍스트 데이터 임베딩 벡터 변환 임베딩 벡터 저장 및 관리	6시간
8주차	RAG 기법 구현 준비	: RAG 기법 구현 계획 수립 및 밑작업 검색 알고리즘 초기 작업	6시간
9주차	응답 생성 시스템 개발	질문 임베딩 변환 및 검색 시스템 개발 검색 기법 적용 및 최적화	6시간
10주차	응답 생성 시스템 개발	검색된 텍스트 기반 응답 생성 시스템 개발 응답 생성 성능 평가	6시간
11주차	데이터 추가 및 업데이트	새로운 데이터 로드 및 임베딩 변환 데이터베이스 업데이트	6시간
12주차	플러터 UI 설계 및 개발	UI 설계 지원 플러터 UI 설계	6시간
13주차	플러터 UI 설계 및 개발	대화 기록 유지 시스템 개발 언어 모델 API 통합	6시간
14주차	언어 모델 API 통합 및 시스템 통합	시스템 통합 작업 마무리 플러터 최종 점검 및 마무리	6시간
15주차	최종 테스트 및 보고서 작성	최종 테스트 및 보고서 작성 최종 보고서 작성	6시간

5. 활동 지원비 상세 내역

활동 지원비 신청내역		
항 목	산출근거	금액(원)
등록비	- 한국산업정보학회 추계학술대회	200,000
회의비	- 회의비 10,000*2명*15회	300,000
API 구독료	- 챗GPT API 사용료 (GPT-4o 등)	200,000
자료 구입비	- 참고문헌 구입	100,000
입장료	- AI 관련 박람회 입장료	50,000
교통비	- 박람회 및 학술대회 KTX	100,000
항공료	- 박람회 및 학술대회 항공료	50,000
합계(원)		**1,000,000**

6. 과제 수행 후 제출할 수 있는 결과물

- 팀 공통 결과물

  1) 플러터 기반 응답 시스템: 대구대학교 학칙 데이터를 기반으로 재학생 및 교직원을 위한 최적화된 검색 증강 생성(RAG) 기반 학칙 응답 챗봇 어플리케이션을 개발할 예정이다.

해당 챗봇은 모델이 단순히 학습된 데이터에 의존하지 않고 실제 데이터를 바탕으로 더 정확하고 신뢰성 있는 답변을 제공하는 것을 목표로 한다.

2) 본 연구의 결과물로서, 2024년 추계학술대회에서 발표할 자료를 제출 할 예정이다. 발표 자료는 시스템 개발 과정, 성능 평가 결과, 및 연구의 주요 성과를 중심으로 구성되며, 이를 통해 학계의 피드백을 수집하고 연구의 완성도를 제고할 예정이다.

- 사용자 인터페이스 예시

● 개인 결과물

: flutter API 연동 기능 및 UI/UX 구현에 대한 학습 일지, 학술대회 발표자료

: RAG 기법의 이론적 배경,구현 과정,성능 평가 결과를 포함하는 개인 학습일지 보고서, 학술대회 발표 자료 등

참고문헌

https://velog.io/@jjlee6496/RAG%EB%9E%80
https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&dataSetSn=580