

## 2023-2학기 DU-도전학기 계획서

과제명	생성형AI를 통한 학교챗봇 만들기			
신청 유형	<input type="checkbox"/> 개인		<input checked="" type="checkbox"/> 팀(팀명: 포워드(Forwarder))	
도전 영역	<input checked="" type="checkbox"/> 전공(주전공 또는 복수전공)		<input type="checkbox"/> 일반선택	
신청 학점	3학점			
참여자	성명	소속	학번	비고
	김 [REDACTED]	경영학과(AI소프트웨어)	[REDACTED]	팀장
	변 [REDACTED]	무역학과(AI소프트웨어)	[REDACTED]	팀원
지도교수 의견	<p>최근 생성형 인공지능의 발전과 사회 각 분야에서 활용은 급속도로 증가하고 있습니다. 이에 AI소프트웨어전공을 복수 이수하고 있는 본 팀의 학생들의 도전학기 주제로 생성형 AI기반의 챗봇 서비스 개발은 매우 적절한 주제입니다.</p> <p>그리고 본 팀의 학생들의 선행 학습으로 텍스트 생성형 AI에 대한 사전 지식과 구현 능력을 가지고 있는 것으로 파악하였습니다. 또한 본 도전학기를 수행할 팀은 평소 우수한 학업 성적과 적극적인 대내외 활동으로 23년 2학기 도전학기 과제의 좋은 성과를 거둘 것으로 예상됩니다.</p> <p style="text-align: right;">(소속) AI학부 (성명) 차경애 (서명) </p>			

### 1. 도전 배경

몇 해 전부터 전 세계 경제와 산업 분야는 '4차 산업 혁명'이라는 큰 변화 속도로 전진하고 있다. 분야를 가리지 않고 디지털화와 인공지능(AI) 도입의 시대, 많은 기업들과 연구기관들은 새로운 혁신을 찾고자 노력하며 경쟁이 치열해지고 있다. '4차 산업 혁명'이라는 흐름 속에서 교육 역시 그 영향을 받게 되었으며, 특히 대학교 교육은 이러한 새로운 기술에 대한 요구가 높아졌다. 이렇듯 4차산업혁명 시대를 선도하는 인공지능(AI) 기술은 다양한 분야에서 혁신을 이끌어왔다. 그중 대표적인 사례로 인공지능 기반의 챗봇 기술을 꼽을 수 있다. 전통적인 대화상자 방식에서 벗어나 이용자에게 다양한 정보와 서비스를 제공하는 상호작용 방식을 제공함으로써 사용자 경험을 향상시키는 것은 물론, 시공간 제약 없이 실시간으로 서비스를 제공함으로써 빠른 응대와 업무 효율성을 높이는데 기여하고 있다. 이러한 산업 트렌드를 고려하여 대학교 교육환경에서 인공지능 기반 Q&A 서비스를 도입하면 어떨까라는 생각에서 이 프로젝트는 시작되었다. 그 결과, 경영대학과 AI 소프트웨어학과를 복수전공하며 다양한 관점에서의 이해와 전문성을 가진 두 명의 팀원이 모여 생성형 AI기반 챗봇 프로젝트를 계획하게 되었다. 이 프로젝트를 통해 대학생들의 필요한 정보를 실시간으로 제공하여 업무 효율 낼 뿐만 아니라, 학생들 간의 질문과 답변을 공유함으로써 다양한 학문 분야에 대한 지식을 넓힐 수 있는 일종의 학술

커뮤니티를 만들어내고자 했다. 이제 인터넷이 생활의 필수 요소가 되었다. 특히 대학생들이나 교직원들에게 교육과 업무 효율을 높일 수 있는 창구로서 도운 역할을 할 수 있음에 기대를 모으고 있다. 이러한 도전과 열정으로 팀원들이 도전학기 프로그램에 신청하게 되었다.

## 2. 도전 과제의 목표

가. 팀 목표 : 생성형 AI기반 대학생 챗봇

기존에 주목을 받고 있는 LLM(Large Language Model) 모델을 파인튜닝하여 학교 환경에 최적화된 인공지능 모델 개발 목적: 학교 생활에 필요한 다양한 정보를 제공하고 원활한 소통과 지식 공유 지원, 신입생 및 재학생들의 학교생활 문제에 대한 적절한 답변 제공

기대 효과

효율적인 정보 제공: 학교와 관련된 정보, 행정지원, 학사일정, 추천 도서, 동아리, 실내외 시설 및 경로 등 시간과 공간의 제약 없이 효율적인 정보를 제공

학업 및 연구 지원: 학생들의 학업 수행과 연구를 위한 지원으로 업무 효율을 높임

내부망 활용성: 외부 인터넷 연결이 필요 없어 보안 강화 및 민감한 정보 유출 위험 최소화

나. 개인목표

1(김■■■■) 딥러닝 모델 학습을 통한 전공 역량 강화: 복잡한 인공지능 기반 챗봇 프로젝트를 진행하며 딥러닝 모델에 대한 이해도와 활용 능력을 키울 수 있습니다. 신경망 구조, 학습 방법, 최적화 알고리즘 등 딥러닝의 다양한 요소를 공부하고, 이를 실제 문제 해결에 적용할 수 있는 전문성을 갖추게 됩니다. 더불어 최신 딥러닝 트렌드 및 전문가들의 연구를 지속적으로 파악하여 학문적 기반을 더욱 탄탄히 다지게 됩니다.

2(■■■■) 크롤링 및 데이터 전처리를 통한 전공 역량 강화: 이 프로젝트를 통해 웹 크롤링 기술을 익히고 다양한 데이터 출처에서 필요한 정보를 수집하는 능력을 향상시킬 것입니다. 또한, 수집된 데이터를 효율적으로 활용하기 위해 데이터 전처리 및 분석 기법을 적용하여 더 깊은 지식을 습득할 수 있습니다. 이를 통해 빅 데이터를 분석하고 다양한 문제 해결에 활용하는 전공 역량을 강화하게 됩니다.

## 3. 도전 과제 내용

1. 웹 크롤링 및 데이터 수집: 웹 사이트나 온라인 자료에서 필요한 정보를 자동으로 추출하는 웹 크롤링 기술을 활용하여 관련 데이터를 수집합니다. 이 과정에서 다양한 데이터 소스를 탐색하고, 필요한 정보를 효과적으로 추출하기 위한 크롤러를 개발합니다.

2. 데이터 전처리: 수집된 데이터를 분석 및 학습에 용이한 형태로 변환하기 위해 데이터 전처

리 작업을 수행합니다. 이 과정에서 불필요한 데이터를 제거하고, 결측치를 처리하며, 데이터의 형식을 통일합니다.

3. 데이터 탐색 및 분석: 전처리된 데이터를 자세히 살펴보고, 데이터의 속성과 특징을 분석하여 학습에 활용할 수 있는 인사이트를 도출합니다. 이를 통해 적절한 학습 데이터를 구성하고 전공 역량을 높이는 데 도움이 됩니다.

4. 딥러닝 모델 구성 및 학습: 기존 LLM(Large Language Model)을 참고하여 적절한 딥러닝 모델을 설계하고, 이를 학습 데이터를 활용해 훈련시킵니다. 이 과정에서 다양한 하이퍼파라미터를 조정하고, 모델 구조를 변경해가며 최적의 학습 결과를 도출합니다.

5. 모델 성능 평가: 학습된 모델의 성능을 평가하기 위해 검증 데이터를 이용합니다. 정확도, 정밀도, 재현율, F1 스코어 등 다양한 평가 지표를 사용하여 모델의 성능을 측정하고, 개선할 수 있는 부분을 찾아 냅니다.

6. 모델 최적화 및 파인튜닝: 성능 평가를 바탕으로 모델의 최적화를 진행합니다. 이를 위해 하이퍼파라미터를 수정하거나 전처리 과정을 개선하며, 딥러닝 모델을 학교 환경에 최적화시킵니다.

7. 서비스 구축 및 테스트: 최종적으로 개발한 AI 모델을 실제 사용자가 이용할 수 있는 서비스로 구현합니다. 이를 위해 챗봇 인터페이스를 설계하고, 사용자의 질문에 대한 답변을 처리하는 로직을 개발합니다. 서비스가 구축되면 실제 사용자들이 테스트를 진행하면서 이슈와 개선 사항을 찾아냅니다.

8. 유지보수 및 지속 개선: 서비스가 구축된 후에도 지속적인 관리와 개선이 필요합니다. 사용자들의 피드백을 반영하고, 데이터를 업데이트하며, 기능을 추가해가면서 서비스를 지속 개선해 나갑니다. 이 과정에서 전공 역량을 계속적으로 강화할 수 있습니다.

팀원 성명	소속	담당 업무
김 [REDACTED]	경영학과	데이터수집1, 전처리1, 모델학습1, 보고서작성1
변 [REDACTED]	무역학과	데이터수집2, 전처리2, 모델학습2, 보고서작성2

#### 4. 도전 과제 추진일정

주차	활동 목표	활동 내용	투입 시간
1주차	프로젝트 기획 및 계획 수립	김 [REDACTED] (팀장) : 프로젝트 목표 및 범위 설정	6시간
		변 [REDACTED] (팀원) : 이해관계자 인터뷰 및 요구사항 수집	
2주차	데이터 수집을 위한 웹 크롤링 기술 학습	김 [REDACTED] (팀장) : 웹 크롤링 관련 라이브러리 및 툴 학습	6시간
		변 [REDACTED] (팀원) : 대상 웹사이트 및 자료 조사	
3주차	웹 크롤링 및 데이터 수집	김 [REDACTED] (팀장) : 크롤러 개발	6시간
		변 [REDACTED] (팀원) : 데이터 수집 작업 진행	
4주차	데이터 전처리	김 [REDACTED] (팀장) : 데이터 정제, 결측치 처리	6시간
		변 [REDACTED] (팀원) : 형식 변환 등 수행	
5주차	데이터 탐색 및 분석	김 [REDACTED] (팀장) : 데이터 분석 및 시각화 도구 활용	6시간

		변	팀원) : 인사이트 도출	
6주차	딥러닝 관련 지식 습득	김	팀장) : 기존 LLM 모델 및 딥러닝 구조 학습	6시간
		변	팀원) : 선행 연구 및 자료 조사	
7주차	딥러닝 모델 설계	김	팀장) : 초안 설계	6시간
		변	팀원) : 구조 설정	
8주차	데이터셋 준비 및 모델 학습 시작	김	팀장) : 학습 및 검증 데이터셋 구성	6시간
		변	팀원) : 모델 학습 진행	
9주차	모델 학습 및 최적화	김	팀장) : 학습 과정 모니터링	6시간
		변	팀원) : 최적화 방법 적용	
10주차	모델 성능 평가	김	팀장) : 검증 데이터 활용	6시간
		변	팀원) : 성능 평가 지표 활용	
11주차	모델 최적화 및 파인튜닝	김	팀장) : 하이퍼파라미터 수정	6시간
		변	팀원) : 전처리 방법 개선	
12주차	챗봇 인터페이스 및 로직 개발	김	팀장) : 인터페이스 설계	6시간
		변	팀원) : 사용자 입력 처리 로직 구현	
13주차	서비스 테스트 및 피드백 수집	김	팀장) : 사용자 테스트 진행	6시간
		변	팀원) : 피드백 및 이슈 처리	
14주차	서비스 개선 및 완성	김	팀장) : 피드백 반영	6시간
		변	팀원) : 서비스 완성도 향상	
15주차	프로젝트 결과 보고 및 마무리	김	팀장) : 결과 보고서 작성	6시간
		변	팀원) : 프로젝트 마무리 및 후속 일정 계획	

## 5. 활동 지원비 상세 내역

활동 지원비 신청내역			
항 목	산출근거		금액(원)
교통비	SRT	동대구역 - 수서역 왕복 SRT 기차비 38,000 X (왕복) X 2명 X 2회 = 304,000원 박람회 2회 참가 예정	304,000원
	회의비	팀 회의비	200,000원
	박람회 입장권	인당 10,000원 x 2명X 2번	40,000원
	자료구입비	관련 도서 구입	120,000원
	구글 코랩 모델 학습비용	월 49.99\$ X 4개월 = 199.96\$ + 알파(컴퓨팅 추가 가능) = 약 329,524원	329,524원
<b>합계(원)</b>			<b>993,524원</b>

## 6. 과제 수행 후 제출할 수 있는 결과물

### 팀 결과물

최적화된 인공지능 모델: 학교 환경에 적합한 LLM(Large Language Model) 파인튜닝을 거친 인공지능 모델입니다. 이 모델은 학교생활에서 겪는 다양한 문제에 대해 정확한 답변을 제공할 수 있도록 학습되어 있습니다.

챗봇 인터페이스: 사용자가 질문을 쉽게 입력할 수 있는 챗봇 인터페이스를 구현합니다. 이 인터페이스를 통해 사용자는 쉽게 AI 모델에 접근하고 원하는 정보를 얻을 수 있습니다.

예시



## 개인 결과물



개인 공부학습 노트



개인 공부학습 노트

[참고문헌]

<https://github.com/Beomi/KoAlpaca>

<https://aisummit.co.kr/>