

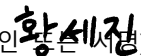
## 2023년 「미래내일 일경험사업」 우수 프로젝트 공모전 참가 신청서

참여 운영기관	한국생산성본부		
프로젝트명	금융관련 질의응답 자연어 처리를 위한 AI 학습모델 기획/제안		
프로젝트 요약	금융권 직원, 고객과 질의 응답하는 오픈 소스 모델 활용 생성형 AI 서비스 기획/제안		
성명(팀장)	황세진	생년월일	1997.06.17.
성명(팀원)	백두현	생년월일	1995.03.19.
성명(팀원)	박태원	생년월일	1997.11.27.
성명(팀원)		생년월일	
성명(팀원)		생년월일	
팀장 연락처 (휴대전화번호)	010-7490-8821	팀장 E-mail 주소	sj970617@naver.com

- 본인은 2023년 「미래내일 일경험사업」 우수 프로젝트 공모전에 참여하고자 위와 같이 신청하며, 기재 내용이 사실임을 확인합니다. ☒
- 본인은 제출 작품이 타 공모전에 기 입상되었거나 조작·편집·표절된 작품이 아님을 확인합니다. ☒
- 본인은 부정한 출품으로 인한 문제 사항, 출품작과 관련된 초상권·저작권 등 사항에 대한 문제 발생 시 본인에게 책임이 있음을 숙지하였습니다. ☒

2023년 10월 31일

공모전 참가 신청자(팀장)

황세진 (인 )

고용노동부 장관 귀하

2023년 「미래내일 일경험사업」 우수 프로젝트 공모전 참가를 위한  
개인정보 수집·이용·제3자 제공 동의서

※ 팀원 전원 작성

□ 개인정보 수집·이용 내역

구분	수집 목적
수집·이용하는 개인정보 항목	성명, 생년월일, 개인 연락처(휴대전화번호), E-mail 주소, 주소
수집 및 이용 목적	공모전 접수, 심사, 발표, 시상, 수상작 공개 시 본인 확인 및 안내, 「미래내일 일경험사업」 홍보
보유 기간	참가신청서 및 관련 자료 : 2024. 3. 31.까지

※ 귀하는 위 개인정보 수집·이용에 대한 동의를 거부할 수 있으며, 동의를 거부할 경우  
공모전 참여에 제한을 받을 수 있습니다.

☞ 위와 같이 개인정보를 수집·이용하는데 동의하십니까? (해당 칸에 체크)

☒ 예, 동의합니다.

☐ 아니오, 동의하지 않습니다.

□ 개인정보 제3자 제공 내역

구분	수집 목적
제공받는 자	2023년 「미래내일 일경험사업」 통합지원센터
제공받는 자의 이용 목적	공모전 접수, 심사, 발표, 시상, 수상작 공개 시 본인 확인 및 안내, 「미래내일 일경험사업」 홍보
제공하는 개인정보 항목	성명, 생년월일, 개인 연락처(휴대전화번호), E-mail 주소, 주소
보유 기간	참가신청서 및 관련 자료 : 2024. 3. 31.까지

※ 귀하는 위 개인정보 제3자 제공에 대한 동의를 거부할 권리가 있으며, 동의를 거부할  
경우 공모전 참여에 제한을 받을 수 있습니다.

☞ 위와 같이 개인정보를 제3자에게 제공하는데 동의하십니까? (해당 칸에 체크)

☒ 예, 동의합니다.

☐ 아니오, 동의하지 않습니다.

2023년 10월 31일

성명

황세진

(서명) 황세진

고용노동부 장관 귀하

[ 안내사항 ] 최종 수상작의 경우 상금 지급을 위해 별도의 수당지급 확인서 및 신분증  
사본과 통장 사본을 요청할 수 있습니다.

2023년 「미래내일 일경험사업」 우수 프로젝트 공모전 참가를 위한  
개인정보 수집·이용·제3자 제공 동의서

※ 팀원 전원 작성

□ 개인정보 수집·이용 내역

구분	수집 목적
수집·이용하는 개인정보 항목	성명, 생년월일, 개인 연락처(휴대전화번호), E-mail 주소, 주소
수집 및 이용 목적	공모전 접수, 심사, 발표, 시상, 수상작 공개 시 본인 확인 및 안내, 「미래내일 일경험사업」 홍보
보유 기간	참가신청서 및 관련 자료 : 2024. 3. 31.까지

※ 귀하는 위 개인정보 수집·이용에 대한 동의를 거부할 수 있으며, 동의를 거부할 경우  
공모전 참여에 제한을 받을 수 있습니다.

☞ 위와 같이 개인정보를 수집·이용하는데 동의하십니까? (해당 칸에 체크)

☒ 예, 동의합니다.

☐ 아니오, 동의하지 않습니다.

□ 개인정보 제3자 제공 내역

구분	수집 목적
제공받는 자	2023년 「미래내일 일경험사업」 통합지원센터
제공받는 자의 이용 목적	공모전 접수, 심사, 발표, 시상, 수상작 공개 시 본인 확인 및 안내, 「미래내일 일경험사업」 홍보
제공하는 개인정보 항목	성명, 생년월일, 개인 연락처(휴대전화번호), E-mail 주소, 주소
보유 기간	참가신청서 및 관련 자료 : 2024. 3. 31.까지


※ 귀하는 위 개인정보 제3자 제공에 대한 동의를 거부할 권리가 있으며, 동의를 거부할  
경우 공모전 참여에 제한을 받을 수 있습니다.

☞ 위와 같이 개인정보를 제3자에게 제공하는데 동의하십니까? (해당 칸에 체크)

☒ 예, 동의합니다.

☐ 아니오, 동의하지 않습니다.

2023년 10월 31일

성명 백두현 (서명) 

고용노동부 장관 귀하

[ 안내사항 ] 최종 수상작의 경우 상금 지급을 위해 별도의 수당지급 확인서 및 신분증  
사본과 통장 사본을 요청할 수 있습니다.

2023년 「미래내일 일경험사업」 우수 프로젝트 공모전 참가를 위한  
개인정보 수집·이용·제3자 제공 동의서

※ 팀원 전원 작성

□ 개인정보 수집·이용 내역

구분	수집 목적
수집·이용하는 개인정보 항목	성명, 생년월일, 개인 연락처(휴대전화번호), E-mail 주소, 주소
수집 및 이용 목적	공모전 접수, 심사, 발표, 시상, 수상작 공개 시 본인 확인 및 안내, 「미래내일 일경험사업」 홍보
보유 기간	참가신청서 및 관련 자료 : 2024. 3. 31.까지

※ 귀하는 위 개인정보 수집·이용에 대한 동의를 거부할 수 있으며, 동의를 거부할 경우  
공모전 참여에 제한을 받을 수 있습니다.

☞ 위와 같이 개인정보를 수집·이용하는데 동의하십니까? (해당 칸에 체크)

☒ 예, 동의합니다.

☐ 아니오, 동의하지 않습니다.

□ 개인정보 제3자 제공 내역

구분	수집 목적
제공받는 자	2023년 「미래내일 일경험사업」 통합지원센터
제공받는 자의 이용 목적	공모전 접수, 심사, 발표, 시상, 수상작 공개 시 본인 확인 및 안내, 「미래내일 일경험사업」 홍보
제공하는 개인정보 항목	성명, 생년월일, 개인 연락처(휴대전화번호), E-mail 주소, 주소
보유 기간	참가신청서 및 관련 자료 : 2024. 3. 31.까지

※ 귀하는 위 개인정보 제3자 제공에 대한 동의를 거부할 권리가 있으며, 동의를 거부할  
경우 공모전 참여에 제한을 받을 수 있습니다.

☞ 위와 같이 개인정보를 제3자에게 제공하는데 동의하십니까? (해당 칸에 체크)

☒ 예, 동의합니다.

☐ 아니오, 동의하지 않습니다.

2023년 10월 31일

성명

박태원

(서명)

고용노동부 장관 귀하

[ 안내사항 ] 최종 수상작의 경우 상금 지급을 위해 별도의 수당지급 확인서 및 신분증  
사본과 통장 사본을 요청할 수 있습니다.

2023년 미래내일 일경험 우수 프로젝트 공모전  
**프로젝트 보고서**

2023. 10. 31.

프로젝트명

금융관련 질의응답 자연어 처리를 위한 AI 학습모델 기획/제안



고용노동부



대한상공회의소

## 프로젝트 개요서(요약)

프로젝트 유형	<input checked="" type="checkbox"/> 기업제안 <input type="checkbox"/> 청년주도 <input type="checkbox"/> 자율공모
프로젝트명 (참여기업명)	금융관련 질의응답 자연어 처리를 위한 AI 학습모델 기획/제안 (대구은행)
팀명	비긴어게인
팀 구성원	(팀장) 황세진, (팀원) 백두현, 박태원 ※ (멘토) DGB대구은행 - 손현오 계장
수행기간	2023년 9월 18일 ~ 10월 19일
프로젝트 분야	IT
프로젝트 소개	대구은행의 디지털 전환 가속을 위한 AI 기반 서비스 제안 프로젝트로 3명의 팀원으로 이루어진 비긴어게인 팀이 한 달간 수행함. 황세진은 AI 모델 코드 구현을 맡았으며, 백두현은 아이디어 도출과 도메인 지식 역할을, 박태원은 모델 탐색 및 학습 데이터 설계 역할을 함. 거대언어모델(LLM)을 기반으로 한 생성형 AI 모델을 파인튜닝하는 프로젝트를 진행함.
추진배경 및 필요성	ChatGPT의 등장과 함께 생성형 AI를 통한 업무처리 수요 폭증으로 은행 산업에서의 보안과 신뢰성을 보장하고 한국어 특화 모델로 한국어를 잘 이해하여 질의응답이 가능해 고객에게 양질의 서비스(빠른 업무 처리, 정확한 질문 의도 파악과 답변 생성)를 제공하는 생성형 AI 기반 서비스를 제안하고자 프로젝트를 추진하게 되었음.
프로젝트 특징점	은행의 생성형AI 도입이 어려운 이유 중 하나가 다른 산업에 비해 개인정보 및 내부정보 유출에 더욱 민감하다는 점임. 이 문제를 해결하기 위해 오픈소스 모델이자 한국어 특화 모델을 사용하여 로컬 환경을 구축해 폐쇄적 공간에서 보안에 신경을 쓰고, 한국어 자연어 처리를 잘 수행하는 것을 특징점으로 한 신기술 구체화 프로젝트임.
주요 기능	오픈 소스 모델이자 한국어 거대언어모델 Polyglot-ko 12.8B 에 도메인 특화 데이터를 추가 학습함. 이때, 소수의 매개 변수만 미세 조정하여 계산 및 저장 비용을 크게 줄이는 PEFT 기법과 4-bit로 양자화된 모델로 LORA 기법을 활용해 파인튜닝하는 QLORA 기법을 사용함.
기대효과 및 활용 분야	향후 은행에서 생성형 AI 서비스 도입 시 행내 약관 및 상품설명서를 사전 언어 모델의 다운스트림(도메인) 데이터로 사용한다면, 오픈소스 모델로써 비용을 절감하고 고객 데이터가 유출되지 않는 직원 전용과 고객 전용 각각의 생성형 AI 챗봇 및 AI은행원 서비스로 고객 니즈를 만족시키고 직원 업무 효율성을 증대할 것이라 기대함. 도메인 특화 데이터에 따라 은행권 뿐만 아니라 전 산업에 활용 가능함.

# 프로젝트 보고서(본문)

## 1 추진배경 및 필요성

### 1. 환경 분석

- 사물인터넷(IoT), 인공지능(AI), 클라우드 등 금융권에도 신기술을 적극적으로 활용하고 있는 추세
- 프로젝트 추진 배경: 생성형 AI를 통한 업무처리 수요 폭증
- 국내 5대 은행 생성형 AI 동향
  - 국민은행: KB-GPT 데모 웹사이트 개설하여 검색, 채팅, 요약, 문서작성, 코딩 등 내부 직원 업무 효율성 제고에 적용 검토
  - 신한은행: 생성형 AI의 금융서비스 적용 전담 태스크 포스(TF) 출범
  - 하나은행: 자체 금융 특화 버티컬 거대언어모델(LLM) 개발
  - 우리은행: 금융 특화 생성형 AI 서비스를 준비 중, 비정형데이터 자산화해 금융언어모델에 적용
  - 농협은행: 생성형 AI 기반 영업점 AI 은행원 서비스 도입 고려, 내부직원 업무 효율화
- 외국 은행
  - JP모건: AI 투자 상담사 서비스 '인덱스GPT'의 상표 출원
- 스타트업
  - 아티피셜소사이어티: AI 디지털 교과서 대응 'eduGPT' 서비스 출시
  - 뽀빠이테크놀로지스(챗뽀빠이: 챗GPT+하이퍼클로바) : 사람의 글쓰기를 돕는 서비스 '휴먼 크리에이티비티(Creativity)'
  - 스켈터랩스(AIQ+Chat: 온프레미스, SaaS)
- 거대언어모델 서비스
  - OpenAI의 ChatGPT, HyperCLOVA X의 cue, Google의 Bard 등

## 2. 목적 및 목표

- 프로젝트 추진 목적: 현존하는 직원전용 수신 및 여신 전문 상담업무 서비스와 챗봇 한계를 생성형 AI를 통해 극복할 수 있는 서비스를 제안하기 위함.
  - 대구은행 행원 전용 챗봇인 'IM챗봇' 존재하지만 생성형 AI 서비스가 아닌 등록된 지식 기반의 챗봇이기에 질문의 의도를 정확히 파악하지 못하고 원하는 대답을 주지 못하는 문제점이 있음.
  - 'IT119' 직원 전용 여,수신 전문 상담원 존재하지만 한정된 상담사 수로 인해 긴 대기시간이 발생하여 빠른 업무처리가 되지 않는 문제점이 있음.
- 프로젝트 필요성:
  - 개인정보 유출 방지하는 AI 모델 필요
  - 사전학습 모델에 금융(은행)도메인 데이터를 학습한 특정 분야 거대언어모델 필요
- SWOT 분석
  - STRENGTHS(강점): 업무 효율성 증가, 비용 절감, 보안
  - WEAKNESSES(약점): 기술적 한계, 예산 불명확
  - OPPORTUNITIES(기회): AI시대로 진입, 시중은행화, 확장성
  - THREATS(위협): 규제 및 법적 제약사항, 기술 변화 및 업데이트 필요성
- 프로젝트 목표: 보안과 신뢰성을 보장하고 한국어를 잘 이해하여 질의응답이 가능한 한국어 특화 모델로 고객에게 양질의 서비스(빠른 업무 처리, 정확한 질문 의도 파악과 답변 생성)를 제공하는 생성형 AI 기반 서비스



## 2

## 프로젝트 수행 방법

### 1. 팀 구성

- 팀장 황세진
  - 환경책임, 대표 의사결정 담당
- 팀원 백두현
  - 실행과 일정관리 담당
- 팀장 박태원
  - 프로젝트 위기관리(위기 감지 및 해결), 디자인 의사결정 담당

※ DGB대구은행 멘토 - 손현오 계장: 생성형AI 및 금융 데이터 관련 자문 수행

### 2. 수행 일정

○ 담당자	📅 1주차(2023/09/18~09/24)	📅 2주차(2023/09/25~10/01)	📅 3주차(2023/10/02~10/08)	📅 4주차(2023/10/09~10/15)
황세진(팀장)	- 대규모언어모델(LLM) 외부 동향 조사 - 대구은행 현행 조사 - 프로젝트 주제 선정 - 사전 학습	- 현행 AI 학습모델 분석 - OpenAI의 Playground 활용 - 외부 멘토와 멘토링	- 금융 도메인 학습 가능한 생성형 AI 학습모델 기획 (Polyglot 사전학습 모델 선정) - 대구은행 멘토(손현오 계장) 회의	- 지속적인 모델 학습 및 파인튜닝
박태원	- 대규모언어모델(LLM) 외부 동향 조사 - 대구은행 현행 조사 - 프로젝트 주제 선정 - 사전 학습	- 학습 데이터 조사 및 설계 (XML파일 → json형식) - 외부 멘토와 멘토링	- AWS(Amazon SageMaker)를 활용한 Llama2 모델 학습 - 업무 활용 방안 구상 - 대구은행 멘토(손현오 계장) 회의	- 기획서 및 발표 자료 작성
백두현	- 대규모언어모델(LLM) 외부 동향 조사 - 대구은행 현행 조사 - 프로젝트 주제 선정 - 사전 학습	- 아이디어 도출 - 법령정보데이터 openAPI 승인 - 외부 멘토와 멘토링	- Gradient를 활용한 Llama2 모델 학습 - 학습 결과 확인 및 파라미터 조정 - 대구은행 멘토(손현오 계장) 회의	- 발표 준비 및 발표 자료 작성

### 3. 기타 수행 방법

- 한컴오피스: 회의록 작성
- OpenAI Playground(ChatGPT): GPT모델 프롬프트 엔지니어링
- 썬트렌드: 외부 동향 조사
- Discord, Zoom, Google Meet: 비대면 회의 툴(채팅, 음성, 화면 공유)
- 미리캔버스: 대구은행 발표회 발표 자료 작성
- AWS(Amazon SageMaker): LLAMA2 모델 파인튜닝

- Google Colaboratory: Polyglot-ko 12.8B 모델 파인튜닝
- Google Sheets: 모델 파라미터 및 성능 기입 및 확인

## 1. 소개 및 구성도

### ○ 프로젝트 소개

- 대구은행의 디지털 전환 가속을 위한 AI 기반 서비스 제안 프로젝트로 3명의 팀원으로 이루어진 비긴어게인 팀이 한 달간 수행함. 황세진은 AI 모델 코드 구현을 맡았으며, 백두현은 아이디어 도출과 도메인 지식 역할을, 박태원은 모델 탐색 및 학습 데이터 설계 역할을 함.
- 오픈 소스 거대언어모델(LLM)을 기반으로 한 생성형 AI 모델을 파인튜닝하는 프로젝트를 진행함.

### ○ 생성형 AI 모델의 파인튜닝 프로세스 Workflow



## 2. 기능 및 기술

### ○ 기능

- 은행의 고유 데이터를 파인튜닝(기존 학습되어 있는 모델에 추가 학습)하여 원하는 답변을 해주는 기능
- 오픈 소스 거대 언어 모델(LLM) Polyglot-ko 12.8B 모델을 활용한 생성형 AI
  - TUNiB에서 수집한 1.2TB의 한국어 데이터를 수집한 한국어 특화모델
  - 오픈 소스 모델 파인 튜닝으로 보안과 신뢰성 가능
- 업무 생산성 향상을 위한 직원전용 챗봇 서비스와 고객 만족도 향상을 위한 고객 전용 금융 AI 서비스에 쓰일 모델의 문장 생성 기능

### ○ 기술

- 자연어 처리(토큰화): Huggingface의 AutoTokenizer로 사전 학습 모델의 토큰나이저 불러와서 토큰화
- Fine tuning: 기업 내부 데이터를 사전 언어 모델에 추가 학습하여 기업 전용 모델을 생성함으로써 질문에 대한 기업에 특화된 답변 능력 향상
  1. PEFT(Parameter Efficient Fine-Tuning)  
: 모델의 전체 파라미터를 다 학습 시키지 않고 매우 적은 양의 파라미터만 학습하여 빠른 시간 내에 새로운 문제를 거의 비슷한 성능으로 학습시키는 방법.  
계산 및 저장 비용을 크게 줄임.
  2. QLORA(Efficient Finetuning of Quantized LLMs)  
: 4-bit로 양자화된 모델을 LoRA 기법을 활용해 파인튜닝하는 방법으로, NF4(4-bit NormalFloat)과 이중 양자화(Double Quantized), 페이지드 옵티마이저(Paged Optimizer) 등을 사용하여 메모리 소모를 크게 줄임.
  3. 하이퍼 파라미터 조정(생성 옵션)
    - max\_steps: 도메인 특화 데이터 훈련 단계 수 조정
    - temperature: 환각(Hallucination) 낮춤
    - logging\_steps: 업데이트 단계 수 조정
    - maximum length: 생성할 토큰 수 길이 조정

### 3. 문제점 및 해결방안

#### ○ 모델 선정

- 보안 문제를 해결하기 위해 로컬 모델로 Llama2를 쓰려고 했지만 GPU와 파인튜닝 과정에서 문제가 발생함. → 해결: Polyglot-ko을 대체 모델로 지정하여 프로젝트를 진행함.

#### ○ 환각현상(Hallucinations)

- AI가 사실이 아닌 오류에 대해 사실인 것처럼 응답 → 해결: 학습 데이터의 품질 향상, 반복적인 개선 및 테스트, Temperature 생성 옵션 환각 낮게 설정

#### ○ 데이터 수집

- 대구은행 고객 데이터 수집에 대한 어려움 → 해결: 프로젝트에서는 국가법령 정보 센터의 현행법령(소관부처: 금융위원회) 데이터와 법령정의사전 데이터를 사용했음. 향후 은행에서 생성형 AI 서비스 도입 시 행내 약관 및 상품설명서를 사전 언어 모델의 다운스트림(도메인) 데이터로 권장

### 4. 결과물

#### ○ 코드 결과물

##### - 학습

```
data = data.map(
    lambda x: {
        'text':
            f"""User: {x['법령용어명']}

Assistant: {x['법령용어상세검색']}<|endoftext|>"""
    }
) #instruction, input, output
```

Map: 100%  1000/1000 [00:00<00:00, 7765.04 examples/s]

```
[row['text'] for row in data['train']]
```

```
['User: (Instrument Meteorological Condition: IMC)Assistant: 시계비행 기상상태외의 기상상태를 말한다. <|endoftext|>',
 'User: (Terminal Operating Company)라 한다)Assistant: 항만시설운영자와 임대계약을 체결한 자(이하 "임대계약자"라 한다)로서 선석·보관시설·하역시설 등 부두시설 일체를 전용(專用)사용하는 민간업체를 말한다. 다만, 컨테이너부두와 개발부두의 임대계약자, 항만시설운영자가 인정하는 임대계약자는 제외한다. <|endoftext|>',
 'User: (이하Assistant: 인터넷을 통한 화확물질 불법유통 및 폭발물 제조방법 등 유해정보를 모니터링, 신고하는 활동에 자발적으로 참여하도록 위촉된 자를 말한다. <|endoftext|>',
 'User: (지리적)위치(Position geographical)Assistant: 지구 표면상의 특정 지점에 대한 위치를 규정하는 수학적 기준단원체를 기준으로 표시한 일련의 좌표집합(위도 및 경도)을 말한다. <|endoftext|>',
 'User: 0 정지방식 비상정지장치Assistant: 액츄에이터(actuator)의 전원차단방식에 의하여 정지되는 비상정지장치를 말한다.12. 사출성형기의 "주요 구조부"는 다음 각 목과 같다.가. 형체기구 및 사출기구나, 프레임(frame)다. 방호장치라. 유·공압 계통다. 제어반 <|endoftext|>',
 'User: 0 정지방식(stop category 0) 비상정지장치Assistant: 액츄에이터(actuator)의 전원차단방식에 의해 정지되는 비상정지장치를 말한다. <|endoftext|>',
 'User: 080착신과금서비스Assistant: 착신과금서비스를 제공하는 기간통신사업자를 말한다. <|endoftext|>',
 'User: 080착신과금서비스Assistant: 「전기통신사업법」 제2조제11호에 따른 기간통신업무 중 공통서비스 식별번호 080을 사용하여 수신자요금부담서비스를 제공하는 서비스를 말한다. <|endoftext|>',
```

## - 결론

```
from transformers.generation import GenerationConfig

gen_cfg = GenerationConfig.from_model_config(model.config)
gen_cfg.pad_token_id = tokenizer.eos_token_id

def gen(x):
    gen_cfg.temperature = 0.1
    gen_cfg.max_new_tokens = 256

    gened = model.generate(
        **tokenizer(
            f"User: {x}\n\nAssistant:",
            return_tensors='pt',
            return_token_type_ids=False
        ).to('cuda'),
        do_sample=True,
        generation_config=gen_cfg
    )

    result = tokenizer.decode(gened[0])
    answer = result.split('Assistant:')

    return answer[1].strip()
```

## - 결과

```
gen('한글날은?')
```

매년 10월 9일을 말한다. 한글의 우수성을 널리 알리고 세종대왕의 성덕과 위업을 추모하기 위한 기념일이다. 1926년에 음력 9월 29일로 지정한 '가갸날'이 그 시초이며 1928년에 명칭을 '한글날'로 바꾸었다. 1949년에 '관공서의 공휴일에 관한 규정'에 따라 공휴일로 공휴일에 관한 규정'에 따라 공휴일로 지정되었으며, 2005년에 국경일로 지정되었다. 한글날을 전후한 주간에 정부·지방자치단체·민간단체 등이 주관하는 각종 기념행사를 개최한다. 한글날의 제정 및 공포는 일제강점기 이후의 일이지만, 한글 창제를 기념하는 날로서의 한글날은 민족사에서 그 의미가 깊다. 한글날은 훈민정음의 반포 일을 기준으로 삼아야 한다는 주장도 있다. 한글날은 1926년에 음력 9월 29일로 지정한 '가갸날'이 그 시초이며 1928년에 명칭을 '한글날'로 바꾸었다. 1949년에 '관공서의 공휴일에 관한 규정'에 따라 공휴일로 지정되

gen('080착신과금서비스와 080착신과금사업자에 대해 알려주세요')

\*080착신과금서비스는 080번호를 이용하여 전화를 걸면, 전화를 받는 사람이 요금을 부담하는 서비스를 말한다. 080착신과금사업자는 080착신과금서비스를 제공하는 사업자자를 말한다. 080착신과금서비스를 제공하는 사업자는 080착신과금서비스를 제공하는 사업자자를 말한다. 080착신과금서비스를 제공하는 사업자는 080착신과금서비스를 제공하는 사업자자를 말한다. 080착신과금서비스를 제공하는 사업자는 080착신과금서비스를 제공하는 사업자자를 말한다. 080착신과금서비스를 제공하는 사업자는 080착신과금서비스를 제공하는 사업자자를 말한다. 080착신과금서비스를 제공하는 사업자는 080착신과금서비스를 제공하는 사업자자를 말한다.

```
gen('3종 금융투자업자?')
```

“다음 각 목의 어느 하나에 해당하는 금융투자업자를 말한다.가. 집합투자업자나, 신탁업자다. 투자매매업자라. 투자중개업자 1) 집합투자업자는 다음 각 목의 어느 하나에 해당하는 자를 말한다.가) 「자본시장과 금융투자업에 관한 법률」 제9조제1항제1호에 따른 투자매매업자다. 「자본시장과 금융투자업에 관한 법률」 제9조제1항제2호에 따른 투자중개업자다) 「자본시장과 금융투자업에 관한 법률」 제9조제1항제3호에 따른 투자일임업자라) 「자본시장과 금융투자업에 관한 법률」 제9조제1항제4호에 따른 투자신탁업자다) 「자본시장과 금융투자업에 관한 법률」 제9조제1항제5호에 따른 투자유한책임사원(투자유한책임회원)2) 신탁업자는 다음 각 목의 어느 하나에 해당하는 자를 말한다.가) 「자본시장과 금융투자업에 관한 법률」 제9조제1항제1호에 따른 투자매매업자나) 「자본시장과 금융

```
gen('1인 자연업자?')
```

[illegible]

### 1. 질문: '한글날은?'

- 질문 의도: '한글날' 단어를 추가 학습하진 않았지만 사전 학습 모델이 잘 구축되어 있는지 보기 위함.

#### - 결과:

- 잘한 점: 세 번째 문장까지는 정확하게 대답함으로써 사전학습 모델이 어느 정도 잘 구축되어 있음을 알 수 있음.
- 못한 점: 네 번째 문장부터 틀린 내용을 포함하고 있음. (환각 증상으로 예상)

### 2. 질문: '080착신과금서비스와 080착신과금사업자에 대해 알려주세요'

- 질문 의도: 두 개의 질문을 하나의 질문으로 합쳐 질문했을 때 잘 대답하는지 보기 위함.

#### - 결과:

- 잘한 점: 두 개의 질문에 각각 정확하게 대답함.
- 못한 점: 추가적으로 학습을 시킨 데이터에 출력 확률을 높인 대답인지 모호함.

### 3. 질문: '1인 자영업자?'

- 질문 의도: 추가 데이터를 잘 학습했는지 보기 위함.

#### - 결과:

- 잘한 점: 추가로 학습시킨 데이터를 의도에 맞게 잘 대답함
- 못한 점: 대답이 짧을 때는 같은 말을 반복함.  
»(해결) no\_repeat\_ngram\_size 옵션을 추가함으로써 반복을 안 하도록 설정

## 5. 프로젝트를 통해 느낀 점

#### ○ 프로젝트 개선 방안

- 학습 데이터의 양과 질을 고려하여 학습을 진행할 필요가 있음
- '단어'로 학습시키기보다는 자연스럽게 '질문'으로 변형해서 말뭉치 데이터 구축하기 (다른 거대언어모델 이용 시 비식별화 필수)
- 생성시킬 때 프롬프트 엔지니어링을 통해 자연스러운 답변이 되도록 하기

## ○ 인사이트

- 금융권 신기술/신사업 트렌드 분석을 통한 금융 업계의 디지털 트랜스포메이션 (DX) 흐름 파악
- 데이터의 품질과 중요성에 대한 깨달음
- 보안 및 규제 문제에 대한 중요성 인식

## ○ 느낀점

- ChatGPT와 같은 편리한 서비스가 나오기 위해 전문가들이 얼마나 고심했을지 조금이나마 이해할 수 있었음.
- 신기술에 대해 관심이 있었는데 이번 프로젝트를 계기로 깊게 연구하고 싶어졌음.
- 실제 업무에 적용할 수 있는 프로젝트를 해보면서 사내 업무를 간접적으로 체험해 볼 수 있어서 좋았음.
- 환각 현상 문제의 중요성을 느낄 수 있었음.
- 우리의 프로젝트를 통해 폐쇄된 환경에서의 개인정보 유출 방지 등 보안 가능성에 대해 확인할 수 있었음.
- gpt4, 라마2 그리고 하이퍼클로바X와 같이 잘 알려진 모델뿐만 아니라 Polyglot-ko 등 다양한 모델을 공부하면서 모델 각각의 특징을 이해하고 비교할 수 있게 되었음.
- 은행 산업이 생성형 AI 도입하는 데에 있어 고민하는 부분을 알게 되었고 같이 생각해 볼 수 있었음.
- 미래내일 일경험 사업으로 취업 준비생임에도 원하는 기업과 프로젝트를 진행하는 소중한 기회를 가지게 되었음.
- 프로젝트 지원금을 받아 colab 등 소프트웨어 및 도서 구입에 대한 비용 걱정 없이 프로젝트를 진행할 수 있었고 참가기업인 대구은행에 가서 발표하는 등 유의미한 경험이 되었음.



## 1. 기대효과

- 직원 업무 효율성 증대
- 고객 니즈 만족
- AI 시대로의 진입
- 금융권 신기술/신사업 트렌드 분석을 통한 DX인식 제고
- 비용 절감
- 고객 데이터 보안

## 2. 활용분야

- 업무 적용(활용) 방안
  - 초개인화 서비스
    - : 고객 정보 데이터를 학습한 특화 모델을 통해 생성 모델에 장기 기억 메모리로 활용함. >> 나만의 자산관리, 맞춤형 대출 상품 추천 등 초개인화된 서비스 구현
  - 음성인식을 활용한 고객 맞춤 업무
    - : 거대언어모델 파인튜닝에 쓰일 추가 데이터로 텍스트로 변환한 고객 상담 데이터를 이용. 실시간 상담 내용을 ASR(자동 음성 인식)을 통한 질문(프롬프트)으로도 입력 가능.
  - 업무 관련 여러 형태 데이터 적용
    - : 웹의 XML형식 뿐만 아니라 PDF, CSV 등 다양한 파일 데이터도 적용 가능
- 활용 분야
  - 직원 전용 생성형 AI 챗봇 서비스
  - 고객 전용 생성형 AI 챗봇 서비스, AI 은행원 서비스

※ 도메인 특화 데이터에 따라 은행권 뿐만 아니라 전 산업에 활용 가능함.

- 활용 효과
  - 직원 업무 생산성 향상(빠른 업무 처리)
  - 고객 만족도 향상

### 3. (필요 시) 구체적 활용실적

○

-

# 저작권, 특허 등록, 판촉, 기업에서 활용 등

## 향후 계획

## 1. 향후 보완 계획 및 진로 방향

- 프로젝트 보완 계획
  - 데이터 주기적 업데이트(추가 및 보완)
  - 추가적인 생성형AI와 거대언어모델 공부
  - 웹과 앱에서 구현
- 진로 방향
  - 금융권에서 생성형 AI 서비스 개발 업무에 적용하기 위한 프로젝트로 활용
  - IT/데이터 분석 진로 프로젝트로 활용
  - 뉴스 기사 활용: <https://n.news.naver.com/article/003/0012159244?sid=102>

**NEWSIS**

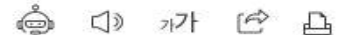
## DGB대구은행, 2023 미래내일 일경험 사업 성료

입력 2023.10.20. 오후 3:18 기사원문

 박준 기자

3

댓글



DGB대구은행, 2023 미래내일 일검험 사업 성료 \*재판매 및 DB 금지

[대구=뉴시스] 박준 기자 = DGB대구은행은 청년 일경험 수요 증가에 대응해 민관협업을 통한 다양한 일경험 기회를 제공하는 2023 미래내일 일경험 사업에 참여, 1개월간의 프로젝트를 성료했다고 20일 밝혔다.

기사 일부 내용