CHOI GAPJU

최갑주 AI Engineer

교육 운영의 문제를 검색·LLM 연동과 데이터 처리, 내부 API로 해결해 온 개발자입니다.
FAQ 검색 방식을 비교하는 평가 스크립트, 출결·설문·채팅 기반 위험도 산출, 운영 이슈 기록·공유 시스템을 구현했습니다. 현업 담당자와 데이터 취합 절차를 협의하고 결과를 실제 업무에 연결한 경험을 바탕으로, LLM 응용 서비스와 데이터 기반 업무 도구를 개발합니다.

xchlrkqwnx@naver.com · 010-9122-1038 · GitHub

Python · SQL · FastAPI · PostgreSQL · BigQuery · Git

(주)멋쟁이사자처럼 · Manager · 정규직

2024.07 – 2026.08

교육 운영 업무 및 데이터·AI 기반 업무 개선 프로젝트 수행

Selected Projects

한국어 임베딩 검색 · RAG 품질 평가 전담 개발

기술스택
Python · ko-SBERT · Pinecone · Claude API
2025.12 – 2026.02

구어체·오탈자 등 표현 변화에 취약한 FAQ 검색을 의미 검색과 근거 기반 답변 생성 구조로 재설계했습니다.

기존 키워드·문자열 유사도 검색과 임베딩 검색 비교 평가

문제·구현
구어체·답변본문·축약 질의에서 기존 검색이 놓치는 사례를 확인했습니다. FAQ 68건을 ko-SBERT로 임베딩해 Pinecone에 인덱싱하고, top-K 검색 결과를 Claude API의 답변 근거로 연결했습니다.

평가 방법
7가지 질문 유형의 초기 10건으로 두 검색 방식을 비교했습니다. 정답 FAQ의 질문 텍스트와 일치하거나 문자열 유사도 0.65 이상이면 적중으로 판정하고 Hit@K·MRR과 질의별 결과를 기록했습니다.

결과·범위
해당 판정 기준에서 Top-1 7/10 → 10/10, MRR 0.75 → 1.00을 기록했습니다. 소규모 검색 평가 결과이며 생성 답변의 정확도와는 구분합니다. 답변에 faq_id·카테고리를 표시하고 외부 정보 사용을 제한했습니다.

행동 데이터 기반 이탈 위험 스코어링 전담 개발

기술스택
Python · Pandas · Scikit-learn
2026.02 – 2026.05

출결·설문·채팅 신호를 통합해 수강생의 이탈 위험을 산출하고 운영진의 면담 우선순위를 제공했습니다.

특정 기수 47명 중 잠재 위험군 7명 추가 식별 · 면담 지원

데이터 확보·전처리
부서별 고객 정보 접근 제약을 해결하기 위해 데이터 책임자와 취합·처리 절차를 협의했습니다. Slack 텍스트 파싱, 학번 정규화, 설문·출결 조인을 자동화했습니다.

스코어링 설계
출결·설문·채팅 점수를 가중 결합하고 4단계 위험 구간을 구성했습니다. 채팅 참여가 적다는 이유만으로 위험도가 과대 반영되지 않도록, 채팅 단독 위험 신호에는 점수 상한을 적용했습니다.

운영 적용·평가 범위
특정 기수 47명 중 출결 기준에서 놓친 잠재 위험군 7명을 추가 식별하고 Danger 4명의 면담을 지원했습니다. 당시 실제 이탈 양성 라벨이 없어 예측 정확도가 아닌 면담 우선순위 산출 결과로 활용했습니다.

Selected Projects

데이터 통합 · 품질 관리 파이프라인 TF · 팀 수행

기술스택
BigQuery · PostgreSQL · Looker Studio
2025.12 – 2026.03

11개 소스를 통합하는 TF에서 고용24 API·교육기획 부서 시트를 수집·정제하고 BigQuery와 KPI 대시보드를 연동했습니다.

TF 성과: 리포트 확인 3~5일 → 30분 이내

> **직접 담당 · 수집·정제**
> 고용24 API로 교육 회사별 IT 직무교육 수강 리스트·인원을 시트에 수집(Bronze)하고 정제(Silver)했습니다. 교육기획 부서의 수강 현황 시트에서 단위기간별 출석률과 강의별 수료 인원을 정리했습니다.
>
> **직접 담당 · 데이터 활용**
> 정제한 데이터를 BigQuery에 연동하고 KPI에 맞춰 Looker Studio 대시보드로 연결했습니다.
>
> **TF 공동 결과**
> 일일 ETL과 통합 리포팅 체계를 구축하고 Data Steward 역할, 데이터 품질 체크리스트·갱신 SOP를 문서화했습니다.

운영 이슈 기록·공유 API 시스템 전담 개발

기술스택
Python · FastAPI · PostgreSQL · Slack API
2025.02 – 2025.04

구두로 전달되던 교육 운영 이슈를 기록·저장하고 Slack에 자동 공유하는 내부 시스템을 개발했습니다.

구두·사후 보고 → 조회·추적 가능한 당일 자동 공유

> **API·업무 연동**
> FastAPI·PostgreSQL로 이슈 입력·저장 흐름을 구현하고 Slack API를 연결했습니다. 담당자와 의사결정자의 일정 조율을 줄이고, 누락 가능성이 있던 구두 전달을 기록 가능한 업무로 전환했습니다.

프로그래머스 인공지능 데브코스 6기 · K-Digital Training

OCR 기반 가격표 정보 추출·매핑 팀 프로젝트

기술스택
Python · Naver CLOVA OCR · Roboflow · React
2023.12 – 2024.01

촬영된 가격표에서 상품명과 가격을 추출·매핑하는 오프라인 최저가 비교 서비스를 개발했습니다.

가격표 영역 검출 · OCR 좌표 기반 상품명–가격 매핑

> **담당 역할·구현**
> 이미지 노이즈 제거, Roboflow API 기반 가격표 영역 검출, CLOVA OCR 연동과 React 화면을 담당했습니다. OCR 결과의 y좌표로 상품명·가격을 그룹화했고, 기울어진 사진에서는 같은 행도 y좌표가 달라지는 한계를 확인했습니다.

복습 튜터용 음성 생성 PoC 사내 개인 탐색 프로젝트

기술스택
Python · PyTorch · Qwen3-TTS · Whisper · Gradio
2026.03

수강생 복습 튜터에 사용자 목소리로 수업 요약을 읽어주는 기능의 적용을 검토하며, 음성 생성 부분을 개인 PoC로 구현했습니다.

참조 음성 5초 추출 → 대본 자동 인식 → 한국어 음성 생성

> **구현 범위**
> Claude Code를 활용해 Qwen3-TTS 1.7B Base·Whisper-small을 연결하고 참조 대본 입력을 자동화했습니다. librosa 전처리, Gradio 업로드·녹음·재생과 Mac MPS/CPU 추론을 구현했습니다. 요약 시스템 연동 전의 음성 생성 PoC입니다. 기반 모델: Qwen3-TTS Technical Report (2026).

Activities

BearStew ML 논문 리뷰 스터디 Notion 2024.01 – 2026.03

- 매주 논문 리뷰와 발표에 참여해 구성원들과 총 65편의 논문을 공동 리뷰했습니다.

PyTorch Korea 오픈소스 컨트리뷰터 2024.07 – 2024.11

- PyTorch 공식 NLP 튜토리얼 한국어 번역에 기여하고 오픈소스 코드 리뷰 워크플로를 경험했습니다.

프로그래머스 인공지능 데브코스 6기 2023.08 – 2024.01

- K-Digital Training ML/DL 집중 과정을 수료하고 OCR 기반 가격 비교 서비스를 개발했습니다.

Education

전남대학교 글로벌비즈니스학부 국제통상학전공 / 산업공학 부전공

2018.03 – 2026.08