강민규

0000.00.00 (만 00세)

Phone 010-0000-0000 | Email kanghy6220@gmail.com

Address - | GitHub https://github.com/RainDrop3

학력

2019.03 – 2022.02	화봉고등학교 졸업
2022.03 – 現	부산대학교 정보컴퓨터공학부 3학년 재학 학점 0.00/4.5

병역사항

2023.02 – 2024.08	육군 병장 만기전역

수상내역

2025.12 – 2026.03	**PNU AI BOOSTER 우수상** - Frontend, 4인팀. 원칙 준수를 돕는 주식 투자 훈련 플랫폼 개발 - Next.js로 회원가입페이지, 마이페이지, 사이드패널 등 제작
2026.04 – 2026.08	**제7회 PNU 창의융합AI해커톤 우수상(DRB동일대표이사상)** - Backend, 5인팀. 도심농부와 공실 제공자를 중개하는 스마트팜 중개서비스 개발 - Swagger 문서 구축, RESTful API 설계
2026.08	**카카오테크캠퍼스 아이디어톤 우수상** - 6인팀. 개인사업자를 위한 종합소득세 경비 판정 서비스 기획 - 에이전트 사용성 및 RAG 기반 데이터 해자 확보 아이디어 기획

프로젝트

2026.04 – 2026.06

갤러리 사진의 카테고리 분류 앱 개발

- TensorFlow Lite 활용, 구글의 EfficientNetB0모델을 특징추출 - 미세조정의 단계로 이미지 2912장 전이학습. 검증 0.94, 테스트 0.90의 정확도 도출.
- RandomZoom, RandomContrast 등으로 데이터 증강하여 과적합 최소화.
- 특징추출 단계에서 베이스 모델 동결 후 0.1%학습률로 30epoch 학습. 이후 미세조정 단계에서 베이스의 상위 30개 층을 풀고 0.001%학습률로 추가 학습.
- AI가 epoch초기값으로 20을 추천하여 20epoch로 학습 진행. 성능이 기대에 못 미쳐서 30으로 늘려보니 검증 loss값이 0.0860 -> 0.0769로 약 10%, 검증 accuracy값이 0.9062 -> 0.9353로 약 3% 개선되는 모습을 보임.
- 카테고리 분류를 위해 최초에는 softmax + sparse_categorical_crossentropy를 사용. 하지만 기존 카테고리 클래스들로는 분류가 어려운 사진도 어떻게든 분류할려고 틀린 클래스로 매핑하는 경향성 확인. 따라서 sigmoid + binary_crossentropy로 변경하고 'Others'카테고리를 추가하여 모든 클래스의 확률이 0.65미만이면 Others로 판정하도록 설정.

2026.09 – 2026.11

개인사업자용 종합소득세 경비 판정 서비스 개발

- AI Lead, 6인팀.
- 국가법령정보 OPEN API로 소득세법 등 법령, 행정규칙, 심판례, 해석례 원문 약 35500건 수집 후 55500개 청크로 임베딩 색인. 조문 구조를 테이블로 다루면서 임베딩도 함께 저장하고 백엔드와 같은 DB를 쓰도록 PostgreSQL(pgvector) 선택. EC2 systemd timer로 매주 바뀐 조문만 재색인하고 변경 요약과 영향받는 룰카드를 Discord로 알리는 코퍼스 자동 최신화 파이프라인 구축.
- 판정 못한 항목의 근거 조문을 검색해 룰카드 초안을 만드는 RAG 규칙 후보 생성 시스템 구축. 검색 기법으로 임베딩 벡터 검색과 pg_bigm 키워드 검색을 RRF로 결합한 하이브리드 검색 채택. 벡터 단독 검색과 비교 실험한 결과, 키워드 검색이 접대비 -> 기업업무추진비처럼 법의 개정으로 바뀐 용어를 잡아서 적합한 조항 인용이 15/60회 -> 40/60회로 증가함을 확인.
- 기존 분류기로 업종 분류하지 못한 가맹점의 업종 추천 모델 연결. Jev와 GPT-6 Luna(/decisions API)를 상권정보 표본 등 3개 평가셋 837건으로 비교, 분류가 어려운 505건에서 top1 정확도가 각각 0.366, 0.414으로 앞선 Luna 채택. 자동 확정 대신 되묻기 화면에 top-3 추천.
- tool 13개를 쓰는 tool-calling 기반 에이전트 비서 AI 서버 개발. Langfuse로 LLM 호출 단계별 토큰, 비용, 지연 추적.
- 에이전트 비서 대상 프롬프트 인젝션 공격 6종(화면 값 주입, 대화 이력 위조, 가맹점명 지시문 주입, 숫자 검증 우회 등) 직접 실험. 시스템 프롬프트 보강과 허용된 화면 경로만 받는 입력 검증으로 6종 모두 차단 확인.

2026.07 – 現	**데이터 엔지니어링 학습 프로젝트** - DataTalksClub Data Engineering Zoomcamp 개인 학습. NYC 택시 운행 데이터로 모듈별 파이프라인 실습 - Docker Compose로 PostgreSQL과 pgAdmin을 띄우고 택시 CSV를 청크 단위로 적재하는 수집 스크립트 컨테이너화. Terraform으로 GCS 버킷과 BigQuery 데이터셋 생성 - Kestra로 PostgreSQL에 적재하는 ETL flow와 GCS, BigQuery에 적재하는 ELT flow 구성. 스케줄 트리거로 2019년 green 택시 데이터 backfill - dbt Core와 DuckDB로 2019~2020년 택시 데이터를 staging, intermediate, marts 세 계층으로 변환. 두 데이터의 컬럼 차이(trip_type, ehail_fee)로 생긴 union 오류 해결, 월별 지역 매출 리포팅 모델 작성

🏆 동아리 및 대외활동

2025.03 – 2025.08	**코드코치(교육봉사 연합동아리)** - 초등학생 대상 아두이노 교육봉사. 담당 학생 2명의 주간 학습지 채점과 질문 응답, 매주 수업 실습 보조 - 초음파 센서와 서보모터로 자동으로 열리는 쓰레기통 만들기 수업 1회 직접 진행, 학생 집중을 위한 퀴즈 직접 제작
2025.09 – 現	**PULSE(알고리즘 학술동아리)** - 정기적으로 AtCoder(온라인 PS대회)에 참가하여 문제해결 후 회고 공유 - Union-Find, 위상정렬 등 중/고급 알고리즘과 시간복잡도 계산법 학습 - 2025 ICPC 예선 참가
2026.05 – 2026.11	**카카오테크캠퍼스 4기 - Agentic AI course** - GPT-2 Transformer 구조를 통한 LLM 내부 동작 원리 학습 - LangChain 기반 일정 관리 에이전트를 6주간 단계별 구현. 일정 CRUD tool, Pydantic 구조화 출력, RAG 검색 tool, MCP tool 호출, supervisor가 하위 에이전트 2개에 작업을 위임하는 멀티 에이전트 구조 - 에이전트가 잘못된 tool을 고른 사례를 trace로 추적해 시스템 프롬프트 수정. 비슷한 tool끼리 사용 기준을 명시해 저장 일정 조회, 기억 검색, 지난 대화 검색의 tool 오호출 3건 해결