MakinaRocks

김정훈

Forward Deployed Engineer - LLM 지원

이즈파크 R&D Center · 서울
010-9559-9471 · bandlagom0927@gmail.com
portfolio.jhkim3ispark.work
github.com/thuygom

영업 데이터의 뜻을 정리하고, 모델의 답을 실제 조회 결과로 확인했습니다.

TieMing의 주력 AI 기능을 맡아 현업 데이터에서 먼저 무엇이 잘못 선택되는지 살폈습니다. CRM에서는 DB 설명과 테이블 관계를 보완하고 SQL 실행 결과로 모델을 평가했습니다. 명함 정보 추출은 지식 증류와 LoRA로 개선하고, 회의 음성은 업무 용어와 처리 경로를 제품에 맞췄습니다. 학습부터 API·권한·배포까지 직접 연결한 경험을 마키나락스의 FDE 업무로 이어가고자 합니다.

경력·학력

이즈파크 R&D Center · AI Researcher 2025.08–현재

일학습병행 인턴 2025.08–2026.02 / 정규직 2026.03–현재

대한민국 공군·정보체계 특기 전문하사 2021.12–2022.12 · 1년

Linux 서버 관리·쉘 스크립트·로그 분석

한성대학교 컴퓨터공학부·공학사 2019.03–2026.02

평점 3.95/4.5 · IRIS Lab 학부연구원 2024–2026

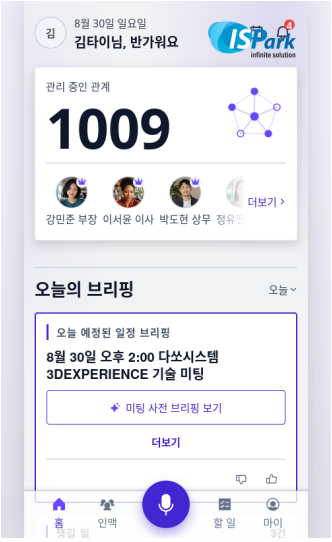
회사 내 사람·거래 정보를 연결하는 AI 제품

TieMing 애플리케이션의 **주력 AI 기능**을 담당하며 명함 정보 추출, 사람·회사 검색, 회의 음성 전사, 영업 데이터 자연어 조회와 음성 입력을 개발했습니다.

명함을 다시 입력하고 기록을 따로 찾던 업무를 줄일 수 있도록, 사람·회사 정보를 중심으로 AI 기능을 통합했습니다. 기능별 모델 개선부터 API 연결과 배포까지 직접 맡았습니다.

AI Summit Seoul 2026에서 공개 시연했습니다. Android 비공개 베타와 Play Store 등록을 준비하고 있습니다.

공개 데모는 내부 운영 데이터와 분리된 인스턴스입니다.



프로젝트 2–5쪽·논문 6쪽·특허·수상·자격 7쪽

영업 데이터 질문 조회

2026.06–현재 · 선행 자연어 조회 연구 · 특허 준비 2025.12–

모델 후학습으로 조회 결과 일치율을 26.6%에서 76.6%로 높였습니다.

영업 담당자가 자연어로 사내 데이터를 조회하도록 검색 · SQL 생성 · 실행 검증을 설계했습니다. 사내 약어와 누락된 테이블 관계를 보완하고, 질문에 필요한 데이터만 찾는 입력 구성을 마련했습니다.

관련 데이터 검색	SQL 생성 · 검증	사용자별 권한 적용

모델 학습과 입력 개선을 함께 맡았습니다

교사 모델의 응답으로 학습 자료를 구성하고, 학생 모델에 LoRA 후학습을 적용했습니다. 내부 CRM 64문항의 실행 결과 일치율을 기준 모델 26.6%에서 후학습 모델 76.6%로 개선했습니다.

내부 CRM 64문항	기준 모델	후학습 모델
실행 결과 일치율	26.6%	**76.6%**
기준 결과와 일치한 질문	17/64	**49/64**

같은 평가에서 +50%p 개선. 행 순서를 제외하고 숫자 · 날짜를 정규화해 비교했습니다.

모델을 전달하기 전 실행 예제를 재검증했습니다. SQL이 정상 실행돼도 잘못된 날짜를 집계할 수 있어, 불필요한 테이블 설명을 줄이고 입력 스키마와 예제를 수정했습니다.

모델의 조회 능력과 데이터 접근 권한을 분리했습니다

모델이 만든 SQL에 사용자 조건이 빠지더라도 권한 밖의 데이터가 반환되지 않도록 PostgreSQL Row Level Security(행 수준 보안, RLS)를 설계했습니다. 읽기 전용 실행, 허용된 표 · 항목과 조회량 검사도 함께 구성했습니다.

접근 정책 설계와 격리 환경 검증을 주도했습니다. 두 계정으로 같은 질문을 실행해 타인 행 노출 0건을 확인하고, 권한 회수 · 복원에 따른 결과를 검증한 뒤 운영 담당자와 실제 시스템 연결을 진행했습니다.

데이터 설명 보완 방법을 특허로 발전시켰습니다

조회 전에 표 · 항목 설명과 관계를 보완해 잘못된 SQL 생성을 줄이는 전처리 방법을 설계하고 특허를 출원했습니다. 모델 후학습, 입력 개선, 실행 검증을 각각 점검해 원인을 구분했습니다.

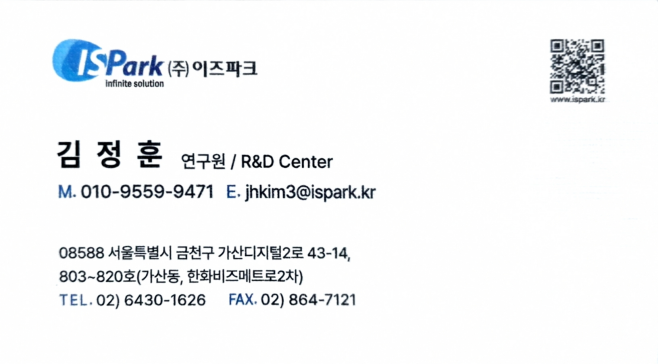
공개 데모는 내부 CRM 모델 평가와 별도로 구성했습니다.

명함 정보 추출

2026.04–현재 · 학습 데이터 구성 · 모델 개선 · 항목별 평가

실제 업무 명함 약 5,000장,
핵심 7개 항목 추출 정확도 94%.

명함을 앱에서 바로 활용할 수 있는 고객 정보로 바꾸기 위해, 모델 학습부터 항목별 평가 · 출력 보정까지 맡았습니다. 이름과 회사의 혼동, 잘못된 출력 형식은 서로 다른 원인으로 나눠 개선했습니다.



추출할 정보
이름 · 김정훈
소속 · 이즈파크 R&D Center
직함 · 연구원
연락처 · 전화 · 이메일

추출 모델을 실제 업무에 적용했습니다

실제 업무 OCR 적용 과정에서 약 5,000장 규모의 명함을 다뤘고, 핵심 7개 항목 정보 추출 정확도는 94%였습니다. 후학습 평가는 별도 검수 자료를 학습 · 조건 선택 · 고정 평가 세트로 나누고, ID와 이미지 해시 중복을 확인해 진행했습니다.

지식 증류와 LoRA로 추출 모델을 개선했습니다

문자 인식(OCR)과 핵심 정보 추출(Key Information Extraction, KIE)을 나눠 평가하고, 교사 모델의 지식을 학생 모델에 전달하는 지식 증류(Knowledge Distillation, KD)와 LoRA 미세조정을 직접 적용했습니다. 값 오류는 후학습으로 개선하고, 앱이 읽을 수 없는 형식은 해당 결과만 다시 생성하도록 설계했습니다.

별도 후학습 평가	기존	후학습	출력 보정
항목 값 정확도	30.2%	60.5%	**65.3%**

후학습과 출력 보정을 결합해 별도 평가의 항목 값 정확도를 30.2%에서 65.3%로 높였습니다. 항목별 변화도 추적해 추가 보완이 필요한 주소 추출을 다음 개선 대상으로 정했습니다.

학습 · 출력 조건은 별도 검증 자료에서 선택하고, 고정 평가 세트에 적용했습니다. 실제 업무 적용 지표와 후학습 평가 결과는 구분해 관리했습니다.

명함을 입력하고 항목별 추출 결과를 확인할 수 있습니다.

회의 음성 전사·사용자별 기억

2026.03–현재·한국어 평가·업무 용어 반영·작업 복구

한국어 회의 전사와 화자 분리를 제품에 적용하고, 장시간 음성 처리 성능을 개선했습니다.

음성 인식·화자 분리 모델을 비교·통합해 TieMing의 회의 기록 기능을 개발했습니다. 영업·제조 용어를 반영하고, 회의가 길어져도 처리 결과가 끝까지 남도록 실시간 세션과 장시간 작업을 분리했습니다.

한국어 업무 회의에 맞춰 품질을 평가했습니다

내부 한국어 회의에서 단어 오류율(WER) 약 19%, 화자 구분 오류율(DER) 약 23%를 확인했습니다. 별도로 화자 1–10명이 등장하는 고난도 공개 음성 10건을 클로바노트·리멤버노트와 같은 조건으로 비교했습니다.

고난도 공개 음성 10건	iScribe	클로바노트	리멤버노트
단어 오류율 (WER)	29.40%	25.94%	26.88%
화자 구분 오류율 (DER)	33.25%	25.71%	40.83%

각 약 8분·낮을수록 오류가 적음·내부 업무 회의 지표와 별도 평가

음성 처리시간을 36.9% 줄였습니다

말소리 구간 감지(VAD)와 배치 추론을 적용해, 30분 음성 1건의 서버 처리시간을 122.6초에서 77.4초로 단축했습니다. 기존 순차 처리와 같은 파일을 비교한 결과입니다.

서버는 처리를 끝내도 HTTP 요청이 먼저 만료되는 문제는 비동기 작업으로 해결했습니다. 단계별 진행 상태와 중간 결과를 저장하고, 재시작 뒤 이어서 처리하도록 복구 경로를 구현했습니다.

업무 용어와 사용자별 맥락을 다음 회의에 반영했습니다

사내 회의록에서 영업·제조 용어를 사전에 반영하고 사용자가 추가한 표현도 누적했습니다. Honcho 기반 사용자별 메모리로 개인 단어와 회의 맥락을 분리해 보존했습니다. 로그인 계정과 조직 권한에 따라 작업 접근도 제한했습니다.

체험 계정: demo / 비밀번호: demo

검색 근거·기록 재사용·운영

검색 근거, 작업 이력과 서비스 상태를 확인하는 시스템

AI가 검토할 근거를 보존하는 검색 API를 개발했습니다

AI가 인용할 자료를 추적할 수 있도록 검색 결과의 원문·출처·수집 시점을 함께 보존했습니다. 요약과 실제 페이지 내용이 다른 경우에도 원문을 다시 검토할 수 있는 검색 API를 개발했습니다.

사이트별 접근 정책과 요청 제한을 지켰습니다. 로그인이나 추가 인증으로 읽을 수 없는 페이지는 성공 결과에 섞지 않고, 접근 제한 상태로 기록했습니다.

개인의 개발 기록을 팀이 재사용하는 도구로 만들었습니다

개인 파일에 흩어진 AI 코딩 기록을 팀이 검색하고 재사용하는 기능을 개발했습니다. 사용자가 선택한 기록과 판단 근거를 출처와 함께 가져오고, 반복해서 참고할 내용은 팀 위키로 정리하도록 설계했습니다.

기록 선택	변경된 내용만 반영	검색·팀 문서 재사용

조직·팀·개인별로 볼 수 있는 기록을 제한하고, 원본 위치를 확인할 수 있도록 연결 정보를 보존했습니다.

모델 배포부터 GPU 운영까지 직접 구성했습니다

추론 API를 Docker 컨테이너로 배포하고 서비스·컨테이너·GPU 상태를 함께 모니터링하는 운영 환경을 구축했습니다. 지연이 발생했을 때 API 오류와 GPU 자원 경합을 구분할 수 있도록 계측했습니다.

학습 중단 뒤에도 이어갈 수 있도록 진행 상태를 저장했습니다. GPU 온도·전력을 30초 간격으로 기록하고, 82도에 도달하면 저장 후 학습을 멈추도록 구성했습니다.

NIPA A100 환경에서 자사 언어모델을 학습·검증했습니다

정보통신산업진흥원(NIPA)의 고성능 컴퓨팅 지원으로 A100 GPU 2장을 사용했습니다. 자사 경영·인사·재무 업무 데이터를 추가 사전학습한 뒤, 업무별 질문·응답으로 지도 미세조정을 진행했습니다.

학습 손실이 낮아도 실제 제품 질문에 맞는 답을 보장하지는 않았습니다. 내부 제품 질문 440개를 같은 조건에서 평가해 통과한 후보만 남겼습니다. SQL 학습에서도 평가 질문이 섞인 초기 시도를 폐기하고 학습 자료와 고정 평가 자료를 다시 분리했습니다.

사용 기술: PyTorch ·Transformers ·FastAPI ·PostgreSQL ·Docker

멀티모달 연구·논문

학부연구원 2024–2026 ·논문 5편, 제1저자 4편

입력 내용에 따라
정보 결합 방식을 선택하는 법을 연구했습니다.

댓글 감정·주제 분석에서 출발해 영상·음성까지 다루는 멀티모달 연구로 확장했습니다. 입력마다 유용한 정보가 다르다는 점에 주목해, 어떤 정보를 어떤 경로로 결합할지 학습하는 방법을 연구했습니다. 제1저자 4편을 포함한 논문 5편으로 정리했습니다.

논문 5편

멀티모달 정보 결합 경로 최적화·제1저자—Multimodal Routing Optimization via Neural Architecture Search of Fusion Paths

IJIBC 17(3), 144–154 ·KCI 등재·2025.08

KCI 원문

입력에 맞춘 정보 결합 경로 학습·제1저자—암시적 정책 학습 기반 멀티모달 경로 최적화

KSC 2025.12 ·장려상 2026.01 수상

DBpia 원문

소셜 콘텐츠에 대한 사용자 반응 분석·제1저자—멀티모달 인플루언서 지지도 추정을 위한 사용자 반응 분석

KIPS 2025.05

KoreaScience 원문

온라인 반응 기반 창작자 보호 신호·공동저자—창작자 보호를 위한 SNS 정서 리스크 지표화

KIPS 2025.11 ·ACK 2025 은상

KoreaScience 원문

댓글의 감정·주제 기반 인플루언서 인기 예측·제1저자—A Transformer-Based System for Predicting Influencer Popularity via Comment Sentiment and Topic Analysis

AICOMPS 2024

특허·수상·자격

연구와 개발 경험의 증빙

데이터 설명을 보완해
잘못된 조회문 생성을 줄이는 방법을 출원했습니다.

자연어 질문을 데이터베이스 조회문으로 바꾸기 전에 표·항목 설명과 관계를 보완하는 방법을 설계했습니다. 이 방법으로 특허를 출원하고 심사를 청구했습니다.

공식 출원 제목: 데이터베이스 설명 자동 생성 및 환각 억제 기능을 갖는 Text-to-SQL 전처리 방법

출원번호 10-2026-0119188 ·출원일 2026.06.30 ·발명자 김정훈

26. 6. 30. 오후 1:05 특허로

관 인 생 략

출 원 번 호 통 지 서

출 원 일 자	2026.06.30
특 기 사 항	심사청구(유) 공개신청(무)
출 원 번 호	10-2026-0119188 (접수번호 1-1-2026-0792313-96) [redacted]
출원인 명칭	㈜이즈파크(1-2014-041198-4)
대리인 성명	특허법인 영비(9-2018-100201-7)
발명자 성명	김정훈
발명의 명칭	데이터베이스 설명 자동 생성 및 환각 억제 기능을 갖는 Text-to-SQL 전처리 방법

지 식 재 산 처 장

수상

2026	AI 스마트워크톤 은상
2026.01	KSC 2025 장려상
2025.11	ACK 2025 은상
2024.11	한성 SW페스티벌 최우수상
2023.09	한성공학경진대회 대상

자격·병역

SW개발 L5(2026), 정보처리산업기사·리눅스마스터 2급(2022)을 취득했습니다.
공군 정보체계 특기 병으로 21개월 복무했습니다.