



지원동기

사내 영업 데이터를 자연어로 조회하는 기능을 개발하면서, 모델이 SQL을 실행하는 것과 사용자가 의도한 답을 얻는 것은 다르다는 사실을 확인했습니다. 실행은 성공했지만 다른 날짜 열로 매출을 집계한 사례가 있었습니다. 저는 모델을 다시 학습시키기 전에 DB의 약어와 테이블 관계, 질문에 전달하는 스키마 설명부터 살폈습니다. 잘못된 열 선택을 유발하던 정보를 줄이고 예제를 고친 뒤, 실제 실행 결과를 기준으로 평가했습니다.

이 경험이 마키나락스의 Forward Deployed Engineer - LLM 직무에 지원한 이유입니다. 공고는 현장 데이터를 분석해 문제를 정의하고, LLM/VLM과 검색·후학습을 실제 서비스에 적용하는 일을 제시합니다. 저는 TieMing 애플리케이션의 주력 AI 기능을 맡아 모델 학습과 사용자에게 결과가 도달하는 경로를 함께 개발해 왔습니다. 교사 모델의 SQL 데이터와 LoRA 후학습으로 내부 CRM 64문항의 실행 결과 일치율을 **26.6%에서 76.6%로 높였고**, 결과를 행 순서·숫자·날짜 표현을 정규화해 비교했습니다.

명함 정보 추출에서도 글자 인식과 항목 배정 오류를 나눠 보았습니다. 교사·학생 모델의 지식 증류와 LoRA 미세 조정을 적용하고, 형식이 깨진 출력만 다시 생성하도록 바꿨습니다. 업무에서 명함 약 5,000장을 처리하며 핵심 7개 항목의 정보 추출 정확도 **94%**를 확인했습니다. 별도의 고난도 후학습 평가에서는 항목 값 정확도를 **30.2%에서 65.3%로 높였습니다**. 두 수치는 서로 다른 평가라는 점까지 구분해 설명할 수 있습니다.

최근에는 정보통신산업진흥원(NIPA)의 고성능 컴퓨팅 지원으로 A100 GPU 2장에서 자사 경영·인사·재무 데이터 기반 언어모델을 학습했습니다. 추가 사전학습과 지도 미세조정을 거친 후보들을 내부 제품 질문 440개로 비교했습니다. 손실값이 낮아도 실제 질문 평가를 통과하지 못하면 채택하지 않았습니다.

모델의 답을 제품에 연결할 때는 실행 범위도 직접 설계했습니다. SQL 생성과 사용자별 조회 권한을 분리해 PostgreSQL의 행 단위 접근 제어와 읽기 전용 실행을 검증했습니다. 장시간 음성 처리에서는 요청 시간이 먼저 끝나는 문제를 비동기 작업·상태 저장·재시작 복구로 풀었고, 같은 30분 녹음 파일의 서버 처리 시간을 **122.6초에서 77.4초로 단축했습니다**. 앞서 공군 정보체계 특기 전문하사로 1년간 Linux 서버를 관리하며 쉘 스크립트와 로그 분석으로 운영 문제를 추적한 경험이 이 작업의 바탕이 됐습니다.

마키나락스에서 저는 현장 담당자가 실제로 확인하려는 결과와 데이터의 의미를 먼저 파악하고 싶습니다. 그다음 검색 오류, 모델 오류, 실행 오류를 나누어 작은 평가 집합으로 확인하겠습니다. 해결 방법을 학습이나 프롬프트 하나로 미리 정하지 않고, 고객의 업무 흐름에 맞는 시스템으로 만들며 반복해서 개선하겠습니다.