

# AI Engineer 이력서 - 김상겸

검색 시스템 구축 경험을 바탕으로, LLM과 RAG 기술을 학습하고 응용하는 프로젝트를 수행해 왔습니다. 새로운 기술을 도입하여 서비스의 사용자 경험을 실질적으로 개선하는 일에 보람을 느낍니다. 문제를 분석하고 기술적 해결책을 설계해 개선하는 데 강점을 가지고 있습니다.

Birthday: 1992.10.30

Email: [vhf1030@naver.com](mailto:vhf1030@naver.com)

Mobile: 010-9212-0252

Address: 서울특별시 관악구 인현6길 8

GitHub: <https://github.com/vhf1030>

## 기술 스택

언어	Python, R, Shell, Go, SQL
ML / DL	PyTorch, scikit-learn, LightGBM, XGBoost, YOLO, LLaVA, E5
LLM / RAG	LangChain, ChromaDB, FAISS, SBERT, OpenAI, Ollama, Gemma, Llama
데이터	Filebeat, Logstash, Kibana, pandas, numpy, matplotlib, seaborn
백엔드	FastAPI, Django, Elasticsearch, MySQL, Streamlit, WebSocket
운영 / 협업	Docker, Airflow, Grafana, Ray, AWS, Linux, Git, Notion

## 타임라인

기간	소속	역할 / 직무	직위 / 구분
2024.12 - 2025.06	원티드 포텐업	AI 부트캠프	교육생
2022.09 - 2024.08	포바이포	데이터 엔지니어	매니저
2022.06 - 2022.09	카카오	데이터 엔지니어	인턴
2019.02 - 2021.05	마크로젠	데이터 분석가	사원
2018.03 - 2018.07	에이콘 아카데미	ML 부트캠프	교육생
2011.03 - 2019.08	경희대학교	생물학과	학사

# 경력

## 원티드 포텐업

2024.12 - 2025.06 (6개월)

부트캠프 / 교육생

AI 실무 역량 강화를 위한 부트캠프

- LangChain RAG 기반 건설사고 재발방지 대책 모델  
활용기술: Python, RAG, LangChain, ChromaDB, SBERT, Ollama  
- 안전 매뉴얼 검색과 LLM을 통해 건설사고 사례별 재발 방지 대책 RAG 구축  
- 역할: 데이터 전처리, vector 임베딩 및 검색 파이프라인 구축, 로컬 LLM 서빙 환경 설정 및 연동  
- 성과: 문장 단위 계층적 검색기반 QA 시스템 설계 및 RAG 시스템 구현
- KOMI - pose estimation을 활용한 자세 분석 기반 재활 코칭 서비스  
활용기술: Python, FastAPI, Streamlit, WebSocket, OpenCV, YOLO, OpenAI  
- YOLO 기반 포즈 인식과 LLM을 통해 실시간 재활 피드백 제공  
- 역할: 서버/통신 인프라 구축, 모델 서빙/통합, websocket 기반 프레임 처리, streamlit UI 구성  
- 성과: YOLO 실시간 자세 분석, LLM 기반 운동 피드백 시스템 및 UI 통합 구현
- DeepDiary - LLM 기반 감정 일기 생성 서비스  
활용기술: Python, FastAPI, Streamlit, PyTorch, kobert, LLaVA, E5, Gemini API  
- 사진과 대화를 기반으로 AI가 감정 분석 및 맞춤형 일기/음악 추천 수행  
- 역할: 이미지 캡셔닝(LLaVA), 임베딩 기반 음악 추천(E5), FastAPI 서버 및 Streamlit UI 개발  
- 성과: 멀티모달 AI와 LLM을 결합한 감정 기반 인터랙티브 서비스 구현
- Pisces - ML 기반 수산시장 가격 예측  
활용기술: Python, Django, pandas, scikit-learn, LightGBM  
- 수산물 가격 변동성을 예측하여 소비자의 구매 시점 결정에 도움 제공  
- 역할: 데이터 전처리, 모델 설계 및 성능 평가, API 서버 연동 및 배포  
- 성과: LightGBM 기반 예측 모델 구축, 기상/환율/검색 트렌드 등 다양한 변수 반영

## 포바이포

2022.09 - 2024.09 (2년)

데이터 엔지니어 / 매니저

콘텐츠 제작 및 스톡 플랫폼 서비스를 제공하는 디지털 콘텐츠 기업

- 머신러닝 모델 서비스 배포

2024.01 - 2024.08

활용기술: Python, PyTorch, cuda, Ray, FastAPI, Git

- 모델 탐색: ML 후보 탐색 및 내부데이터 연동을 통한 모델별 성능 검증
- 이미지 추론 모델 배포: 이미지데이터 기반 classification, detection, captioning 모델 서비스 배포
- 시맨틱 검색 서비스 적용: 문장 임베딩 모델 적용을 통한 문장 검색 기능 서비스 배포

- KeycutSearchLab - 내부 검색 품질 테스트 도구 개발

2023.10 - 2024.01

활용 기술: AWS, MySQL, Python, Django, Git

- 검색엔진 연동: 서비스 데이터 연동을 통한 검색 기능 검증 및 개선
- 검색 테스트 환경 구축: 옵션 커스터마이징 기능을 제공하여 조직의 신속한 정책 결정에 기여
- 검색 편의기능 검증: 검색 필터 및 쿼리 변환 기능 품질 확인을 위한 솔루션 제공

- KeycutStock - 콘텐츠 플랫폼 검색 운영

2023.01 - 2024.07

활용기술: Elasticsearch, MySQL, Python, Go, Filebeat, Logstash, Kibana, Git

- 검색 품질 고도화: 검색 및 인덱싱 기능 개선을 통한 사용자 경험 향상
- 검색 편의기능 개발: 번역 및 자동완성 파이프라인 구축 및 서비스 연동을 통한 사용자 편의성 향상
- 사용자 로그 파이프라인 개발: 검색 및 웹 로그 데이터 수집 및 적재를 통한 BI 시스템 기반 구축
- 협력 업체 기술 공급: 타사 이미지 플랫폼의 검색 모듈 구축

- 서비스 검색모듈 구축

2022.09 - 2023.01

활용 기술: AWS, Docker, Elasticsearch, MySQL, Python

- 검색 시스템 구축: 데이터 파이프라인 개발 및 검색엔진 마이그레이션
- 검색모듈 API 개발: 서비스 콘텐츠 검색 기능 구현을 위한 API 개발
- 벡터 검색 도입: 단어 임베딩 및 벡터 유사도 기반 검색 알고리즘 적용을 통한 검색 정밀도 10% 향상
- 검색 성능 최적화: 벡터 양자화 통한 응답속도 20% 개선 및 스토리지 사용량 75% 절감

## 카카오

2022.06 - 2022.09 (3개월)

데이터 엔지니어 / 인턴

## 2022년 상반기 채용연계형 인턴십 - Data Engineering

- 데이터 인프라 개선 업무  
활용 기술: hadoop, hive, spark, airflow, druid
  - 데이터 파이프라인 마이그레이션
  - 탐색적 데이터 분석을 위한 DW 적재

## 마크로젠

2019.02 - 2021.05 (2년 3개월)

데이터 분석가 / 사원

유전체 분석과 생명 정보 기반 기술을 중심으로 연구자용 분석 서비스를 제공하는 바이오기업

- 분석 파이프라인 개발 및 운영 및 분석 솔루션 개발  
활용기술: python, R, mysql, shell
  - 고객 데이터를 활용한 리포트 생성 업무
  - 비즈니스 로직 간소화 및 자동화
  - 수동 업무 처리량 50% 단축
  - 통계기반 시각화 서비스 4건 개발

---

## 학력

### 경희대학교

2011.03 - 2019.08 (졸업)

생물학과 / 학사

- 데이터과학 스터디
  - R을 이용한 데이터 핸들링 및 시각화 실습
  - 머신러닝 모델 이론 및 개발 방법 학습
- 학부 졸업논문
  - Drug repositioning 약품탐색
  - NCBI 데이터 수집
  - 코사인 유사도를 활용한 통계분석