

가톨릭대학교 PIPET 데이터 사이언티스트. LLM/생성형 AI 기반 prompt engineering-평가 framework 구축 과 대규모 의료 빅데이터 분석·통계 모델링 을 모두 직접 수행. K-Pop 가상 아이돌 앱 프로젝트에서 챗봇 persona, LLM 모델 선정 벤치마크, 이미지 생성 파이프라인을 end-to-end로 설계·운영. Python·SQL·R로 데이터 추출부터 분석·모델링·LLM 응용까지 전 과정 경험.

SKILLS

Languages	Python, R, SQL, SAS
Libraries	scikit-learn, PyTorch, pandas, NumPy
Tools	Jupyter, RStudio, VS Code, Shiny, Tableau, Git, Linux, GitHub, Notion, Slack

EXPERIENCE

Data Scientist · 가톨릭대학교 PIPET (빅데이터팀) 2024.02 - 현재

- HIRA 전국 의료 빅데이터 (약 5천만 명) 기반 코호트 추출·전처리·통계 모델링 (SQL · R · SAS · Python)
- 다발골수종 (MM) 도메인 — 인과 추론 + 생존 분석으로 사전 진단의 생존 향상 효과 입증, **Blood Cancer Journal (IF 11.6) 출판**
- K-Pop 가상 아이돌 앱 사이드 프로젝트 — **LLM 모델 벤치마크 + MBTI persona priming + 이미지 생성 파이프라인** end-to-end 설계·운영, [portfolio](#) 공개

Data Analyst · 인하대병원 의학통계지원센터 2023.02 - 2023.12

- 다양한 임상 진료과 연구진 협업 — 연구 디자인 자문, 데이터 정제, 통계 모델링, 논문 통계 분석 지원
- 재직 기간 중 **통계 분석 21건 + R 통계 교육 8건** 수행

EDUCATION

MS in Statistics · Inha University 2021.03 - 2023.02

연구 분야 — 머신러닝 · 신경과학 (Neuroscience). 수업 조교 — 프로그래밍 언어, 자료구조론, 데이터 사이언스 개론, 통계분석방법·데이터 마이닝 등

BS in Statistics · Inha University 2016.03 - 2021.02

SELECTED PROJECTS ★ 프로젝트 제목 클릭 시 상세 페이지로 이동

LLM / AI Agent · 생성형 AI 파이프라인 (K-Pop 가상 아이돌 앱)

- LLM Model Benchmark** — EQ-Bench 리더보드 기반 후보 선정 → 프리미엄/경량 Top 5 비교 + 비용 시뮬레이션. MBTI priming + 도메인 큐레이션 질문으로 응답 품질 평가. [[chaeniverse/LLM-bench](#)]
- MBTI Persona Priming** — arXiv 논문 framework 검토 + 16Personalities 60문항 / 7점 Likert / 15회 반복으로 LLM persona 일관성 검증 흐름 설계.
- Casting & Content Image Generation** — 두 단계 (연습생 캐스팅 ~800장, 콘텐츠 scene ~3,200장) 각각 prompt 변수 조합 격자로 생성·정성 검증. Reference-image conditioning + negative constraints로 정체성 일관성 확보. [[chaeniverse/kpop-image-experiments](#)]
- OPT2I + DSG** — Mañas et al. 2024 (arXiv:2403.17804) iterative prompt refinement 알고리즘 구현. DSG (Davidsonian Scene Graph) 로 prompt를 작은 질문들로 쪼개 VQA 채점 — 이미지가 prompt를 얼마나 잘 반영했는지 숫자로 평가. [[chaeniverse/opt2i-dsg](#)]

Healthcare Big Data · ML

- Improved Survival in MM with Early Precursor Detection** — HIRA 약 5천만명 코호트, MGUS/SMM/de novo MM 정의·추출 + IPTW + Marginal Cox 회귀로 사전 진단의 생존 향상 효과 입증. **Blood Cancer Journal (IF 11.6, 2025)**. [[chaeniverse/mm-precursor-survival](#)]
- Long-term Opioid Use in MM** — HIRA 약 5천만명 코호트에서 신규 진단 MM 환자 3,621명 정의·추출, baseline risk score 모형 + 저/중/고위험군 stratification. submit 준비 중 (1저자).
- OvaRisk-ML** — 5종 ML 모델 (RF, LR, XGBoost, CatBoost, MLP) + RFE + Bayesian HPO + Bayes-rule 보정으로 난소암 위험군 분류. **Python Shiny 앱 배포**로 lifetime risk cumulative incidence 출력. Review 중. [[chaeniverse/OvaRisk-ML](#)] [[Shiny App](#)]

Generative AI · Medical Imaging

- Future DaTScan Synthesis with cWDM** — 초기 DaTScan 한 장으로 2년 후 영상을 합성하는 conditional wavelet diffusion 프레임워크. PPMI 746쌍 종단 데이터. **한국통계학회 1저자 포스터**. [[chaeniverse/cwdm](#)]
- Cross-Modality Neonatal Brain Image Conversion** — 신생아 뇌 T2w → T1w 합성 latent diffusion model. **IEEE AIXMHC 2025 oral (공동 교신저자)**.
- Whole-Brain Functional Gradients in Migraine** — 비선형 매니폴드 러닝으로 편두통 뇌 기능 정량화. **Human Brain Mapping (IF 4.7, 2023, 1저자)**. [[chaeniverse/migraine_project](#)]

Statistical Consulting · KPI Modeling

- Customer Satisfaction Index Modeling with SEM (Amos)** — 병원 서비스 품질을 구조방정식 (SEM) 으로 모형화 + 표준화 회귀계수 기반 고객만족지수 (CSI) 산출. 인하대병원 의학통계지원센터 컨설팅.