

Vaniushkina Kseniia

Machine Learning Engineer



+82 10 6514 2124



averksuu@naver.com



Profile

현실의 복잡하고 노이즈가 많은 데이터를 실제로 동작하는 제품으로 구현하는 Applied ML Engineer입니다. 텍스트, 의료 신호, 바이오인포매틱스 데이터를 다루며, Transformer 기반 언어모델을 학습·파인튜닝하고 다국어 텍스트 이해/분류뿐 아니라 임베딩 기반 검색·랭킹 및 문서 처리 파이프라인을 end-to-end로 개발해 왔습니다. 또한 의료 시계열 예측과 바이오 데이터 기반 모델링 경험을 바탕으로, FastAPI 기반 추론 API로 서비스에 통합하고 데이터 품질·정량 평가·배포까지 지표 중심으로 수행합니다.



Main Skills

- CV, calibration
- data preprocessing
- End-to-end ML pipeline
- visualization
- Feature engineering
- contrastive learning
- representation learning
- custom loss design
- medical modeling
- multilingual NLP
- workflow automation
- Imbalanced Learning
- large-scale sparse data handling
- finetuning



Tools & Languages

- Python, SQL
- GoogleCloud
- Git
- Docker
- FastApi
- Pytorch, TensorFlow, HuggingFace, scikit-learn
- Pandas, Numpy, Matplotlib, Seaborn, OpenCV, Scanpy, AnnData



Work Experiences

11/2025 - 01/2026
AIGEN Sciences Inc.
Seoul, South Korea



Research Engineer • Internship

단일세포 전사체 기반 신약 탐색을 위한 2-모듈 Transformer 모델 설계·구현

Drug Candidate 모듈에서 발현 변화 시그니처로부터 타깃 및 구조 (SMILES) 신호를 추론하고

Response Prediction 모듈에서 후보 약물 처리 후 발현 변화(ΔX)를 시물레이션하여 후보를 정밀 랭킹

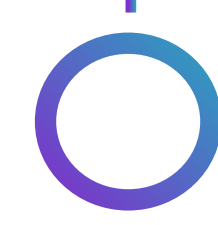
대규모 scRNA-seq perturbation 데이터 전처리(정규화, HVG 선택 등) 및 학습 파이프라인 구축

전체 약물 라이브러리 ≤ 50 개 후보만으로 정답 약물 회수 확률 85–90% 달성, 전수 탐색 대비 스크리닝 비용 약 99% 절감

04/2025 - 07/2025

Giper LLC

Moscow, Russia



Machine Learning Engineer • Contractor

마켓플레이스 내 고객 피드백 및 요청을 분석·자동화하는 AI 시스템을 개발(트리아지, 분류, 라우팅)

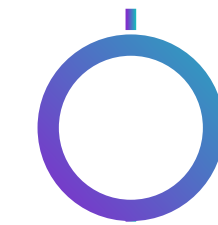
서비스 KPI(예: 처리/해결 시간, SLA 준수율, 품질 점검 지표)를 기반으로 운영 리포팅을 지원하는 상담원/오퍼레이터 성과 평가 로직을 설계·구현

ML 결과물을 사용자 워크플로우에 통합하고, 이해관계자 피드백을 반영한 반복 개선 프로세스를 운영

11/2024 - 01/2025

Endocrinology Online

Moscow, Russia



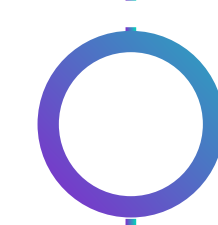
Machine Learning Engineer • Contractor

CGM(연속혈당측정), 인슐린 치료, 식이, 신체 활동 데이터를 분석하여 당뇨 치료 효과를 평가

CGM 시계열 기반 단기 혈당 예측 및 평가를 위해 슬라이딩 윈도우 기반 특징을 설계하고 LightGBM/XGBoost 모델을 구축

실시간 예측, 사용자 데이터 저장, 필요 시 모델 파인튜닝을 지원하는 Telegram 봇을 개발

(WiDS Datathon)



Unraveling the Mysteries of the Female Brain: Patterns in ADHD

fMRI 기반 특징과 설문 데이터를 활용하여 ADHD 진단 및 성별을 예측하는 멀티아웃풋 XGBoost 베이스라인 모델을 구축

클래스 불균형 완화 및 강건성 향상을 위해 SMOTE를 적용

교차검증 및 표준 분류 지표를 기반으로 모델을 선택/평가

Tech: Python, scikit-learn, XGBoost, imbalanced-learn(SMOTE), Pandas, NumPy

아이펠 (모두의연구소)



Paper is all you need

레이아웃을 유지하면서 영어 연구 논문을 한국어로 번역하는 파이프라인을 설계·구현:

PDF를 Docling으로 Markdown/HTML로 파싱 → 불필요 요소 제거 → 섹션 단위 JSON으로 구조화 → GPT-5-mini로 섹션별 번역 → Markdown을 LaTeX로 변환 → 최종 한국어 PDF로 컴파일.

기술 용어, 제목/헤딩, 그림·표, 수식은 원문 형태를 유지했으며, 번역·평가용 프롬프트 설계 및 구현/품질 평가에 기여.



Trainings & Education

아이펠 (모두의연구소)

AI Research / Machine Learning

Seoul, South Korea • 07/2025 - 01/2026

AI Researcher 트랙: 논문 리딩 및 비평,
머신러닝과 딥러닝 고급

University of Artificial Intelligence

Data Science, Machine Learning & Artificial Intelligence

Moscow • 04/2024 - 02/2025



CERTIFICATIONS

Machine Learning Specialization

Coursera (DeepLearning.AI & Stanford University)

Natural Language Processing Specialization

Coursera (DeepLearning.AI & Stanford University)

Computer Vision

Yandex ML Bootcamp



LANGUAGES

English C1

Korean C1

Russian