Kseniia Vaniushkina

Uiwan, South Korea • +82 10 6514 2124 • averksuu@naver.com

Machine Learning Engineer

Applied ML Engineer who turns complex, noisy real-world data into reliable, working products. I work across text, medical signals (e.g., CGM time series), and bioinformatics (e.g., single-cell transcriptomics)—training and fine-tuning Transformer-based language models, and building end-to-end pipelines for multilingual text understanding/classification, embedding-based retrieval and ranking, and document processing. I also develop medical time-series forecasting and bio-data modeling solutions, integrating models into user-facing services via FastAPI and driving data quality, quantitative evaluation, and deployment with a metrics-first approach.

WORK EXPERIENCE

AIGEN Sciences Inc. **11/2025 - 01/2026**

AIGEN Sciences is a South Korea–based biotech company developing AI-driven solutions for drug discovery and biomedical research.

Research Engineer • Internship **Seoul**

- Developed GeneSpeak, a two-module Transformer pipeline for drug discovery from single-cell transcriptomics
- Drug Candidate module (inverse) proposes targets/structural signals from expression deltas, and Response Prediction module (forward) simulates post-treatment expression changes for candidate ranking.
- Processed large-scale scRNA-seq perturbation data: log-normalization, Δ(treated–DMSO) signatures, and HVG selection for efficient training.
- Trained with retrieval-focused objectives (e.g., BCE, InfoNCE ranking, CLIP-style contrastive alignment).
- Achieved 85–90% probability of retrieving the true drug within ≤50 candidates, reducing screening by ~99% versus exhaustive search.

LLC Hyper **04/2025 - 07/2025**

Machine Learning Engineer • Contractor **Moscow**

- Developed an AI system to analyze and automate client feedback and requests on marketplaces (triage, categorization, and routing).
- Built operator performance evaluation logic using service KPIs (e.g., resolution time, SLA adherence, quality checks) to support operational reporting.
- Integrated ML outputs into a user-facing workflow and maintained an iteration loop based on stakeholder feedback.

LLC Endocrinology Online **11/2024 - 01/2025**

Endocrinology Online LLC is a Moscow-based digital endocrinology/diabetology organization operating the Diabetology Online portal (diabet.ru), which provides online consultations and educational resources for adults with diabetes.

Machine Learning Engineer • Contractor **Moscow**

- Analyzed CGM, insulin therapy, diet, and physical activity data to assess diabetes treatment effectiveness.
- Built LightGBM/XGBoost models using sliding-window features for time-series prediction and evaluation.
- Developed a Telegram bot for real-time predictions, user data storage, and on-demand model fine-tuning.

EDUCATION

in AI Research / Machine Learning
Aiffel (Modulabs) Seoul, South Korea • 07/2025 - 01/2026

AI Researcher track: paper reading & critique, experiment design, model evaluation, technical writing.

in Data Science, Neural Networks, Machine Learning & Artificial Intelligence
University of 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

PROJECTS

Otitis Image Classifier
Personal Project

- Built a binary CNN classifier for otoscopic image analysis
- preprocessed images and applied augmentation
- trained/validated the model and achieved 75% test accuracy
- Tech: Python, TensorFlow/Keras, NumPy, PIL, Matplotlib

Unraveling the Mysteries of the Female Brain: Sex Patterns in ADHD
(WiDS Datathon)

- Built a multi-output XGBoost baseline to predict ADHD diagnosis and sex using fMRI-derived features and questionnaire data
- applied SMOTE to address class imbalance and improve robustness
- used cross-validation and standard classification metrics for model selection
- Tech: Python, scikit-learn, XGBoost, imbalanced-learn (SMOTE), Pandas, NumPy

Paper is all you need
Aiffel

- Designed and implemented a pipeline to translate English research papers into Korean while preserving layout
- parsed PDF with Docling to Markdown/HTML, removed non-essential parts (authors/references), structured content into section-level JSON, translated section-wise with GPT-5-mini into Markdown, converted Markdown to LaTeX, and compiled the final PDF
- Maintained technical terms, headings, figures/tables, and equations in original form

- contributed prompts for translation and evaluation plus implementation and quality assessment

SKILLS

- AI service development
- CLIP-style learning
- cloud deployment & workflow automation (MLOps basics)
- CNN-based image classification
- contrastive learning
- custom loss design
- data preprocessing & visualization
- Docker
- document AI pipelines (PDF parsing with Docling)
- experiment design & ablation studies
- FastAPI
- gene & cell-line embeddings
- Git
- Hugging Face
- HVG selection
- imbalanced learning
- large-scale sparse data handling
- layout-preserving translation
- LightGBM
- Linux
- medical time-series modeling
- metric learning
- multilingual NLP
- Open CV
- Python
- PyTorch
- representation learning
- retrieval & ranking
- single-cell RNA-seq perturbation modeling
- SQL
- Telegram bot integration
- TensorFlow/Keras
- text classification
- Transformers
- XGBoost
- YOLO

LANGUAGES

- English C1
- Korean C1
- Russian (native)