Nadezhda ZHUKOVA

M.Sc. Candidate in Machine Learning

Paris, France | nadezhda.zhukova@dauphine.eu | nadezhdazhukova | zhukovanadezhda

Master's student in Machine Learning with experience analyzing high-dimensional scientific data and building rigorous, reproducible pipelines. I enjoy projects at the intersection of applied and theoretical research, combining real-world impact with methodological depth. I am currently interested in exploring representation learning and generative modeling, especially diffusion and foundation models, but I remain open to other areas as well.

Education

Université PSL (Dauphine • ENS • Mines Paris) — Paris, France
MSc in Machine Learning — Sept 2025 – Sept 2026

- Awarded the PR[AI]RIE–PSAI Excellence Scholarship for academic merit
- Coursework: Foundations of ML, Optimization, RL, LLMs, Computer Vision, etc.

Université Paris Cité — Paris, France
MSc in Bioinformatics — Sept 2023 – Sept 2025

- GPA: 16.26/20 (highest honors)
- Ranked 1st in Optimization, Signal and Image Processing

Université Paris Cité & Moscow State University — Paris & Moscow
BSc in Bioinformatics — Sept 2018 – June 2023

- GPA: 14.7/20 (honors)
- Ranked 1st in Combinatorics, Mathematical Analysis, Algorithms & Data Structures, Python

Presidential Lyceum of Physics & Mathematics №239 — Saint-Petersburg, Russia
Major in Mathematics — Sept 2013 – May 2017

Experience

INSERM — Institut de Psychiatrie et Neurosciences de Paris — Paris, France
Research Intern (Statistical Modeling & ML for Precision Medicine) — Sept 2023 – Sept 2025
Python, R, Scikit-learn, Snakemake, SLURM, GLM, Longitudinal Models

- Processed high-dimensional biomedical data into analysis-ready formats
- Conducted statistical modeling and machine learning to identify candidate biomarkers
- Designed, documented, and automated end-to-end analysis pipelines ready for scalling-up

CNRS — Institut Jacques Monod — Paris, France
Research Intern (Scientific Software & Reproducibility) — Feb 2023 – July 2023
Python, REST APIs, Snakemake, Data Engineering, Text Mining

- Designed a Snakemake pipeline for text mining and text extraction from scientific publications
- Automated data retrieval and cleaning via PubMed API
- Implemented archiving protocols in Software Heritage, ensuring reproducibility

FRAKTAL Educational Center — Saint-Petersburg, Russia
Mathematics Olympiad Mentor — Sept 2014 – Aug 2021
Mathematics, Probability, Problem solving, Pedagogical Design

- Trained national award-winning schoolchildren while teaching them advanced mathematics (combinatorics, probability, graph theory, number theory, etc.)
- Created training programs across seven centers, reaching 300+ participants (2020)
- Recognized by the State Education Committee (2018) for my contribution to the development of mathematics among schoolchildren

Conferences and Trainings

European Conference of Human Genetics (poster) — Berlin, 2024
Journées Ouvertes en Biologie, Informatique et Mathématiques (poster) — Toulouse, 2024
Eastern European Machine Learning Summer School (AIslovakIA & DeepMind) — Košice, 2023

Projects

Learning and Evaluating Generative Adversarial Networks

PyTorch, GANs, SLURM, Generative Models

- Studied GAN training stability and precision–recall trade-offs using PR-divergences
- Implemented PRD metric and analyzed generator diversity on MNIST
- Explored hyperparameter effects and different sampling strategies

Collaborative Filtering for Recommender Systems

NumPy, scikit-learn, Optuna, Optimisation
github

- Built a matrix-factorization recommender with side metadata and hyperparameter tuning
- Conducted ablation studies showing metadata improves RMSE and ranking metrics

Protein Structure Prediction via Threading

NumPy, Double-Dynamic Programming
github

- Reimplemented and tested a threading-based algorithm for protein structure prediction
- Computed DOPE-like energy scores and statistically validated them with z-score testing

Interpretable Pathogenicity Prediction with LLMs

LLMs, Explainability, Prompt Engineering

- Built a temporally-stratified dataset integrating structural and functional features
- Designed feature-based and rule-guided prompts to evaluate LLM reasoning vs ML baselines (RF, RuleFit)

Sleep State Detection Using Neural Networks (Kaggle Competition)

Deep Learning, TensorFlow/Keras, Time-Series, SMOTE

- Developed a preprocessing pipeline for accelerometer signals and engineered temporal features
- Trained and compared deep models (MLP, CNN, GRU, LSTM) on sleep-state classification

Water Shortage Prediction (Hi!ckathon at Institut Polytechnique)

XGBoost, CNN, Feature Engineering
github

- Predicted groundwater levels using weather, hydrology, and socio-economic data
- Achieved an F1-score of 58%, finishing in the top-25% of the leaderboard

Skills

Programming languages: Python, Bash, R, C, SQL, JavaScript
Pipelines & HPC: Snakemake, Nextflow, SLURM, Conda
Version control: Git, GitHub
Languages: English (fluent), French (fluent), Russian (fluent)

Interests

Tennis (weekly) • **Piano** (weekly) • **Mountain skiing** (20 years) • **Volunteering** (mentoring international students)