

ARMAN BOLATOV

+77782687150 [✉ arman.bolatov\[at\]mbzuai.ac.ae](mailto:arman.bolatov[at]mbzuai.ac.ae) [🎓 google scholar](#) [🐙 github](#) [k kaggle](#)

EDUCATION

Mohamed bin Zayed University of Artificial Intelligence

Master of Science in Machine Learning (CGPA: 3.80/4.00)

Aug. 2024 – Jun. 2026

Masdar City, Abu Dhabi, UAE

Nazarbayev University

Bachelor of Science in Computer Science (CGPA: 3.75/4.00)

Aug. 2019 – Jun. 2024

Astana, Kazakhstan

RELEVANT COURSEWORK

- Formal Languages
- Advanced Linear Algebra
- Algorithmic Problem Solving
- Industrial ML
- Statistical Learning
- Probability
- Deep Learning
- Continuous Optimization
- Artificial Intelligence
- Bioinformatics

EXPERIENCE

Wonk AI

NLP Engineer

Sep. 2023 – Aug. 2024

- Developed a Retrieval-Augmented chat-bot to assist students with their studies by integrating OpenAI GPT-3.5, Pinecone, Fasttext, and HuggingFace endpoints (Kaz-RoBERTa conversational, RuBERT-cased).
- Designed a FastAPI backend microservice to parse PDFs, segment them into chunks, and retrieve relevant textbook excerpts based on user queries.
- Leveraged ChatCompletion function calls and Langchain retrieval to aid teachers in generating quizzes.

Mohamed bin Zayed University of Artificial Intelligence

Undergraduate Research Intern

May 2023 – Jul. 2023

- Built reinforcement learning environment with 'gymnasium' and trained PPO, DDPG, TD3, A2C from stable-baselines3.
- Achieved a 3-5% accuracy increase through extensive Bayesian hyperparameters search using WandB sweeps.
- Contributed to a workshop submitted to IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids, 2023.

Institute of Smart Systems and Artificial Intelligence

Computer Vision Research Engineer

May 2022 – July 2023

- Collected, processed, and annotated the Central Asian Food Dataset consisting of 16,000 images of Central Asian dishes.
- Successfully trained CAFD on VGG, ResNet, EfficientNet, ViT, and DenseNet models using designated seeds.
- Achieved a 1-2% accuracy increase through deep data augmentation using a personalized Latent Diffusion Model.
- Explored ways to visualize and send a weekly reports of project progress to team members.

National Analytic Center of Kazakhstan

Intern Data Scientist

Dec. 2021 – Feb. 2022

- Created Python scripts for calculating death rates & probabilities, infant mortality rates, life expectancy, population by age and building life tables from the Republic Database of Private Entities.

Beyond Curriculum Public Foundation

Head of Scoreboard

Dec. 2020 – Mar. 2022

- Created a database of 50+ Olympiad results on 15 school subjects cumulatively achieving 8000+ participants.
- Incorporated Python scripts to aggregate raw CSV and Excel Olympiad results into JSON arrays and to load the latest into scoreboard.kz.
- Led team of 15+ members to expand the Olympiad database, delegated tasks and ran weekly meetings to oversee progress in essential parts of the project.

PUBLICATIONS

- A. Bolatov, S. Horváth, M. Takáč, E. Gorbunov. *Byzantine-Robust Optimization under (L_0, L_1) -Smoothness*. CPAL 2026 (Proceedings Track) Poster, 2026.
- A. Riabinin, A. Veprikov, A. Bolatov, M. Takáč, A. Beznosikov. *Where Does Warm-Up Come From? Adaptive Scheduling for Norm-Constrained Optimizers*. arXiv:2602.05813, 2026.
- V. Kovalchuk, D. Son, A. Bolatov, M. Guizani, S. Horváth, M. Panov, M. Takáč, E. Gorbunov, N. Kotelevskii. *Who to Trust? Aggregating Client Knowledge in Logit-Based Federated Learning*. arXiv:2509.15147, 2025.
- A. Veprikov, A. Bolatov, S. Horváth, A. Beznosikov, M. Takáč, S. Hanzely. *Preconditioned Norms: A Unified Framework for Steepest Descent, Quasi-Newton and Adaptive Methods*. arXiv:2510.10777, 2025.

I. Rodkin, D. Orel, K. Smirnov, **A. Bolatov**, et al. *Beyond Memorization: Extending Reasoning Depth with Recurrence, Memory and Test-Time Compute Scaling*. arXiv:2508.16745, 2025.

A. Bolatov, A. Legg, I. Melnykov, A. Nurlanuly, M. Tezekbayev, Z. Assylbekov. *Overspecified Mixture Discriminant Analysis: Exponential Convergence, Statistical Guarantees, and Remote Sensing Applications*. arXiv:2510.27056, 2025.

P. Skrzypacz, **A. Bolatov**, A. Dziedzic, G. Ellis, K. Tangirbergen, B. Pruchnik, P. Putek. *Fourier Approximation of magMEMS Oscillations: Neural Network Space Handling*. Micromachines 16 (12), 1355, 2025.

A. Pak, M. Tezekbayev, **A. Bolatov**, I. Melnykov, K. Dauletbek, Zh. Assylbekov. *Classifier Performance on Long-Tail Distributions*. Statistical Analysis and Data Mining: An ASA Data Science Journal 18 (4), e70037, 2025.

A. Legg, I. Melnykov, **A. Bolatov**, A. Pak, Z. Assylbekov. *Convergence of the EM Algorithm in KL Distance for Overspecified Gaussian Mixtures*. Statistical Papers 66 (6), 125, 2025.

L. Zhao, J. H. He, P. S. Skrzypacz, **A. Bolatov**, et al. *Optimizing dynamic pull-in threshold and periodic trajectories for magnetically actuated MEMS (magMEMS) in wearable sensors*. Frontiers in Physics 13, 1638299, 2025.

R. Takhanov, M. Tezekbayev, A. Pak, **A. Bolatov**, Z. Assylbekov. *Gradient Descent Fails to Learn High-frequency Functions and Modular Arithmetic*. Machine Learning 114 (4), 117, 2025.

R. Takhanov, M. Tezekbayev, A. Pak, **A. Bolatov**, Zh. Kadyrsizova, Zh. Assylbekov. *Intractability of Learning the Discrete Logarithm with Gradient-Based Methods*. Asian Conference on Machine Learning (ACML), 1321-1336, 2024.

A. Karabay, **A. Bolatov**, H. A. Varol, M. Y. Chan. *A Central Asian Food Dataset for Personalized Dietary Interventions*. Nutrients 15 (7), 1728, 2023.

A. Bolatov, M. Tezekbayev, I. Melnykov, A. Pak, V. Nikoulina, Zh. Assylbekov. *Long Tail Theory Under Gaussian Mixtures*. European Conference on Artificial Intelligence (ECAI), 2023.

S. N. Longinos, **A. Bolatov**, P. Skrzypacz, R. Hazlett. *Application of Image Processing in Evaluation of Hydraulic Fracturing with Liquid Nitrogen: A Case Study of Coal Samples from Karaganda Basin*. Applied Sciences 13 (13), 7861, 2023.

A. Bolatov, K. Dauletbek. *Empirical Analysis of the AdaBoost's Error Bound*. arXiv:2302.00880, 2023.

PROJECTS

Neural Defense: Lyrics Generator | Python, Streamlit, Inference API, PyTorch December 2022

- Web-scraped lyrics of Yegor Letov's songs and fine-tuned a Russian GPT-3-Large language model on lyrics and title pairs.
- Deployed a web application that generates lyrics from titles using Streamlit and the HuggingFace Inference API.

Gesture and Keyword Control of Smart House | Python, Picovoice Porcupine, OpenCV, CV Zone May 2022

- Developed a program that authorizes the user's face, reads hand gestures, spots keywords and, depending on this, opens/closes the door, changes the temperature level, or turns off/on the light in the rooms.
- Built a prototype of the smart house and integrated the program into the Arduino circuit.

DNA Sequence Analyzer | Python, PySimpleGUI Sep. 2021

- Created algorithms that can detect and alter the restriction sites of user's choosing in the given DNA sequence.
- Utilized PySimpleGUI to build an interface which shows the restriction sites within the DNA sequence and annotates the precise name and position of the restriction site.

TECHNICAL SKILLS

Languages: Python, C++.

Developer Tools: VS Code, Jupyter Notebook, Google Colab, Git, Docker.

Technologies/Frameworks: Pandas, OpenCV, PyTorch, Linux, scikit-learn, Wandb, \LaTeX .

HONORS AND AWARDS

International Math Competition for Undergraduate Students | First prize August 2022

Kaggle American Express – Default Prediction | Silver medal August 2022

Samsung Innovation Campus Hackathon | First place July 2023

HackNU/22, HackNU/23 and HackNU/24 | Second place Fall 2022, Spring 2023 and Spring 2024

Dean's List | Nazarbayev University Fall 2022

Google Developers Student Club – Fire Recognition Hackathon | First place November 2022

JAS AI Hackathon | Third place November 2023

National Math Olympiad | Bronze medal March 2017

International Zhautykov Olympiad in Mathematics | Silver medal January 2018