Ravil Mussabayev, PhD

+7 (962) 9587359 • Moscow, Russia • ravmus@uw.edu
rmusab.github.io • GitHub • LinkedIn • Google Scholar

Education

University of Washington — Seattle, WA
Doctor of Philosophy (PhD) in Mathematics [diploma, transcript], GPA: 3.81/4.0. — Sep 2018 – May 2024

- *Advisor*: Prof. Gunther Uhlmann.
- *Dissertation*: "How to Use K-means for Big Data Clustering?".
- *Topic*: Novel efficient, effective, scalable, parallel, and simple big data clustering algorithms, conceptualized under multimodal optimization metaheurstics.
- *Key publications*: "How to Use K-means for Big Data Clustering?" (Q1 "Pattern Recognition", IF 8.52), "Superior Parallel Big Data Clustering through Competitive Stochastic Sample Size Optimization in Big-means" (ACIIDS 2024, Core B, Best Paper Award), "High-Performance Hybrid Algorithm for Minimum Sum-of-Squares Clustering of Infinitely Tall Data" (Q1 "Mathematics", IF 2.3), "Optimizing Parallelization Strategies for the Big-Means Clustering Algorithm" (OPTIMA 2023).
- Improved state of the art by more than 93.3% in accuracy and 96.3% in time for large datasets.

University of Washington — Seattle, WA
Master of Science in Mathematics [diploma], GPA: 3.81/4.0. — Sep 2018 – Mar 2021

Kazakh-British Technical University — Almaty, Kazakhstan
Bachelor of Engineering in Mathematical and Computer Modeling, GPA: 3.93/4.0. — Aug 2015 – Jun 2018
[transcript+diploma]

- *Advisor*: Nurlan S. Dairbekov.
- *Thesis topic*: Reinforcement learning-based (Q-learning, Deep Q-Network, Actor-Critic) road intersection controller and curvature-based clustering [thesis, presentation].
- *Key publications*: "Reinforcement Learning Intersection Controller" (ICECCO 2018), "Colour-based object detection, inverse kinematics algorithms and pinhole camera model for controlling robotic arm movement system" (ICECCO 2015).

Research Interests

H-index: 4 (Google Scholar).

AI in education, including the aspects of digitalization, personalization, adaptiveness, automation.

Algorithmic foundations of computer science: multimodal generative AI (GenAI), AI-based decision support systems, augmented intelligence, multi-agent systems, big data technologies, clustering, multimodal global optimization and soft computing (metaheuristics), operations research, streaming algorithms, distributed and parallel algorithms, natural language processing, and AI for software engineering.

Work Experience

Senior Research Scientist — Jan 2025 – current
Sber AI Lab — *Moscow, Russia*

- Work on industry projects related to natural language processing (NLP).

Research Scientist — Oct 2023 – current
Satbayev University — *Almaty, Kazakhstan*

Project "Automated Construction of a Multilingual Ontology for Empowering the Kazakh Language Through Advanced AI Technologies" (BR21882268).

- Focus on developing advanced search methods for large-scale optimization of machine learning algorithms.
- Application areas: big data clustering, construction of optimal multilingual vocabularies and ontologies.

Research Engineer — Jun 2022 – Jun 2024
Huawei Russian Research Institute — *Moscow, Russia*

- Achieved top 0.55 F1 score performance of method name prediction model for Java and C++; developed novel natural language generation metric [paper].

- Implemented, trained, and evaluated deep learning models, graph neural networks, and large language models (LLMs) (prompt engineering, in-context learning) for source code vulnerability detection [paper].
- Collected noiseless real-world vulnerability dataset for Java; fine-tuned open-source code LLMs for in-depth vulnerability analysis using next token prediction and binary classification regimes; achieved state-of-the-art AUC score of 0.69; prepared and submitted research paper [paper].

Research Engineer Mar 2018 – Aug 2018
Satbayev University *Almaty, Kazakhstan*

Project "Development of the software package for 3D modelling, continual monitoring and forecasting of the level of air pollution in densely populated and industrial areas" (BR05236316).

- Applied reinforcement learning algorithms (Q-learning, Deep Q-Network, Actor-Critic) to traffic light control problem; published results [paper, pdf].
- Designed novel reward formula leading to convergence of models in traffic simulation package SUMO; visualized results.
- Trained junior university students in probability and reinforcement learning basics; mentored them through research agenda.

Research Engineer Mar 2015 – Dec 2017
Uniline Group LLP *Almaty, Kazakhstan*

- Resolved robotic arm control problem using Python and single video camera input; published results [paper, pdf]. Delivered 2 international conference talks on findings.
- Engineered backend for Kazakh text-to-speech system in Microsoft Visual C++ to aid individuals with disabilities.
- Executed Fujisaki model for pitch contour generation, enhancing naturalness of synthesized speech.

Freelance Software Developer, *Almaty, Kazakhstan* Sep 2010 – Sep 2014

- Developed and launched Android app "Physics Lab" on Google Play; achieved over 17,000 installations in 3 months.
- Developed commercial application for new cryptographic algorithm in Embarcadero Delphi.
- Designed and developed desktop and Android apps for English vocabulary learning using text-to-speech technologies and MySQL database.

Academic Experience

Associate Professor Sep 2024 – current
Satbayev University *Almaty, Kazakhstan*

- Teach an NLP course "Applied Text Processing" (CSE2562).
- Supervise a doctoral student.
- Apply for scientific grants.

Teaching Assistant & Predoctoral Instructor Sep 2018 – Jun 2022
University of Washington *Seattle, WA*

- Taught introductory courses for sections of 40 students: Matrix Algebra with Applications, Introduction to Differential Equations.
- Conducted practical college-level math classes for 40 students; led in-class discussions, held office hours, compiled and administered exams: Precalculus, Calculus with Analytic Geometry I, II, III, Introduction to Differential Equations, Linear Optimization.
- Tutored students in challenging math concepts; helped develop problem-solving and analytical thinking skills.

Teaching Assistant Jan 2017 – Dec 2017
Kazakh-British Technical University *Almaty, Kazakhstan*

- Conducted calculus classes with focus on problem solving.
- Compiled, administered, and graded written exams.

Published Articles

1. Ravil Mussabayev. «WRDScore: New Metric for Evaluation of Natural Language Generation Models». In: *2024 20th International Asian School-Seminar on Optimization Problems of Complex Systems (OPCS)*. 2024, pp. 20–23. DOI: 10.1109/OPCS63516.2024.10720439.
2. Ravil Mussabayev and Rustam Mussabayev. «Comparative Analysis of Optimization Strategies for K-means Clustering in Big Data Contexts: A Review». In: *Applied Soft Computing* (2024). Preprint submitted to "Applied Soft Computing". arXiv: 2310.09819 [cs.LG].
3. Ravil Mussabayev and Rustam Mussabayev. «High-Performance Hybrid Algorithm for Minimum Sum-of-Squares Clustering of Infinitely Tall Data». In: *Mathematics* 12.13 (2024). ISSN: 2227-7390. DOI: 10.3390/math12131930. URL: https://www.mdpi.com/2227-7390/12/13/1930.
4. Ravil Mussabayev and Rustam Mussabayev. «Optimizing Parallelization Strategies for the Big-Means Clustering Algorithm». In: *Advances in Optimization and Applications*. Ed. by Nicholas Olenev, Yuri Evtushenko, Milojica Jaćimović, Michael Khachay, and Vlasta Malkova. Petrovac, Montenegro: Springer Nature Switzerland, 2024, pp. 17–32. ISBN: 978-3-031-48751-4. DOI: 10.1007/978-3-031-48751-4_2.
5. Rustam Mussabayev and Ravil Mussabayev. «Superior Parallel Big Data Clustering Through Competitive Stochastic Sample Size Optimization in Big-Means». In: *Intelligent Information and Database Systems*. Ed. by Ngoc Thanh Nguyen, Richard Chbeir, Yannis Manolopoulos, Hamido Fujita, Tzung-Pei Hong, Le Minh Nguyen, and Krystian Wojtkiewicz. Lecture Notes in Artificial Intelligence. Ras Al Khaimah, UAE: Springer Nature Singapore, 2024, pp. 224–236. ISBN: 978-981-97-4985-0. DOI: 10.1007/978-981-97-4985-0_18.
6. Rustam Mussabayev and Ravil Mussabayev. «Variable Landscape Search: A Novel Metaheuristic Paradigm for Unlocking Hidden Dimensions in Global Optimization». In: *Annals of Operations Research* (2024). Preprint submitted to "Annals of Operations Research". arXiv: 2408.03895 [math.OC].
7. Rustam Mussabayev, Nenad Mladenovic, Bassem Jarboui, and Ravil Mussabayev. «How to Use K-means for Big Data Clustering?» In: *Pattern Recognition* 137 (2023), p. 109269. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2022.109269. arXiv: 2204.07485 [cs.LG]. URL: https://www.sciencedirect.com/science/article/pii/S0031320322007488.
8. Gulnur Tolebi, Nurlan S. Dairbekov, Daniyar Kurmankhojayev, and Ravil Mussabayev. «Reinforcement Learning Intersection Controller». In: *2018 14th International Conference on Electronics Computer and Computation (ICECCO)*. 2018, pp. 206–212. DOI: 10.1109/ICECCO.2018.8634692.
9. Ravil Mussabayev. «Colour-based object detection, inverse kinematics algorithms and pinhole camera model for controlling robotic arm movement system». In: *2015 Twelve International Conference on Electronics Computer and Computation (ICECCO)*. 2015, pp. 1–9. DOI: 10.1109/ICECCO.2015.7416879.

Working Papers & Reports

1. Ravil Mussabayev and Rustam Mussabayev. «Boosting K-means for Big Data by Fusing Data Streaming with Global Optimization». 2024. arXiv: 2410.14548 [cs.LG].
2. Alexey Shestov, Rodion Levichev, Ravil Mussabayev, Evgeny Maslov, Anton Cheshkov, and Pavel Zadorozhny. «Finetuning Large Language Models for Vulnerability Detection». 2024. arXiv: 2401.17010 [cs.CR].
3. Ravil Mussabayev. *Structure-Aware Code Vulnerability Analysis With Graph Neural Networks*. 2023. arXiv: 2307.11454 [cs.CR].

Conferences

- Talk and paper "WRDScore: New Metric for Evaluation of Natural Language Generation Models" at XX International Asian Seminar "Optimization Problems of Complex Systems" (OPCS), *Cholpon-Ata, Kyrgyzstan.* Jul 2024
- Talk and "Best Paper Award" "Superior Parallel Big Data Clustering through Competitive Stochastic Sample Size Optimization in Big-means" at XVI Asian Conference on Intelligent Information and Database Systems (ACIIDS 2024, Core B ranking), *Ras Al Khaimah, UAE.* Apr 2024
- Talk and paper "Новый параллельный алгоритм кластеризации больших данных Big-means с конкурентной стохастической оптимизацией размера выборки" at XXI International Conference "Mathematical Methods of Pattern Recognition" (MMPR-21), *Moscow, Russia* Dec 2023
- Reviewer, talk, and paper "Optimizing Parallelization Strategies for the Big-Means Clustering Algorithm" at XIV International Conference "Optimization and Applications" (OPTIMA-2023), *Petrovac, Montenegro.* Sep 2023

- Talk and paper "Colour-Based Object Detection, Inverse Kinematics Algorithms and Pinhole Camera Model for Controlling Robotic Arm Movement System" at XII International Conference on Electronics, Computer, and Computation (ICECCO-2015) [presentation], *Almaty, Kazakhstan.* Sep 2015
- Talk and paper "Распознавание цветных объектов, обратная кинематика и модель камеры-обскуры для управления рукой-манипулятором" at XI International Asian Seminar "Optimization Problems of Complex Systems" (OPCS), *Cholpon-Ata, Kyrgyzstan.* Aug 2015

Peer Review

- XIV International Conference "Optimization and Applications" (OPTIMA-2023). Sep 2023
- Elsevier "Pattern Recognition" journal. May 2024
- Elsevier "Information Fusion" journal. Nov 2024
- Elsevier "Information Fusion" journal. Nov 2024
- Elsevier "Information Fusion" journal. Jan 2025
- XV Asian Conference on Intelligent Information and Database Systems (ACIIDS 2025, Core B ranking). Jan 2025

Skills

- **Programming:** Python (PyTorch, NumPy, Pandas, OpenCV, Scikit-learn, Matplotlib), C/C++ (OpenCV), Scala, MATLAB, R, Java, HTML/CSS, Databases (Microsoft SQL Server, MySQL), TensorFlow, Keras, Hadoop, Spark, Docker, Git, Linux shell scripts, LaTeX (wrote $>$ 1000 pages of math).
- **Machine learning:** supervised learning (classification, regression, Naive Bayes, SVM, random forests), reinforcement / imitation learning, unsupervised learning / clustering (including big data), neural networks (RNNs, CNNs, GANs, autoencoders), natural language processing, transformers, BERT, graph neural networks, large language models (LLMs): prompting, in-context learning, chain-of-thought reasoning, finetuning, LoRA, PEFT.
- **Relevant coursework:** Programming Languages, Algorithms and Data Structures, Theory of Probability and Mathematical Statistics, Inroduction to Machine Learning, Introduction to Artificial Intelligence, Natural Language Processing, Deep Learning, Introduction to Cryptography, Distributed Systems, High-Dimensional Probability in Data Science, Networks and Combinatorial Optimization, Convex Analysis and Nonsmooth Optimization, Optimal Transport, Real Analysis, Topological and Smooth Manifolds.
- **Electrical engineering:** computer hardware, microcontrollers (Arduino, Raspberry Pi), servo motor actuators.

Summer Schools

- "Best Project Award" at AIRI Summer School 2024. Aug 2024
- MSRI Mathematics of Machine Learning summer graduate school. Jul – Aug 2019
- ADSI Summer Workshop on Algorithmic Foundations of Learning and Control. Aug 2019

Leadership and Awards

- Winner (among 251 submitted manuscripts from 43 countries) of "Best Paper Award" at ACIIDS 2024 conference (Core B ranking). Apr 2024
- Winner (among 300 participants) of $7000 for summer research abroad by Yessenov Foundation based on leadership and research potential. Jun – Aug 2017
- Held fellowship named after First President of Republic of Kazakhstan for academic excellence.
- Elected member of university-wide student parliament, 10 out of 2200 students.
- Winner in nominations "Best GPA 2016, 2017".
- Winner of district (2011 – 2014), city (2012, 2013), and national (2013) Olympiads in physics.
- Graduated with honors from MIPT's Distance Learning School of Physics and Technology (Moscow, Russia).

References

- Prof. Kenneth Bube [letter]

- Prof. Nurlan S. Dairbekov [letter]
- Prof. Gunther Uhlmann