

Ravil Mussabayev

+1 (206) 9391844 • Seattle, WA • ✉ ravmus@uw.edu
🌐 rmusab.github.io • 🐙 [GitHub](#) • 🔗 [LinkedIn](#) • 📄 [Google Scholar](#)

Education

University of Washington

Seattle, WA

Doctor of Philosophy (PhD) in Mathematics, GPA: 3.81/4.0.

Sep 2018 – May 2024

- *Advisor*: Prof. Gunther Uhlmann.
- *Dissertation*: “How to Use K-means for Big Data Clustering?”.
- Improved state of the art by more than 93.3% in accuracy and 96.3% in time for large datasets.
- Published paper “[How to Use K-means for Big Data Clustering?](#)” in peer-reviewed Q1 journal “Pattern Recognition” with impact factor 8.52.

University of Washington [\[diploma\]](#)

Seattle, WA

Master of Science in Mathematics, GPA: 3.81/4.0.

Sep 2018 – Mar 2021

Kazakh-British Technical University [\[diploma\]](#)

Almaty, Kazakhstan

Bachelor of Engineering in Mathematical and Computer Modeling, GPA: 3.93/4.0.

Aug 2015 – Jun 2018

- *Advisor*: Nurlan S. Dairbekov.
- *Thesis*: “Self-learning neural network based traffic signal controller for an isolated intersection and construction of a new clustering algorithm in unsupervised machine learning” [[thesis](#), [presentation](#)].

Research Interests

Data Science, Clustering, Big Data, Global Optimization, Metaheuristics, and Natural Language Processing.

Work Experience

Research Scientist

Jun 2022 – current

Huawei Russian Research Institute

Moscow, Russia

- Implement and improve method name prediction algorithms for Java and C++, reaching state-of-the-art 0.55 F1 score.
- Devise, train, and evaluate novel deep-learning models, graph neural networks, and large language models (LLMs) for code vulnerability and malware detection in Java, achieving state-of-the-art AUC score of 0.69 on real-world Java vulnerability dataset.
- Prepare and submit a research paper that encompasses new vulnerability and malware benchmark dataset collection, optimization of the vulnerability and malware detection paradigm, application of prompting and in-context learning of LLMs to vulnerability and malware detection, and fine-tuning of SOTA LLMs for in-depth analysis [[pdf](#)].

Research Assistant

Mar 2018 – Aug 2018

Satbayev University

Almaty, Kazakhstan

- Applied reinforcement learning algorithms (Q-learning, Deep Q-Network, Actor-Critic) to traffic light control problem; published results [[ieee](#)][[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 [[ieee](#)][[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.

Teaching Experience

Teaching Assistant & Predoctoral Instructor
University of Washington

Sep 2018 – Jun 2022
Seattle, WA

- Taught introductory courses in differential equations and linear algebra for sections of 40 students.
- Conducted practical college-level math classes for 40 students; led in-class discussions, held office hours, compiled and administered exams.
- Tutored students in challenging math concepts; helped develop problem-solving and analytical thinking skills.

Teaching Assistant
Kazakh-British Technical University

Jan 2017 – Dec 2017
Almaty, Kazakhstan

Publications

1. Ravil Mussabayev и Rustam Mussabayev. “Optimizing Parallelization Strategies for the Big-Means Clustering Algorithm”. B: *Advances in Optimization and Applications*. Под ред. Nicholas Olenev, Yuri Evtushenko, Milojica Jaćimović, Michael Khachay и Vlasta Malkova. Cham: Springer Nature Switzerland, 2024, с. 17—32. ISBN: 978-3-031-48751-4. DOI: [10.1007/978-3-031-48751-4_2](https://doi.org/10.1007/978-3-031-48751-4_2).
2. Rustam Mussabayev и Ravil Mussabayev. *Superior Parallel Big Data Clustering through Competitive Stochastic Sample Size Optimization in Big-means*. 2024. arXiv: [2403.18766](https://arxiv.org/abs/2403.18766) [cs.LG].
3. Alexey Shestov, Rodion Levichev, Ravil Mussabayev, Evgeny Maslov, Anton Cheshkov и Pavel Zadorozhny. *Finetuning Large Language Models for Vulnerability Detection*. 2024. arXiv: [2401.17010](https://arxiv.org/abs/2401.17010) [cs.CR].
4. Ravil Mussabayev. *Dissecting Code Vulnerabilities: Insights from C++ and Java Vulnerability Analysis with ReVeal Model*. 2023. arXiv: [2307.11454](https://arxiv.org/abs/2307.11454) [cs.CR].
5. Ravil Mussabayev и Rustam Mussabayev. *Optimizing K-means for Big Data: A Comparative Study*. 2023. arXiv: [2310.09819](https://arxiv.org/abs/2310.09819) [cs.LG].
6. Ravil Mussabayev и Rustam Mussabayev. *Strategies for Parallelizing the Big-Means Algorithm: A Comprehensive Tutorial for Effective Big Data Clustering*. 2023. arXiv: [2311.04517](https://arxiv.org/abs/2311.04517) [cs.DC].
7. Rustam Mussabayev, Nenad Mladenovic, Bassem Jarbouli и Ravil Mussabayev. “How to Use K-means for Big Data Clustering?” B: *Pattern Recognition* 137 (2023), с. 109269. ISSN: 0031-3203. DOI: <https://doi.org/10.1016/j.patcog.2022.109269>. arXiv: [2204.07485](https://arxiv.org/abs/2204.07485) [cs.LG]. URL: <https://www.sciencedirect.com/science/article/pii/S0031320322007488>.
8. Gulnur Tolebi, Nurlan S. Dairbekov, Daniyar Kurmankhojayev и Ravil Mussabayev. “Reinforcement Learning Intersection Controller”. B: *2018 14th International Conference on Electronics Computer and Computation (ICECCO)*. 2018, с. 206—212. DOI: [10.1109/ICECCO.2018.8634692](https://doi.org/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”. B: *2015 Twelve International Conference on Electronics Computer and Computation (ICECCO)*. 2015, с. 1—9. DOI: [10.1109/ICECCO.2015.7416879](https://doi.org/10.1109/ICECCO.2015.7416879).

Conferences

- Talk and paper “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), *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) [[pdf](#)], *Almaty, Kazakhstan*. Sep 2015
- Talk and paper “Colour-Based Object Detection, Inverse Kinematics Algorithms and Pinhole Camera Model for Controlling Robotic Arm Movement System” at XI International Asian Seminar “Problems on the optimization of complex systems”, *Cholpon-Ata, Issyk-Kul, Kyrgyzstan*. Aug 2015

Peer Review

- XIV International Conference “Optimization and Applications” (OPTIMA-2023). Sep 2023
- Elsevier “Pattern Recognition” journal. May 2024

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, Introduction 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.

Leadership and Awards

- Winner (among 251 submitted manuscripts from 43 countries) of “[Best Paper Award](#)” at [ACIIDS 2024](#) conference. 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). Jun 2014

References

- [Prof. Kenneth Bube](#) [[letter](#)]
- [Prof. Nurlan S. Dairbekov](#)
- [Prof. Gunther Uhlmann](#)