

Anya Korsakova

anya@calicolabs.com | codehelix.ai | github.com/anyakors

Skills

Programming: python, TensorFlow, pyTorch, bash, C++ , openCV

Tools: slurm, Google Cloud Platform, Docker, LaTeX

Machine Learning (ML): transformers, variational autoencoders (VAEs), sparse autoencoders (SAEs), energy-based models, convolutional neural networks (CNNs), decision trees, input importance attribution

Experience

Postdoctoral researcher, Calico Life Sciences LLC (Alphabet) – San Francisco, CA May 2023 – Present
ML architecture design and data preparation for biological ML models.

- Enabled genetic variant effect prediction for structural variants using align-and-stitch method at inference and created new benchmarks for models while improving AUROC performance on existing benchmarks by 9% [1].
- Built a framework for mechanistic interpretability and extracted sequence insights from DNA sequence-based ML models using SAEs.

Postdoctoral Researcher, Cancer Science Institute – Singapore Oct 2022 –May 2023

- Devised and implemented a probabilistic mutational signature assignment method ALPS
gitlab.com/PittGenomics/alps.
- Collaborated on building ensemble approaches to mutational signature assignment [2].

Quantitative Finance Developer, Juniper Investment Pte Ltd – Singapore Feb 2022 – Sept 2022

- Programmed high yield multi-timeframe, multi-instrument trading bots for foreign exchange markets in C++ (MQL5) using technical and fundamental analysis.

PhD Scholar, Nanyang Technological University – Singapore Aug 2017 – July 2022

Developed machine learning frameworks for structural and cellular biology.

- Built a G-quadruplex DNA structure prediction model using convolutional neural networks and auxiliary inputs that achieved state-of-the-art performance [3].
- Implemented data pipelines and devised the architecture of an RNA splicing prediction framework with an energy-based model augmented with RNA-binding protein levels [4].

Junior Researcher, Lebedev Physical Institute RAS – Moscow, Russia Jan 2014 – Aug 2017

- Modeled diffusive-thermal instabilities in hydrogen-air flames via solving systems of partial differential equations in Mathcad and MATLAB; reported instability onset regimes in journal publications [5].

Junior C++ Developer, NRNU MEPhI – Moscow, Russia Dec 2012 – Dec 2013

- Implemented an algorithm for eye iris recognition and tracking in live video stream with C++ and openCV.

Projects

Attractors in Neural Networks codehelix.ai/blog/attractors

- Visualizing and analyzing attractors in feedforward neural network circuits, featured in Towards Data Science.

Education

Nanyang Technological University, Singapore – PhD in Biophysics July 2022

NRNU MEPhI, Moscow, Russia – MS in Applied Mathematics and Physics (First Class Honors) June 2016

NRNU MEPhI, Moscow, Russia – BS in Applied Mathematics and Physics June 2016

Awards

SINGA Scholarship Award – Nanyang Technological University, Singapore 2017 – 2021

Best Student Award – NRNU MEPhI, Moscow, Russia 2015

Peer Review

Nature Machine Intelligence ISSN: 2522-5839, reviewer, 2024

Research Internships

Wan Lab of RNA structuromics , Genome Institute of Singapore	Mar 2021 - June 2021
RNA G4 structure probing in long synthetic RNA with NAI-SHAPE and nanopore sequencing.	
Lab of Biology and Applied Pharmacology , École Normale Supérieure Paris-Saclay	June 2018 – Aug 2018
Geometrical improvement of NMR RDC data usage for G4 structure resolution.	
Institute of Technical Thermodynamics , Karlsruhe Institute of Technology.	Oct 2015 – Nov 2015

Teaching and Supervision

NTU School of Physical and Mathematical Sciences, Singapore

- **Undergraduate student mentor** 2018, 2020, 2021
Supervised students on an ML-genomic project and a nanopore sequencing project.
- **From synthesis to quantification of DNA using UV absorption, circular dichroism and fluorescence spectroscopy (PH3399)** 2019, 2020
- **Silicon Charge Particle Detectors (PH3199)** 2018, 2019, 2020
- **Fabrication Laboratory (PH3199)**

Conference Presentations

- Korsakova A et al., “**Shift augmentation for improved indel scoring in DNA sequence-based ML models**”, ASHG 2024 Annual Meeting, Denver.
- Srivastava D, Yuan H, Korsakova A et al., “**Borzoï-guided fine mapping improves variant and gene prioritization in GWAS**”, ASHG 2024 Annual Meeting, Denver.

Selected Publications

- [1] Korsakova A et al., “**Shift augmentation for improved indel scoring in DNA sequence-based ML models**” to be submitted to Nat Methods.
- [2] Wu AJ, Perera A, Kularatnarajah L, Korsakova A, Pitt JJ, “**Mutational signature assignment heterogeneity is widespread and can be addressed by ensemble approaches**” in Briefings in Bioinformatics, 2023, DOI:10.1093/bib/bbad331.
- [3] Korsakova A, Phan AT, “**Prediction of G4 formation in live cells with epigenetic data: a deep learning approach**” in NAR Genomics and Bioinformatics, 2023, DOI:10.1093/nargab/lqad071.
- [4] Chan A, Korsakova A et al., “**RNA alternative splicing prediction with discrete compositional energy network**” at ACM CHIL ’21, DOI:10.1145/3450439.3451857.
- [5] **Google Scholar**: scholar.google.com/citations?user=5A3VUGMAAAAJ.