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): ML system design, transformers, variational autoencoders (VAEs), convolutional neural networks (CNNs), energy-based models, decision trees, input importance attribution, dictionary learning

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.

- Mined data, devised and built a G-quadruplex DNA structure prediction model using convolutional neural networks that achieved state-of-the-art performance [3].
- Implemented data mining pipeline and devised architecture of an RNA splicing prediction framework with an energy-based model and augmented input sources [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

Metaheuristic portfolio optimization Confidential Project

- Designed a metaheuristic portfolio optimization system for daily strategy selection using genetic algorithms, decision trees and regression methods for a cryptocurrency liquidity provider.

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.

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.