Christopher Rohlicek

crohlicek | crohlicek.github.io | ✉ rohlicek@alumni.harvard.edu | ☎ 617.686.3089

SUMMARY

Machine Learning and Data Scientist trained in pure and applied mathematics with research experience applying deep learning (AI/ML) at leading laboratories, including the Broad Institute of MIT and Harvard and MIT Lincoln Laboratory. Collaborative team member with a demonstrated ability to take initiative and build interdisciplinary relationships. Skilled communicator with a special interest in applications of statistical machine learning to biological systems.

EDUCATION

2020 - 2021 **Brown University** – ScM in Data Science
Relevant Coursework: Probability, Statistics, and Machine Learning; Data Engineering; Statistical Learning; Deep Learning and Special Topics in Data Science; Modern Applications of Probability and Statistics

2016 - 2020 **Harvard University** – AB in Applied Mathematics
Relevant Coursework: Matrix Methods in Data Analysis, Signal Processing, and Machine Learning; Decision Theory; Abstraction and Design in Computation; Computer Networks; Optimization; Honors Linear Algebra and Multivariable Calculus and Real Analysis; Theory of Groups and Vector Spaces; Vector Space Methods for Differential Equations

WORK EXPERIENCE

Generate:Biomedicines August 2024 - present
Scientist I, Data Science – Computational Protein Generation

- Data scientist on Computational Protein Generation team, building research infrastructure to facilitate in silico experimentation in computational therapeutic design workflow.
- Researching Bayesian methods for learning uncertainty-aware protein fitness landscapes.
- Designing and implementing Bayesian methods for identifying error in experimental assay data.
- Work with wet-lab teams to implement automated analysis pipelines for experimental data, reducing person-hours spent by an order of magnitude.

Broad Institute of MIT and Harvard September 2021 - July 2024
Computational Associate II – Popic Lab

- Researched deep learning approaches to genomic structural variant detection; Led investigation of complex variant detection and benchmarking
- Designed and implemented software infrastructure for scientific pipelines using Python and Bash, processing large-scale genomic/biological datasets for analyses used in publication
- Developed simulation engine for modeling context-aware complex genomic variation and creating synthetic genomes for method development and evaluation
- Led investigations of generative adversarial network (GAN) applications to histology imaging

Brown University – Carney Institute for Brain Science June 2020 - August 2021
Research Assistant to Prof. Jason Ritt

- Investigated dynamical dimensionality reduction of neurophysiological brain models, using recurrent neural networks (RNNs) to solve inverse problem posed by empirical neural spike data
- Implemented custom RNN architectures in PyTorch to test hypotheses of effective forms modeling the dynamical systems underlying the data

MIT Lincoln Laboratory May 2019 - August 2019

Research Intern with Dr. Michael Brandstein

- Applied signal processing and speech recognition techniques to identify bird species from flight call data
- Designed and implemented deep learning speech recognition pipeline to train a classifier on a large dataset of audio recordings
- Combined a CNN image encoder and k-nearest neighbors classifier to process spectrogram image data; Improved on best published accuracy in species-identification benchmark task from 94% to 99%

O&R Patent Law May 2018 – August 2018

Software Intern

- Designed and developed Python-based software tools for enterprise to automate manual tasks in patent law workflow, saving hundreds of person-hours
- Used Python and SQL to automate entity resolution and summarization of records for financial analysis

Publications and Presentations

Popic, V., **Rohlicek, C.**, Cunial, F., Garimella, K., Meleshko, D., Hajirasouliha, I. (2023) **"Cue: a deep-learning framework for structural variant discovery and genotyping."** *Nature Methods.*

Brendel, M. Honigsberg. R, Maharjan S., **Rohlicek, C.** "Weakly-supervised tumor purity prediction from H&E stained slides." Poster presented at 2022 Starr Cancer Consortium Retreat; April 2022; Cold Spring Harbor, NY.

"A deep learning approach to structural variant discovery" and **"Cue: A framework for cross-platform structural variant calling and genotyping with deep learning,"** presented with V. Popic at Models, Inference& Algorithms Initiative interinstitutional seminar series, Broad Institute of MIT and Harvard, February 2022.

Select Projects

insilicoSV: a framework for structural variant simulation

Popic Lab, Broad Institute of MIT and Harvard

Led development of genomic variant simulation engine, providing flexible simulation features for generating synthetic genomes with known mutations.

Molecular Toxicity Prediction

Deep Learning and Special Topics in Data Science, Brown University

Completed research on the use of CNNs and BERT-inspired models for the processing of SMILES molecule representations; implemented a novel strategy for predicting toxicity.

Neural Network Parameter Reduction Using Pruning and Matrix Decomposition

Matrix Methods, MIT (cross-registered)

Conducted experiments in PyTorch comparing effects of magnitude-based pruning techniques to matrix decomposition methods of network size reduction.

Technical Skills

Programming Languages	Python, Java, OCaml, SQL, R, MATLAB, Julia, Bash
Libraries	PyTorch, Tensorflow, Keras, Pysam, Numpy, Pandas, Scikit-learn