Neil Thomas

https://people.eecs.berkeley.edu/~nthomas/

1601 Allston Way
Berkeley, CA 94703

nthomas@berkeley.edu
(310) 699-7973

EDUCATION

University of California, Berkeley 2017 - Present
PhD Student, Computer Science

- My research focuses on developing machine learning methods for learning meaningful representations of proteins, with the aim of enabling applications in protein design, functional annotation, homology detection, and structure prediction.

University of California, Berkeley 2011 - 2015
BS, Engineering Mathematics & Statistics, *High Honors*

Independent University of Moscow, Russia Spring 2014
Study-abroad program: *Math in Moscow*

SELECTED COURSEWORK

Organic Chemistry, Microbial Ecology, Statistical Mechanics, Convex Optimization

PUBLICATIONS

Nicholas Bhattacharya*, **Neil Thomas***, Roshan Rao, Justas Dauparas, Peter K. Koo, David Baker, Yun S. Song, Sergey Ovchinnikov "Interpreting Potts and Transformer Protein Models Through the Lens of Simplified Attention" *Pacific Symposium on Biocomputing (2022)*

Lucy Colwell, **Neil Thomas** "Minding the gaps: The importance of navigating holes in protein fitness landscapes" *Cell Systems (2021)*

Samantha Petti, Nicholas Bhattacharya, Roshan Rao, Justas Dauparas, **Neil Thomas**, Juannan Zhou, Alexander M. Rush, Peter K. Koo, Sergey Ovchinnikov "End-to-end learning of multiple sequence alignments with differentiable Smith-Waterman" *bioRxiv (2021)*

Roshan Rao*, Nicholas Bhattacharya*, **Neil Thomas***, Yan Duan, Xi Chen, John Canny, Pieter Abbeel, Yun S. Song "Evaluating Protein Transfer Learning with TAPE" *Advances in Neural Information Processing Systems 32 (NeurIPS 2019).* Selected as **Spotlight Talk**. (2.4% of submissions)

WORK EXPERIENCE

Research Intern Aug 2021 - Dec 2021
Google, Remote

- Developed open source codebase for benchmarking protein sequence design on synthetic fitness landscapes. https://github.com/google-research/slip

AI Resident March 2021 - Aug 2021
X (formerly Google X), Remote

- Part of an early-stage protein engineering project at the intersection of machine learning and high-throughput experimental characterization. Working to facilitate a shift towards a bioeconomy, low carbon growth and advances in human health.

Software Engineer Aug 2015 - June 2017
23andMe, Mountain View, CA

- Implemented IBD (Identity by Descent) pipeline in Apache Spark to take advantage of data locality and parallelism. Relieved backlog of customer processing.
- Automated imputation pipeline using Minimac3 and Luigi. Imputed over 1 million individuals on custom reference panel with improved accuracy.
- Deployed to AWS to scale computation pipelines to handle massive customer demand.
- Built haplotype phasing application using modified BEAGLE algorithm. Wrote Python wrappers for extracting and deploying legacy research C++ code.
- Responsible for collaborative project planning - distilling input from multiple teams.

Interviewed candidates. Mentored and onboarded new engineers.

TEACHING

Graduate Student Instructor Summer 2022
Introduction to Artificial Intelligence (CS 188), UC Berkeley
- Topics include: Search, Games, Graphical models, HMMs, RL, Machine Learning
- Led 2 weekly discussion sections.

Graduate Student Instructor Fall 2020
Mathematical Statistics (STAT 135), UC Berkeley
- Core upper division course for statistics majors. Topics include: parameter estimation, hypothesis testing, linear regression.
- Led 3 weekly discussion sections, totalling 40 students.

AWARDS

NIH Genomics Training Grant 2018 - 2020
Best Lightning Talk - UC Berkeley Computational Biology Retreat 2019
Winner - 23andMe Hackathon 2016
- Part of a 5-person team that, in 1.5 days, built an interactive way to explore 23andMe genetic reports using the human body.

Honorable Mention (top 40%) - COMAP Mathematical Contest in Modeling 2015
- **Neil Thomas**, Lukas Whaley-Mayda, Miles Rusch. "Leveraging the Criticality of Outbreaks to Eradicate Ebola." February 2015. Simulated vaccine effect on an Ebola outbreak using site-percolation model.

LEADERSHIP & SERVICE

Board Representative, Berkeley Student Cooperative Aug 2020 - March 2021
Member of executive board of $14M/year nonprofit that provides affordable housing to students. Crafted policy and planned budget as part of Capital and Finance Committee.

Network Manager, The Convent, Berkeley Student Cooperative Jan 2018 - Aug 2022
Manage IT infrastructure for 25-person dwelling.

Organizer, Protein ML Reading Group Nov 2019 - Dec 2020
Weekly online reading group focused on machine learning for biology. Facilitated discussion and recruited speakers.

LANGUAGES

Python, R, Java

TOOLS

AWS, Ansible, Apache Spark, Azure, Git, Jenkins, JIRA, LaTeX, MySQL, Packer, PyMol, PyTorch, Tensorflow, Vagrant

FOREIGN LANGUAGES

Russian, *Fluent*

HOBBIES

Ultimate Frisbee, Piano, Cycling, Hiking, Rap, Improv, Dunking†
† - *in progress*