

BRIAN J. ARNOLD, PHD

CONTACT

651-206-2522

brianjohnarnold@gmail.com

brian-arnold.github.io

SKILLS

Python

AI/ML

Pipelines

Data visualization

Git

Docker

Cloud computing

Communication

Collaboration

Detail-oriented

EDUCATION

Harvard University

2009-2015

PhD in Organismic and
Evolutionary Biology

University of Oulu –
Finland

2008-2009

Fulbright Student

University of Minnesota –
Twin Cities

2004-2008

BS in Plant Biology

OBJECTIVE

Cross-functional data scientist with over 9 years of professional experience in bioinformatics, looking to apply expertise in creating data and machine learning pipelines to develop innovative diagnostics that enhance disease understanding and patient outcomes.

EXPERIENCE

Senior Data Scientist, Princeton University

2020-Present

Enabled and accelerated data-science research across Princeton: initiated long-term collaborations with 8 academic groups, designed and taught 6 workshops, provided consulting to 29 graduate students, and managed 2 researchers. *Resulted in 5 peer-reviewed publications (incl. Science) and 1 preprint.*

- Co-developed, optimized, and deployed novel encoder-decoder model that enabled analyses of biological gradients in spatial gene expression data
- Co-engineered and deployed machine-learning pipeline to more accurately detect DNA duplications/deletions in bulk tumor sequencing data; deployed on AWS, GCP
- Spearheaded machine learning analyses and visualizations of 3D videos to map neural activity to animal behavior; identified critical experimental biases and devised statistical approach to correct them
- Optimized and re-tooled data pipelines for two other large-scale projects with 700+ whole-genome and 40+ epigenomic-profiling samples totaling over 10 TB

Senior Bioinformatician, Harvard University

2018-2020

Independent contributor on a team of three scientists facilitating large-scale computational research for biologists: collaborated with 4 academic groups and taught 2 workshops. *Resulted in 3 peer-reviewed publications (incl. Nature Reviews Microbiology).*

- Engineered and published pipeline that massively parallelized software for detecting mutations in whole-genome sequences, enabling >20x speedup
- Designed novel machine learning classifier to detect horizontal gene transfer in whole-genome alignments, revealing significant genetic exchanges in a parasitic flower

Postdoctoral Researcher, Harvard School of Public Health

2015-2018

Designed theoretical and computational approaches to understand how bacterial pathogens adapt to novel environments. *Resulted in 8 peer-reviewed publications and 1 preprint.*

- Awarded competitive, 3-year NIH F32 postdoctoral fellowship
- Designed fast, memory-efficient C++ code to simulate unexplored mechanisms of bacterial adaptation; implemented Bayesian Optimization approach to fit these models to data and quantify rates of adaptation