

Education

University of Washington

Ph.D. in Applied Mathematics

From Mechanism to Practice: Evolutionary Forecasting for SARS-CoV-2

Seattle, WA

12/2024

University of Chicago

B.S. in Mathematics

Chicago, IL

12/2019

Work Experience

Google (via Embedded Resources Group/Yoh)

Senior Research Engineer (Contract)

Remote

5/2025 – present

- Designing and implementing state-of-the-art generative and multimodal model architectures for training, post-training, and inference in JAX, emphasizing modularity, cross-platform (TPU + GPU) performance, and ease of experimentation.
- Building and maintaining a research-driven JAX model garden supporting optimized inference, controlled experimentation, structured post-training (e.g., fine-tuning, quantization), and reproducible benchmarking.
- Integrating model infrastructure into prompting, evaluation, and benchmarking workflows in collaboration with researchers.

Microsoft Research

Research Intern, AI Interaction and Learning Group

Redmond, WA

6/2024 – 5/2025

- Developed interpretable embedding techniques using pretrained language models, aligning them with human-defined concepts (e.g., domain, user intent) to improve retrieval and semantic alignment.
- Built scalable infrastructure to classify and analyze millions of weekly human-LLM interactions, supporting research on interaction patterns, user intent inference, and behavioral loss signals.
- Worked with researchers, data scientists, and product managers to convert large-scale human-LLM interaction logs into insights for agent alignment and behavior evaluation in knowledge work and productivity.

Fred Hutchinson Cancer Center

Post-doctoral Research Fellow, Bedford Lab

Seattle, WA

12/2024 – present

- Founded and led a subgroup on biologically grounded AI, designing an interdisciplinary research program that uses and extends generative language models for phylogenetic analyses, evolution forecasting, pathogen risk assessment, and vaccine design.
- Developed diffusion-based generative models for sequence evolution conditioned on context (e.g., time, geography), enabling controllable forecasting, ancestral reconstruction, and sequence editing.
- Built interpretable, evolution-aware embeddings from genomic LLMs for real-time pathogen tracking via Nextstrain, offering a scalable deep learning alternative to phylogenetics.
- Fine-tuned genomic foundation models for monitoring evolution in biological sequences and natural language; explored interpretable deployment for public health monitoring and vaccine targeting.

Graduate Research Fellow, Bedford Lab

9/2020 – 12/2024

- Led development of ML pipelines and evaluation frameworks for real-time evolutionary forecasting to inform vaccine strain selection.
- Designed and deployed real-time predictive models for epidemic trends and evolution in respiratory viruses, improving existing methods to scale analyses to 2+ million genetic sequences globally, surpassing prior scalability limits of 10,000 sequences.
- Mentored 8 junior researchers in mathematics, epidemiology, and biology on open-source applied scientific projects, emphasizing reproducibility, collaborative development, and production-grade ML workflows (CI/CD, version control, unit testing).

Technical Skills

Programming & Tools: Python (expert), R (advanced), Golang (intermediate), C++ (beginner), Git, Docker, AWS, and Azure.**Machine Learning & AI:** Supervised, unsupervised, and reinforcement learning including fine-tuning diffusion models and LLMs.**Statistics:** Bayesian inference, causal inference, survival analysis, time series analysis, and hierarchical modeling.**Applied mathematics:** Optimization methods, approximation theory, algorithm development, and numerical methods.**Data & Feature Engineering:** Designed and implemented data pipelines with SQL, Pandas, Polars, PySpark.**Evaluation & Experimentation:** Back-testing, cross-validation, model comparison, sensitivity analysis, and A/B testing.

Software and Open-source Projects

[1] **forecasts-ncov:** Real-time Pathogen Forecasting (github.com/nextstrain/forecasts-ncov)

- Led development of automated data curation, analysis, and prediction pipelines for estimating SARS-CoV-2 variant transmission advantages from millions of sequences.
- Collaborated with Nextstrain software engineers and scientists to build an open-source global surveillance tool for public health.

[2] **evofr:** Evolutionary Forecasting Toolkit (github.com/blab/evofr)

- Created and led development of evofr, a production-level Python library for statistical modeling, uncertainty quantification, and Bayesian inference on genetic and epidemiological data using JAX and NumPyro.
- Designed for real-time inference on large genomic data sets (> 2 million sequences) and time series with a focus on usability in public health settings.

[3] **attrembed:** Compact, attribute-aware text embeddings (github.com/marlinfiggins/attrembed)

- Developed a Python library for compressing pretrained embeddings into interpretable, attribute-aligned subspaces, enabling concept-based retrieval, clustering, and classification across flat and hierarchical metadata.