Jonathan Deaton

Website | LinkedIn | GitHub | Scholar

1125 Park Place, Apt 215 San Mateo, California
jonpauldeaton@gmail.com | (413) 531-1568

EXPERIENCE

Research Scientist (previously Research Engineer)   November 2021 – Present
X Development (Google X)   *Mountain View, California*

- Delivered diverse machine learning approaches for real-world biomolecular sequence design problem including:
    * Fine-tuned BERT-like model which significantly outperformed our team's existing methods in molecular design by increasing experimental hit-rate from 58% to 96%
    * Self-supervised pre-training of BERT on large biological nucleotide sequence datasets, eliminating dependency on data collection from an unreliable experimental assay through transfer learning
    * Decoder-only LM with RLHF fine-tuning to accelerate generation of biologically plausible candidates
    * Investigated zero-shot capabilities of RNA language model in screening loss-of-function mutations
- Designed and built scalable distributed training system in JAX supporting Language Models. Accelerated rapid prototyping of ML algorithms for molecular design by integrating with T5 architectures, speeding up training by 2-5x, and increasing algorithmic expressiveness and data-scaling

Software Engineer   September 2019 – November 2021
Google Health / Google Research   *Palo Alto, California*

- Led development of software to validate computer-vision models for dermatological condition classification
- Contributed to productionization of the DermAssist ML model thereby achieving regulatory approval as a CE-marked Class I medical device
- Conducted research advancing state-of-the-art computer-vision models and published research findings on long-tail class-imbalance

Graduate Research Assistant   January 2018 – April 2018
Stanford Network Analysis Platform   *Stanford, CA*

- Transformed 2TB of Reddit user data by building a highly parallelized processing system. Enabled construction of user-interaction graph

Undergraduate Bioengineering Researcher   June 2014 – June 2017
Quake Lab   *Stanford, CA*

- Identified novel bacteriophages in metagenomic sequence data by developing data-analysis, visualization, and ML software in Python
- Performed molecular biology laboratory experiments including cell culture, qPCR, DNA sequencing, microfluidic device fabrication
- Communicated scientific findings in academic conferences, lab meetings, and publications

Research Associate Intern   June 2016 – September 2016
Illumina, Protein Engineering Group   *San Diego, California*

- Developed algorithm in Python for quantifying decay in DNA sequencing data quality
- Performed laboratory experiments to optimize assays for human exome sequencing

EDUCATION

Stanford University   January 2018 – June 2019
MS in Computer Science, Artificial Intelligence   *Stanford, California*

- GPA: 3.957 / 4.0
- Course Assistant: CS 140: Operating Systems, CS 161: Design and Analysis of Algorithms
- Coursework: Probabilistic Graphical Models, Machine Learning Theory, Deep Learning, Reinforcement Learning

Stanford University   September 2012 – March 2018
BS in Bioengineering with Honors   *Stanford, California*

- GPA: 3.839 / 4.0
- Research thesis: Identified Bacteriophage genomes in metageonmic sequencing data with machine learning

Publicaitons

- Big Self-Supervised Models Advance Medical Image Classification. Azizi S, Mustafa B, Ryan F, Beaver Z, Freyberg F, **Deaton J**, Loh A, Karthikesalingam A, Kornblith S, Chen T, Natarajan V, Norouzi M. International Conference on Computer Vision 2021.

- Underspecification Presents Challenges for Credibility in Modern Machine Learning. D'Amour A, Heller K, et al. Journal of Machine Learning Research, 2020.

- Addressing the Real-world Class Imbalance Problem in Dermatology. Weng WH, **Deaton J**, Natarajan V, Elsayed GF, Liu Y. NeurIPS Machine Learning for Health Workshop 2020.

- Mini-Metagenomics and Nucleotide Composition Aid the Identification and Host Association of Novel Bacteriophage Sequences. **Deaton J**, Yu F, Quake S. Advanced biosystems, 2019.

Projects

Transformer Pointer-Generator Networks for Summarization   Stanford CS 224N Course Project

- Implemented fusion of Transformer Encoder/Decoder and pointer-generator network for text summarization
- Successfully trained novel deep network architecture with low compute resources

AgarLE: Competitive Multi-Agent RL Environment   Personal/FOSS Project

- Build high-performance multi-agent reforcement learning environment in C++
- Benchmarked state-of-the-art RL methods such as Proximal Policy Optimization and Deep-Q Learning

Heat Stroke Risk Monitor   Bioengineering Senior Capstone

- Led team of three in designing, building, and testing wearable heat stroke risk monitor
- Developed and validated machine learning algorithm to predict user risk in real time
- Featured at Rice 360 Global Health Competition, 2017 Tau Beta Pi Engineering Showcase