

Ross Everett Altman

+1-516-375-4505 | ross.e.altman@gmail.com | github.com/knowbodynos | linkedin.com/in/ross-altman
bit.ly/ross-altman-publications | bit.ly/ross-altman-patents

Summary

Data scientist and full-stack ML engineer with 6+ years of industry experience. PhD computational and mathematical physicist by training. Self-starter, lifelong learner, and teller of bad puns. Passionate about working alongside people with diverse experience to build interdisciplinary knowledge and deliver impactful results. Solving complex problems through data, algorithms, teamwork, and strong research fundamentals.

Work Experience

Inari Agriculture

Cambridge, MA

Machine Learning Engineer

Jan 2024 - Present

- Led a cross-functional team of ML engineers and biologists to train a foundation LLM for plant genomes and fine-tune LoRA adapters for tissue-specific gene expression, achieving a 56% accuracy gain and directly informing target prioritization for downstream gene editing experiments.
- Partnered with biologist stakeholders to design, deploy, and scale an in-house protein engineering platform in Airflow using AlphaFold and generative models, enabling rapid screening of proprietary CRISPR-Cas enzymes and eliminating a \$50k annual vendor subscription.
- Developed a high-throughput protein-protein interaction AlphaFold pipeline to pre-screen potential pathway interventions for trait improvement in soybean, reducing the number of candidates to be tested in the lab by 100x.
- Implemented a cross-species transfer learning pipeline using a modified variational autoencoder to leverage high-quality RNA sequencing data from well-studied species to augment sparser data in soybean, enabling more accurate gene regulatory network inference for target discovery.
- Developed a novel and robust method to extract quantitative physiological features from hundreds of noisy soybean root images using a combination of computer vision and network analysis methods.

Data Scientist

Jan 2019 - Jan 2024

- Developed a high-throughput computational screen for estimating phenotypic impact of CRISPR-Cas edited alleles in plants using a mixture of protein language models and bioinformatic methods.
- Fine-tuned a protein language model to discover novel stability-enhancing mutations for improving CRISPR-Cas nuclease efficiency for genome editing in plants.
- Spearheaded the application of deep learning for protein sequence-to-function modeling and gene regulatory network perturbation-response modeling, demonstrating its potential and leading to the establishment of a full ML/AI team.
- Mentored 10+ junior colleagues, interns, and data science co-op students.

Insight Data Science

Boston, MA

Data Science Fellow

Sept 2018 - Dec 2018

- Implemented a Word2Vec model to enhance Wikipedia page previews by displaying the most relevant information to the current page.
- Used Wikipedia's bi-directional clickstream data to validate a 17% increase in relevance over default page previews.
- Built a Chrome browser extension served using a jQuery/Flask/AWS stack to achieve high performance on real-time, client-side data.

Northeastern University

Boston, MA

Graduate Research Assistant

Sept 2011 - May 2017

- Designed a distributed algorithm using C++ and Python to systematically compute type-IIB string theory vacuum states on an HPC cluster.
- Developed a custom Python-based load balancer optimized for distributed computing across multiple SLURM clusters with different resource constraints and usage patterns.
- Compiled the world's largest database of 100k string theory vacua, resulting in discovery of candidate universes extending the Standard Model.
- Estimated the number of vacua ($> 10^{10,000}$) in the string landscape using deep neural networks using specialized "equation learner" layers.
- Built and maintained a queryable web UI with MongoDB backend based on feedback from stakeholders in the string theory community.
- Co-founded the String Data group for ML applied to string theory, leading to international conferences and collaboration with industry.

Education

Northeastern University

Boston, MA

PhD in Physics

Sept 2011 - May 2017

- Focus:** String theory/phenomenology, high-dimensional geometry, topology, high-performance computing, machine learning.

Cornell University

Ithaca, NY

MEng in Applied Physics

Sept 2010 - May 2011

Cornell University

Ithaca, NY

BSc in Applied & Engineering Physics

Sept 2005 - May 2009

Skills

Programming Languages

Python, C/C++, Shell, HTML/CSS, JavaScript.

Machine Learning Stack

PyTorch, TensorFlow/Keras, Scikit-learn, OpenCV, Huggingface, Lightning AI, MLFlow.

Engineering Stack

AWS, UV, Docker, Kubernetes, Airflow, Flask/FastAPI, Terraform, MongoDB, CICD.

Technical Skills

LLMs, Transformers, Foundation models, Variational Autoencoders, LoRA, Geometric Deep Learning, NLP, Computer Vision, Network Analysis, Statistics, Bioinformatics.

Soft Skills

Self-motivation, Cross-functional collaboration, Scientific communication, Interpersonal skills, Rapid prototyping.