Ross Everett Altman

+1-516-375-4505 | ✉ ross.e.altman@gmail.com | ⌂ github.com/knowbodynos
linkedin.com/in/ross-altman | 🎓 Publications | Ⓖ Patents

Summary

Principal Machine Learning Engineer and Data Scientist bridging research depth with startup execution. 10+ years building and deploying scalable data and ML systems across physics and biotech, spanning statistical analysis, interpretable data modeling, and production ML. Experienced technical lead for generative design, transfer learning, and foundation model initiatives driving high-stakes scientific and product decisions. Champions the application of interdisciplinary knowledge from mathematics, physics, and biology to adapt mature frameworks to new problems and unlock impactful insights into complex data.

Experience

Inari Agriculture, Cambridge, MA   Jan 2019 – Present
Machine Learning Engineer / Data Scientist

- Expanded scope from applied data science (statistical modeling and inference) to end-to-end ML engineering and technical leadership, setting direction and coordinating cross-functional collaborators across research and engineering.
- Led the training of a biologically-informed foundation LLM for plant genomes and fine-tuning of LoRA adapters, improving gene expression prediction by **25% Spearman correlation** and automating predictive design pipelines across **five research programs**.
- Architected and deployed a secure, in-house protein design platform using Airflow, AlphaFold, and generative models enabling high-throughput generation and efficiency screening of synthetic CRISPR enzymes, saving **$50K annually** in vendor licensing costs.
- Built a high-throughput AlphaFold protein-protein interaction screening workflow that cut experimental resource requirements by **97%**, accelerating novel target discovery.
- Developed a cross-species transfer learning method leveraging multimodal VAEs to augment sparse soybean scRNA-seq datasets by **300%**, improving gene network inference accuracy.
- Launched a high-throughput, zero-shot ML screen for phenotypic impact of CRISPR edits, expanding model-assisted product advancement pipelines to **46%** more plants and reducing the burden on greenhouse capacity.
- Automated quantitative root phenotyping using computer vision and graph analysis, **transforming 1,000+ noisy images** into interpretable feature sets for trait scoring.

Insight Data Science, Boston, MA   Sept 2018 – Dec 2018
Data Science Fellow

- Build a Word2Vec-based pipeline to boost the content relevance of Wikipedia page previews, resulting in a **17% improvement** in clickstream engagement.
- Developed and deployed a Chrome extension using Flask and AWS for real-time user-side data retrieval and ranking.

Northeastern University, Boston, MA   Sept 2011 – May 2017
Graduate Research Assistant, PhD Physics

- Developed high-performance, distributed C++/Python algorithms to construct a massive **database of 100k vacuum state solutions** compatible with Type-IIB string theory, effectively bridging high-dimensional geometry with the Standard Model.
- Applied deep "equation learner" networks to extrapolate the solution space structure of the full string landscape.
- Built and maintained data platforms and search tools used by international research collaborators.
- Co-founded an interdisciplinary research effort focused on applying ML to fundamental physics problems.

Education

PhD, Physics, Northeastern University — 2017
- **Focus:** String theory, topology, machine learning.

MEng, Applied Physics, Cornell University — 2011
BSc, Applied & Engineering Physics, Cornell University — 2009

Skills

Programming:	Python, C++, Shell, SQL, PyTorch, TensorFlow, HuggingFace, OpenCV, FastAPI
Core ML:	Foundation Models, LLMs, LoRA, Transformers, Diffusion Models, VAE, GNN, Decision Trees, Statistical Modeling, Representation Learning, Causal Inference
MLOps:	MLFlow, Airflow, Docker, Kubernetes, AWS, Terraform, CI/CD, UV, Model Deployment, Data Pipelines
Bioinformatics:	Protein Modeling, RNA-seq Analysis, Network Biology, Automated Phenotyping, Gene Regulation Modeling