

# Ryan Engel

(516) 242-9323 | Email: ryan.engel30@gmail.com | Website: ryanengel.info

## Education

**Stony Brook University:** M.S. in Computer Science *Sept 2024 - May 2026 (Expected)*  
*Thesis on Scalable Factor Models for Portfolio Optimization*

**Stony Brook University:** B.S. in Computer Science *Sept 2020 - May 2024*  
*Specialization in Artificial Intelligence and Data Science*

## Research Publications

Ryan Engel, Yu Chen, Pawel Polak, Ioana Boier. 2025. Scaling Conditional Autoencoders for Portfolio Optimization via Uncertainty-Aware Factor Selection. In *Proceedings of the 6th ACM International Conference on AI in Finance (ICAIF '25)*, pages 123–131, Singapore. ACM.

Ryan Engel and Gilchan Park. 2024. Evaluating Large Language Models for Predicting Protein Behavior under Radiation Exposure and Disease Conditions. In *Proceedings of the 23rd Workshop on Biomedical Natural Language Processing*, pages 427–439, Bangkok, Thailand. Association for Computational Linguistics.

## Work Experience

**NVIDIA Corporation:** Software Engineer Intern *May 2025 – September 2025*

- Contributed to the development of scalable, cloud-native microservices for LLM training and evaluation, while optimizing and maintaining distributed GPU clusters deployed with Docker, Helm, and Kubernetes.
- Demonstrated an end-to-end pipeline for deployment, training, and evaluation of the Qwen-3 LLM on the Medical-o1-Reasoning dataset using NeMo Microservices and Nvidia's Multi-LLM NIM.

**Proprietary Hedge Fund (New York):** Quantitative Research Intern *August 2024 - May 2025*

- Designed a scalable framework for asset pricing and portfolio optimization using time-series foundation models and uncertainty quantification.
- Achieved risk-adjusted performance of 2.204 Sharpe ratio, 4.01 Sortino ratio, and 5.952 Omega ratio, delivering 14.37% annualized returns with maximum drawdown of 9.22% across a 25-year backtest (2000-2024).

**Brookhaven National Laboratory:** Machine Learning Research Intern *June 2023 - August 2024*

- Evaluated the Mistral, Llama-2, and Llama-3 LLMs for predicting protein behavior, achieving state-of-the-art results in multiple experiments, with accuracy improvements of up to 10% over baseline models. This work was presented at the BioNLP workshop during the ACL 2024 conference.
- Researched Parameter Efficient Fine-Tuning (PEFT) techniques for LLMs. Compared Low-Rank Adaptation (LoRA), Prompt-Tuning, and Prefix-Tuning of various LLMs on the PubMedQA Dataset.

## Projects

### Fault-Tolerant Distributed Transaction Processing System in Rust

- Engineered a high-performance sharded banking application achieving 3,000+ TPS across 9 nodes, using Multi-Paxos consensus for intra-shard replication and two-phase commit protocol for atomic cross-shard transactions.
- Implemented leader election with view-change protocols and hypergraph-based dynamic resharding to minimize cross-shard overhead across 3 clusters managing 9,000 data items.

### Portfolio Optimization with Time-Series Foundation Model

- Placed as a finalist for the 2024 SQA Alphathon competition (Quantitative Finance Hackathon) out of 77 teams (148 participants) by developing a portfolio optimization algorithm using the Chronos-Bolt TSFM.

### Search Engine for National Electric Codebook

- Built a full-stack semantic search engine using React, Flask, Docker, and FAISS in a retrieval-augmented generation (RAG) framework deployed on Linode cloud services, to query the NFPA 2017 Electric Codebook.

## Technical Skills

- Programming Languages:** Python, Rust, C, Java, Javascript, C++, OCaml, Go, Haskell, MIPS Assembly
- Libraries:** PyTorch, Pandas, Numpy, Transformers, PEFT, Langchain, Scikit-learn, Matplotlib
- Tools & Platforms:** Linux, Docker, Kubernetes, Helm, Nginx, Git, SLURM, QuantConnect