

Brendon Reperttang

ML Engineer & Founder
(619) 895-0022
brendonrt@gmail.com
<https://www.linkedin.com/in/brendon-rt>

EDUCATION

Massachusetts Institute of Technology
Cambridge, MA
Bachelor of Science in Computer Science and Engineering

December 2023

SKILLS

Computer Languages & Software: Python, Mojo, Type/JavaScript, Java, Rust C++, C, C#, \LaTeX , Lua, Julia, Jupyter, Gurobi, RISC-V Assembly, CLIs, GitHub, Docker, Kubernetes, Windmill, CSS, HTML, Svelte, PostgreSQL, MongoDB, Click-House

Specializations: LLMs and Generative AI, Pipelining and Architectural Patterns, Category Theory, Meta-learning, NAS, Reinforcement and Multi-Agent Learning, Ensemble Learning, Forecasting and Anomaly Detection, Computer Vision, Data Synthesis and Feature Engineering, MLOps, Research Paper Implementation, Unit and Integration Testing, Ablation and Incremental Studies

Relevant Coursework: Deep Learning, Computer Vision, Machine Learning, Multi-Agent Learning, Software Construction, Fundamentals of Programming, Linear Algebra, Multi-variable Calculus, Math for CS, Algorithm Design, Computational Structures, Embedded Systems, Computer Systems Engineering, Optimization, Modeling for ML, Software Studio

WORK EXPERIENCE

AI Research Associate
Society for AGI, Pasadena, CA

June 2024 - September 2024

- Discussed and proposed LLM research and evaluation methods from a Category and Control Theory perspective
- Experimented with compute efficiency focused research papers, resulting in iso-FLOP to accuracy improvements of over 40 percent for the best ablated component combination, while also holding parameter count constant
- Enhanced depth of ML architecture expertise using Flax

Zephyr AI Researcher
DAI Lab, Cambridge, MA

August 2023 - December 2023

- Studied the effects of including static covariates on time series data forecasting, increasing accuracy by 16 percent over the baseline XGBoost without covariates
- Deployed Temporal Fusion Transformer models as a pipeline with MLBlocks
- Theoretical propositions and experiments for other potential predictive maintenance improvements on error code predictions

Machine Learning Engineer and Data Analyst
Bamboo Mortgage, Portland, OR

May 2020 - August 2021

- Created a random forest estimator model to find high ROI housing investments
- Informed acquisitions of 6 properties with a hybrid model/heuristic system
- Scraped data and compiled expenses to allow for investment analysis and cash-flow analysis

Web Development Intern and Financial Assistant
Bamboo Mortgage, Portland, OR

June 2016 - July 2020

- Solely reformatted the company website using HTML and JavaScript
- Organized and kept track of hundreds of financial documents for many clients
- Assisted in organizing over a dozen open house events with Xiao Realty

OTHER EXPERIENCE

Niru ML & Automated Machine Learning Operations

March 2024 - Present

Designed and built a production-grade MLOps framework in Flax/JAX featuring a strongly-typed component registry, factory-based pipeline instantiation, and automated model lifecycle management. Developed novel tabulated model representations enabling programmatic architecture manipulation, to perform comprehensive ablation and incremental studies. Integrated external LLM systems for RL-driven evolutionary NAS with multi-objective optimization, balancing model performance metrics against computational efficiency through weighted aggregation of signals. The automation pipeline manages end-to-end workflows from data preprocessing through distributed training to post-training evaluation, ensuring reproducibility through declarative Blueprint configurations that specify model architecture and training constraints. The modular platform supports mixed-precision training, parameter-efficient fine-tuning, and seamless integration of emerging techniques through kernel and JAX JIT implementations. Currently extending toward agentic multi-agent training systems with reinforcement learning optimization. Demonstrates advanced expertise in NLP, distributed training, empirical evaluation, and production DevOps.

Multi-Agent LLM Truth Elicitation

September 2023 - December 2023

In 6.S890, a special subject in Multi-Agent Learning, collaborated with PhD students Charlotte Siegmann and Stewart Slocum, the aim of the project was to verify theoretical claims about applying a Bayesian Truth Serum (BTS) to a multi-agent LLM system. Summarizing the groundwork for this BTS setup, provided by Ray Weaver and Drazen Prelec, given questions $q \in Q$ with a finite number of unique answer choices, a query is made to each agent for their answer as well as a guessed probability mass for every answer option (the predicted chance that others would output that). Leveraging this unique approach to truth elicitation with theoretical guarantees of a Pareto-Dominant Nash Equilibrium, many setups contained statistically significant results, with p-values at most $7 * 10^{-19}$ for the best performing BTS setups compared to the control of querying the question by itself. For context, on the Massive Multitask Language Understanding dataset, the Brier score went from .074 to .035 on GPT-4, a significant improvement on the proper scoring metric. These results were achieved with only a small number of agents at inference time, and there is feasibility of applying this approach to pre-training tasks, which could yield even larger performance gains. This experience provided knowledge about the nuances of modularity and automation, advanced game theory theorems for extensive form games, and statistical analysis techniques.

Recurrent Neural Network for Online Market Futures

February 2023 - May 2023

In 6.S052, a special subject ML project class, a long short-term memory recurrent neural network was created to perform market analysis in multiple decentralized online markets. There exists a major opportunity for larger profit margins in these markets, due to the saturation of algorithmic trading on the stock market. A simulated 60-day trading run with price, demand, and volume data from the Steam Community Market, provided an average profit of over 7,000 USD, starting with only 2,500 USD. To get these results, state-of-the-art RNN techniques were used, along with lesser-known techniques, and even some novel ones. An uncommon technique of encoding important conditional and static features directly into the time-series data is done by affinely transforming the input vector at $t = 0$. This, paired with a novel static feature vector formulation, which provides a unique encoding for the set of commodities focused on (Counter Strike skins), average profit improved by over 2,700 USD over the baseline RNN. For selection, ROI slopes are generated from the outputted price predictions, top- k ROI peaks are selected, and at times corresponding to peaks in this list, the points are regenerated. If any updated slopes from this current point are found to be higher than the prediction, the asset is held, otherwise, it is listed. This project provided experience in designing custom ML pipelines and adapting architectures to novel financial applications.