Brendon Reperttang

(619) 895-0022
brendonrt@gmail.com
6717 Friars Road Unit 85, San Diego, CA 92108

EDUCATION

Massachusetts Institute of Technology December 2023
Cambridge, MA
Bachelor of Science in Computer Science and Engineering

SKILLS

Computer Languages & Software: Python, Mojo, Type/JavaScript, Java, Lua, C++, C, C#, LaTeX, Julia, RISC-V Assembly, Jupyter, Gurobi, Linux/Unix/Windows CLI, GitHub, Docker, Kubernetes, CSS, HTML, React/Angular, PostgreSQL, Prisma, Supabase, n8n, Windmill

Specializations: LLMs and Generative AI, Category Theory and Algebraic Abstraction, Supervised and Unsupervised Learning, Recurrent Neural Networks and Anomaly Detection, Reinforcement and Multi-Agent Learning, Data Synthesis and Processing, MLOps and Pipelines, Computer Vision, Research Paper Implementation, Unit and Integration Testing

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

Languages: English (Fluent), Mandarin (Fluent), Cantonese (Semi-fluent)

WORK EXPERIENCE

Research Collaborator June 2024 - September 2024
AGI Society, Pasadena, CA
- Discussed and proposed LLM research and evaluation methods from a Category and Control Theory perspective
- Implemented and iterated on compute efficiency focused research papers
- Improved depth of ML architecture expertise using functional programming-based library, JAX

Zephyr AI Researcher August 2023 - December 2023
DAI Lab, Cambridge, MA
- Studied the effects of static covariates on time series data forecasting.
- Deployed Temporal Fusion Transformer models as a pipeline with MLBlocks
- Model evaluation and reporting with WandB, data collection with MongoDB
- Theoretical propositions and experiments for potential predictive maintenance improvements on error code predictions

Machine Learning Engineer and Data Analyst May 2020 - August 2021
Bamboo Mortgage, Portland, OR
- Entrusted to analyze the housing market, proposing options for investment
- Created a random forest estimator model to find high ROI options
- Informed acquirements 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 June 2016 - July 2020
Bamboo Mortgage, Portland, OR
- Solely reformatted the company website using HTML and JavaScript
- Organized and kept track of financial documents for many clients
- Assisted in organizing open house events in conjunction with Xiao Realty

OTHER EXPERIENCE

Efficient Transformer Project February 2024 - Present

I am developing a SOTA language model in Flax (A flexible end-to-end ML framework using JAX), utilizing cutting-edge techniques to optimize performance and efficiency. The model architecture is based on the Transformer, with innovations in both pre-training and fine-tuning stages. For pre-training, I implemented a BERT-inspired approach with masked language modeling, utilizing dynamic masking for token obfuscation. This approach, combined with various sampling techniques, and an optimized batch size of 512, significantly improved training efficiency. The model employs Grapheme Pair Encoding (GPE) for tokenization, hypersphere normalization (nGPT), multi-head attention mechanisms, and RELU^2 activations with dropout. The project will also feature mixed-precision training, novel fine-tuning processes, and advanced attention mechanisms such as SageAttention combined with Differential Transformer. My modular approach facilitates experimentation with emerging techniques and enables comprehensive ablation studies. This work demonstrates expertise in advanced NLP, efficient model design, rigorous empirical evaluation, and best practices for large-scale DevOps (a secondary goal of mine for the project).

Multi-Agent LLM Truth Elicitation September 2023 - December 2023

For 6.S890, a special subject in Multi-Agent Learning, I paired up with Charlotte Siegmann and Stewart Slocum, both PhD students, to verify the theoretical claims about applying a Bayesian Truth Serum (BTS) to a multi-agent LLM setup. 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, we query each agent for their answer as well as a 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), we evaluated setups with statistically significant results, with p-score values at most $7 * 10^{-19}$ for outlier BTS setups compared to the control of querying the question by itself. For context, on the Massive Multitask Language Understanding dataset, an important benchmark for STEM knowledge, the Brier score went from .074 to .035 on GPT-4, a significant improvement on the proper scoring metric. This result is only from prompting a small set of agents. and we proved the feasibility of applying this setup to pre-training tasks such as next-word prediction, which provides larger potential for model performance. I am grateful for my amazing coworkers during this time, where I garnered a lot of knowledge about the nuances of modularity and automation in the context of Transformer architectures, advanced game theory theorems for extensive form games, and statistical analysis techniques.

Recurrent Neural Network for Online Market Futures February 2023 - May 2023

For 6.S052, a special subject ML project class, I created a long short-term memory recurrent neural network to perform market analysis in multiple decentralized online markets. I saw a major opportunity for larger profit margins in these markets, due to the saturation of algorithmic trading on the stock market. I simulated a 60-day trading run with price, demand, and volume data from the Steam Community Market, with an average profit of over 7,000 USD starting with 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. I used an uncommon technique of encoding important conditional and static features into the time-series data by transforming the input vector at $t = 0$. Paired with my novel static feature vector formulation, which provides a unique encoding for the set of commodities I focused on (Counter Strike skins), average profit improved by over 2,700 USD. For selection, I generated ROI "slopes" from the price predictions outputted, took the top-k ROIs, and at times corresponding to ROIs in this list, the points are regenerated. If any updated slopes from this current point are found to be higher than the prediciton, the asset is held, otherwise, it is listed. This project played a large role in making me well-informed regarding creating a custom pipeline and solution for nearly any ML architecture.

OTHER EXPERIENCE CONT.

Cost of Living Optimization Analysis October 2022 - December 2022

With rising costs of living being an increasingly primary concern, I performed an detailed and novel CoL analysis with a dataset compiling a broad set of USD-adjusted costs for given commodities and securities for nearly every city in the world. I performed a linear optimization analysis in Julia, using Gurobi, to generate purchasing power scores for each city, assigning 100 to the city with the lowest costs. For a given score x, and spending portfolio, your money gets you x percentage of goods compared to the highest scoring city. I performed the analysis in the context of multiple example portfolios, in which an example percentage budget for each type of expense weights the impact of cost in a given category differently. A strong point of this final project is the high quality and depth of the analysis, providing subsets of cities with high ROI to average income ratios for differing spending portfolios. Given the number of plots, cases, and different formulations needed, this was the first time I started implementing custom functions and scripts to automate generating figures in Matplotlib. Overall, this project greatly improved my data analysis, automation, and LP formulation skills.

Computer Vision Experimental Project March 2022 - May 2022

As a final project in Advances in Computer Vision, I narrowed down my focus to a specific problem for image classification, which is the problem of maintaining accuracy when the number of classes grows arbitrarily large. As an experiment, I tested the accuracy of a modified resnet classification model in the context of the Fashionpedia clothing dataset with a large number of classes and attributes. Knowing that one of the natural bottlenecks for many-class image classification is the quality of the dataset, made clear by Meta AI's Segment Anything, I primarily tested the impact of affine image transformations to create more data points, which ended up improving the mean intersection over union score, which was harshly evaluated class by class, and attribute by attribute. Nearly every pixel contains more than 1 classification, and the evaluation punished both false positives and false negatives. This was one of my first times actually modifying the architecture of deep models, custom data pre-processing and loading, running a pipeline, and performing evaluations, all in a solo project.

PLHS Robotics Team Lead Developer October 2016 - June 2018

I was elected the lead developer in the Point Loma High School Robotics Team, a small team that hovered around 6 members. I learned concurrency, memory management, and general problem-solving in C++ while making a robot fit to compete in the FIRST robotics competition, a difficult task considering the funding and size of our team. In order to reach the deadline while having a fairly competitive robot, I needed to assign and manage tasks best suited to each member who could code. To ensure that we were productive and efficient, later into semesters, I would host online meetings to recap progress on code the night before the first in-person meeting of the week. My early experience leading and learning how to code in a fairly difficult language gave me a rock-solid foundation in coding and group work before my undergraduate studies.