Matthew Harvill

Stanford, CA | 512-578-9123 | mharvill@cs.stanford.edu | GitHub | LinkedIn

EDUCATION

M.S. Computer Science, Stanford University | June 2024 | GPA: 4.03/4.30

Spoken Language Processing, Parallel Computing, Systems for Machine Learning, Deep Reinforcement Learning, Computer Vision, Decision Making Under Uncertainty, ML with Graphs, Natural Language Processing, Mining Massive Datasets

B.S. Computer Engineering, University of Texas at Dallas | Graduated May 2022 | GPA: 3.99/4.00

Operating Systems, Embedded Systems, Computer Architecture, Data Structures and Algorithms

EXPERIENCE

Teaching Assistant - Stanford University. Stanford, CA | September 2022 - June 2024

- Led successful quarterly course-review lectures and weekly discussion/problem sections for Probability and Statistics, and Computer Organization and Systems
- Recognized twice by Professor as an exceptional teacher in Probability and Statistics

Machine Learning Engineer Intern - Hippocratic AI. Palo Alto, CA | June 2023 - September 2023

- Built and oversaw the company's large language model (LLM) evaluation framework on Amazon Web Services (AWS)
- Trained hundreds of healthcare-specific 7B+ parameter LLMs on AWS SageMaker and SLURM
- Researched custom healthcare tokenizer cost/performance tradeoffs, leveraging techniques like TFIDF
- Greatly reduced data processing latencies by migrating pipelines to containers on AWS SageMaker

Software Engineer Intern - Texas Instruments, Inc. Dallas, TX | May 2022 - August 2022

- Founded Texas Instruments's (TI) Wi-SUN Embedded Host application in C/C++, allowing customers to fully integrate the new Wi-SUN communication standard into existing large-scale applications
- Co-created TI's Wi-SUN ReactJS + NodeJS web application for intuitive network management across hundreds of nodes

Test Engineer Intern - Texas Instruments, Inc. Dallas, TX | June 2021 - August 2021

- Generated C++ test plans to verify various aspects of hardware functionality on a lowside driver board
- Detected elusive board issues while running tests on automated testing equipment

PROJECTS

Test-Time Training for Speaker Adaptation in ASR Systems

- Achieved absolute Word Error Rate (WER) reductions of 2-3% on unseen out-of-distribution (OOD) audio samples and 5-20% on seen OOD audio samples using custom Encoder-Decoder ASR System with Test-Time Training

Smarter Atari Agents with Efficient(est)Zero

- Improved EfficientZero's State of the art (SoTA) performance with dynamic parallel Monte Carlo Tree Search (MCTS)
- Researched a novel state-based similarity sharing technique to improve MCTS, but found it is inhibitively slow

Designing a Reliable *Crew* Member

- Designed a MCTS agent for the cooperative trick-taking game, *The Crew*, that significantly outperformed a random agent and was robust to unreliable teammates
- Built the simulation environment for playing *The Crew* and testing agent performance from scratch in Python

Cryptocurrency Fraud Detection with GNNs

- Applied Recurrent Neural Network (RNN)-based Graph Neural Networks (GNNs) to predict illicit crypto transactions
- Discovered that RNN-based GNN models do not improve performance over their vanilla counterparts

Multitasking with BERT Embeddings

- Performed in the top quartile of Stanford class on sentiment classification, paraphrase detection, and semantic textual similarity tasks by fine tuning BERT embeddings with various strategies, including incorporating Jaccard Similarity metrics

Coup for all | https://coup-for-all-8b5265fb626c.herokuapp.com

- Architected a secure, fault-tolerant, real-time, mobile-friendly public website for playing the popular game *Coup*
- Built everything from scratch, including all frontend designs, using ReactJS, NodeJS, websockets, and MongoDB

SKILLS

Programming Languages - Python, C/C++, Bash, CUDA, SQL, JavaScript/TypeScript, Java
Frameworks/Libraries - PyTorch, Numpy, PySpark, OpenMP, Pandas, NodeJS, ReactJS
Platforms/Tools - AWS SageMaker, AWS Simple Storage Service (S3), AWS Elastic Cloud Compute (EC2), AWS Elastic Container Registry (ECR), SLURM, Docker, GitHub, JIRA, Heroku, MongoDB, PostgreSQL
Operating Systems- MacOS, Unix/Linux, Embedded Linux, Windows