https://andrewnc.github.io

Andrew Carr

andrewcarr06@gmail.com

Relevant Experience

Gretel.AI - Senior Applied Research Scientist
Jan 2022 - Present

- Senior scientist on the Machine Learning team working to improve privacy preserving data synthesis for multiple modalities with a focus on natural language and tabular data

OpenAI - Member Of Technical Staff (Fellow)
May - Nov 2021

- Run evaluations and benchmark performance on Codex program synthesis models included in "Evaluating Large Language Models Trained on Code"
- Designed, researched, and created novel datasets in Pyspark resulting in a 2x model performance improvement. These datasets replaced old benchmarks and became the standard training sets used across the entire team
- Improved throughput of model inference by 66x and scaled to hundreds of GPUs using MPI, internal tools, and distributed Pytorch training
- Developed demos, novel visualizations, and internally showcased emergent model behavior using web technologies and the Python data stack
- Led research project from ideation to completion, presenting findings to CTO and chief scientist

Google Brain - Research Science Intern
May - Nov 2020

- Contributed key mathematical and algorithmic insights into a new self-supervised pretraining method that leverages recent advances in differentiable sorting for representation learning
- Defined new state of the art performance for unsupervised audio and vision based tasks
- Streamlined and standardized several scattered experiments across notebooks and code bases. This drastically increased team productivity and we were able to launch multiple large scale experiments daily
- Designed and developed two novel metrics to measure experimental success now used by the team to communicate our findings
- Orchestrated foundational experiments across thousands of GPUs and decreased data loading time from 5 minutes to 300ms
- Explored mathematical relationship between entanglement and optimal transport distance, presented work to team
- Explored the literature and successfully reproduced results from the field which increased our ability to iterate and improve upon existing research
- Published findings in IEEE Journal for signal processing

Lyft, Level 5, Autonomous Vehicles - Software Engineering Intern
June - Aug 2019

- Developed A/B testing platform in high performant C++ to compare prediction models locally and in the cloud greatly increasing my team's development velocity
- Identified predictive features and developed real-time feature extraction system for use in machine learning pipeline
- Explored statistical and neural models for dynamical vehicle motion prediction leading to a 22.5% performance improvement
- Lead 3 engineers in exploratory 20% project for semantic code search
- Presented research to members of my team, explaining relevant topics and mathematics to apply to our technology stack

Qualtrics - Software Engineering Intern
May - Aug 2018

- Achieved ∼96% accuracy with a .005% false positive rate, matching state of the art on phishing detection by researching and implementing system using sophisticated NLP feature engineering and machine learning
- Increased speed of system 3x resulting in a 63% reduction in hardware costs while handling 3 million daily requests by engineering asynchronous API using parallel processing and high performance computing techniques
- Identified, explored, and implemented state of the art emerging topic tracking system which allowed my team to reach their stretch goals for the quarter and led to a **patent**
- Built question similarity tool using sentence embeddings after collecting and curating a dataset of ∼130,000 questions. Improved f1 score from .3 to ∼.7 built using both structured and unstructured datasets
- The final estimated impact of my internship is $300k - 500k in yearly savings

Amazon Alexa Prize: Team Eve - Machine Learning Research Engineer
Jan - Apr 2018

- Member of team Eve for the Alexa prize challenge. One of eight teams selected out of hundreds to research and build a social chatbot system to hold arbitrary conversation for 20 minutes on any topic
- Designed and built an offensive speech filtering system using probabilistic methods, which performed ∼3% better than current industry standards
- Researched and designed a complex sentiment analysis tool that classified sentences as having complex sentiment used for noteworthy knowledge retrieval

EDUCATION

M.S. Computer Science   2020
Brigham Young University   *Provo, UT*

B.S. Applied and Computational Mathematics   2018
Brigham Young University   *Provo, UT*

OTHER EXPERIENCE

Author - Everyday Data Science 2021: Best selling book about using Data Science in daily life, 1500+ copies sold

1st **place BYU AI Club Hackathon 2020**: Built a computer vision controlled robotic hand

2nd **place BYU ACM Hackathon 2019**: Built a computer vision pong game that is controlled with hand detection

2nd **place BI Wolff Hackathon 2018**: Built prescriptive ML solution to predict individual risk of becoming homeless

1st **place BYU ACM Hackathon 2017**: Created Auto Dino program to perfectly play the chrome dino no wifi game

1st **place BYU ACM Hackathon 2016**: Created *Mathify* app using polynomial interpolation to display text as math

1st **place BYU ACM Summer Coding Competition 2018, 2019**

2nd **place Global Legal Hackathon Utah 2018**: Made a chrome extension using NLP to summarize terms and conditions which I turned into a product, grew to 2000 active users, and sold

Python 3.8 Open Source: Fix small doc bug in cpython pull #11683

pyprobml Open Source: A primary contributor for Machine Learning a Probabilistic Perspective v2 Python code with Dr Kevin Murphy

Data Science Blog: 300+ monthly readers. Data science problems solved with esoteric programming languages

Ranked 8th in world: Tetris in spring of 2011