William Wong

C: +1 408 628 8804   Cupertino, CA 95014   w07wong@gmail.com   williamwwong.com   github.com/w07wong

Skills

Languages & Libraries: Python, Java, PyTorch, TensorFlow, JAX, JavaScript, NumPy, pandas, SQL, Node.js, React, HTML, Golang.

Tools: Git, AWS, MongoDB, Google Cloud, Heroku, wandb, Vim, Surge AI, Scale AI, bash, Excel, Flask, Postman, Docker.

Experience

2.2023 – Present; 5.2022 – 8.2022:   *Google DeepMind, Mountain View, CA*
Research Engineer
- Core contributor of Gemini large language model. Researched and productionized post-training, data, and evaluation techniques.
- Led the training, evaluation, and release of the Gemini Pro API in Google Cloud. Coordinated 40+ team members.
- Headed a quarter-long effort to create a human data collection workflow and processing pipeline of SFT and RLHF data.
- Developed RLHF techniques to improve language model robustness as well as training cost and speed by 16x.
- Creating agents with Gemini and releasing through Google Cloud.

- Productionized a video captioning model in 1 month for Google Cloud's Vertex AI platform.
- Owned the serving of a Flamingo-style multimodal model for multiple projects including Google's Lookout app.
- Researched hierarchical reinforcement learning (RL) and multi-agent RL to optimize Google's industrial air-cooling systems.
- Improved Google's cooling systems energy efficiency by 51% with novel reinforcement learning agents built using TensorFlow.

8.2021 – 12.2022:   *Machine Learning Department, Carnegie Mellon University*
Graduate Researcher advised by Professor Zack Lipton
- Investigated risk sensitive reinforcement learning and supervised learning; improved accuracy of classification models by 200%.
- Developed new optimization methods to improve the efficiency of lifelong continual learning models.

5.2021 – 7.2021; 6.2020 – 8.2020:   *Amazon Lab126, Sunnyvale, CA*
Software Development Engineer Intern
- Researched and developed intelligent human-robot interaction capabilities for the first Amazon Astro home robot.
- Trained neural networks in PyTorch, used kernel density estimation, and Bayesian statistics to build tailored user experiences.
- Improved customer interactivity by 12x through new autonomous features that respond to real time user actions and feedback.
- Sped up computation of the robot's behavior and motion planning components by 600%.

10.2018 – 6.2021:   *Berkeley Artificial Intelligence Research Lab, UC Berkeley*
Machine Learning Researcher advised by Professor Ken Goldberg
- Discovered and applied autonomous agricultural robot policies using imitation learning and deep reinforcement learning.
- Created CNN architectures to train deep neural network policies that decreased water by 93% and increased crop yield by 25%.
- Experimented with GANs and attention-based Mask R-CNN for plant key-point identification for domain adaptation.

5.2019 – 8.2019:   *Google, Mountain View, CA*
Software Engineering Intern
- Deployed an end-to-end solution that collects, stores and surfaces server test metrics and metadata.
- Created a Flume Java pipeline that queries, cleans and aggregates 1 to 3 million rows of server test data daily.
- Designed a Spanner database that stores and queries data with 100% accuracy and is robust to pipeline failures.

5.2018 – 2.2022:   *Vectorspace AI, San Francisco, CA*
Staff Engineer
- Utilized SVM, LDA, K-Means and t-SNE for machine learning modeling to create cryptocurrency ETFs based on hidden relationships.
- Created REST APIs for large scale data mining, a live NLP build log, and the company website using JavaScript and Node.js.
- Developed Ethereum smart contracts for over 350 million cryptocurrency tokens including the VXV token.

Education

8.2021 – 12.2022:   *Carnegie Mellon University*   M.S. in Machine Learning
Relevant courses: Deep Reinforcement Learning, Computer Vision, NLP, Convex Optimization, Probabilistic Graphical Models, Statistics

8.2017 – 5.2021:   *University of California, Berkeley*   B.A. Computer Science

Publications

- *Optimizing Industrial HVAC Systems with Hierarchical Reinforcement Learning*. **NeurIPS 2022 Reinforcement Learning for Real Life.**
- *RiskyZoo: A Library for Risk-Sensitive Supervised Learning*. **ICML 2022 Responsible Decision Making in Dynamic Environments**.
- *Learning Seed Placement and Automation Policies for Polyculture Farming with Companion Plants*. **ICRA 2021**.
- *Simulating Polyculture Farming to Tune Automation Policies for Plant Diversity and Irrigation.* **T-ASE 2022, CASE 2020**. **Best Student Paper**.

10 years of work experience since 2014. For information on earlier experience, please visit my personal website williamwwong.com.