Raymond Yoo

Amazon Machine Learning Engineer II

Personal Site + Portfolio: raymondyoo.com

About Me

Technical Skills: LLMs, RAG, Agents, Spark, SQL, Big Data, ML Ops, AWS, ML, Systems, OpenAI, DRL
Languages: Python, Java, Javascript, Haskell, C, Swift
Interests: Music, AI, Math, Mental Health, Philosophy, Classics, Deep Reinforcement Learning

Work Experience

Amazon   NYC (Mar 2022 - Current)
- Machine Learning Engineer
    - Architected and maintained a high-performance ML filtering system processing 100K+ TPS with sub-25ms latency for Amazon's sponsored products
    - Designed and implemented end-to-end ML pipelines using AWS StepFunctions, EMR, and CDK for model training and deployment
    - Led development of a multimodal recommendation engine for Amazon mobile app's "more like this" feature, leveraging computer vision models for visually similar products
    - Enhanced ad relevance by implementing semantic matching using kNN vector search on ad embeddings while maintaining strict latency requirements
    - Collaborated with data scientists to rapidly productionize ML models for A/B tests, enabling data-driven decision making that increased ad relevance by 15%
    - Orchestrated data pipeline optimization using Spark, reducing processing time by 40% while maintaining data quality for critical ML features
- Software Engineer II
    - Engineered scalable data pipelines for ML recommendation services using Spark on EMR, processing terabytes of user interaction data daily
    - Productionizing ML ETLs that applied scientists on the team create using spark SQL, EMR, Redshift, lambda, CDK, and SageMaker.
    - Working on a 100+ API service with a ReactJS frontend that saves account managers thousands of hours a year in creating managed service advertising campaigns.
- Software Engineer I
    - full stack web dev: React, Redux, SQL + NoSQL, AWS, Redshift, EMR, data pipelines, Java backend
    - CI/CD using internal tools, git, REST, Agile

Amazon   NYC (Summer 2021)
- Software Engineer Intern
    - Full stack web development + data engineering for decorating data to present trailing quarters of advertiser performance on internal website

Research Experience - Rutgers University

Machine Learning Lab
Advisor: Sungjin Ahn   Piscataway, NJ (Dec 2020-Dec 2021)
- Undergraduate Research Assistant - Deep Reinforcement Learning (DRL)
    - Implemented OpenAI's Dreamer model from scratch and presented my learnings to the lab
    - Used Dreamer as a baseline model to solve Atari games

Computational Brain Lab
Advisor: Konstantinos Michmizos   Piscataway, NJ (Jan 2020-Dec 2020)
- Undergraduate Research Assistant - ML, Deep Learning, Robotics, Deep Reinforcement Learning (DRL)
    - Developed several state of the art DRL algorithms with spiking actor networks using PyTorch
    - Wrote baselines for measuring the performance of neuromorphic versions of DRL algorithms
    - Worked on debugging deep learning models in PyTorch in a linux environment
    - Accepted publication at CoRL20 + TMLR22 (see below for link to published paper).

Education and Honors

Rutgers University New Brunswick   New Brunswick, NJ (2018-2021)
- GPA: 3.8
- **Publications**: https://arxiv.org/pdf/2010.09635.pdf + https://openreview.net/pdf?id=ZPBJPGX3Bz