

Cyrus Hadavi

Machine Learning Engineer | Ranking & ML Systems

cyrus@hadavi.net | kuros.io | github.com/cyrushadavi1 | linkedin.com/in/cyrus-h-8563b435 | New York, NY

Summary

Staff-level ML engineer with 13+ years building ranking, recommendation, and ML systems at Meta, Google, and Microsoft. Returned to an IC role at Meta in May 2026 to focus on model scaling, transfer learning, and GPU performance.

Experience

Meta | New York, NY

Machine Learning Engineer, Ads Ranking | May 2026 - Present

- Optimize features and representations for large-scale ads-ranking models through signal selection, automated ablations, feature interactions, and embedding improvements
- Develop model-scaling and transfer-learning strategies for advanced ranking architectures, including DHEN and Wukong, adapting learned representations across models, tasks, and traffic surfaces
- Profile and optimize CUDA and Triton kernels for training and inference, improving compute efficiency, memory utilization, and end-to-end ranking performance
- Scale distributed training and inference for large ranking models, balancing model quality, experiment throughput, memory efficiency, and serving latency
- Build agent-native ML workflows that automate feature optimization, architecture tuning, and hypothesis testing across a production fleet of 500+ models

Engineering Manager, Machine Learning | Sep 2024 - May 2026

- Led and technically directed 15 engineers delivering ML systems that rank billions of ads, with a focus on model architecture, feature optimization, and reliable productionization
- Mentored engineers from new grad through staff level and aligned research, infrastructure, and product partners around high-impact model investments

HBO Max | New York, NY

Engineering Manager, Machine Learning Platform | Mar 2022 - Sep 2024

- Architected an end-to-end ML platform on AWS spanning feature storage, orchestration, model registry, deployment, and monitoring using S3, EKS, RDS, Snowflake, and Databricks
- Built a self-service productionization framework that accelerated iteration across deep learning, embeddings, recommendations, churn prediction, and generative AI
- Owned platform architecture and cross-team adoption while leading a distributed team of 4 engineers

Google | Seattle, WA

Senior Software Engineer, Machine Learning | Sep 2020 - Mar 2022

- Built evaluation metrics and training-data tools for complex language tasks, improving model performance by 38%
- Fine-tuned deep learning language models for semantic understanding, named entity recognition, and relation extraction in Google Shopping

Dollar Shave Club | Los Angeles, CA

Lead Machine Learning Engineer | Jan 2019 - Sep 2020

- Built collaborative-filtering and Bayesian-ranking recommendation systems that generated \$6M in incremental revenue
- Built user and product embeddings plus an ensemble churn model that reduced customer attrition by 13%
- Designed an ML feature store that shortened production API delivery from months to weeks

Facebook (Meta) | Seattle, WA

Data Scientist | May 2017 - Nov 2018

- Applied machine learning and analytics in adversarial integrity domains, including hate speech, bullying and harassment, and hate groups

Hulu | Santa Monica, CA

Data Scientist | Oct 2015 - Feb 2017

- Built graph-based ML for content discovery and cold-start onboarding, increasing retention by 8%

Microsoft Azure | Redmond, WA

Data Scientist | Apr 2013 - Feb 2015

- Developed recommendation systems using collaborative filtering and Bayesian ranking

Skills

Core ML: Ranking, Recommendation Systems, Deep Learning, NLP, Reinforcement Learning, Transfer Learning, Representation Learning

Optimization: Feature Engineering, Feature Selection, Model Architecture, Hyperparameter Optimization, Model Scaling, Training Dynamics

Performance: Distributed Training, Low-Latency Inference, High-Throughput Inference, GPU Kernel Optimization, Profiling

Applied AI: Research Tooling, Model Evaluation, Agentic ML Workflows, Generative AI, Contrastive Learning, Embeddings, AutoML, Experimentation

Programming: Python, C++, CUDA, Triton, SQL

Infrastructure: Distributed Systems, Feature Stores, Model Monitoring, AWS, Kubernetes, Databricks, Spark, Snowflake

Education

University of Maryland | BS, Mathematics/Biology | College Park, MD | Aug 2007 - May 2012