

JONATHAN JIN

me@jonathanjin.in
linkedin.com/in/jinnovation
github.com/jinnovation
huggingface.co/jinnovation

SELECT WORK EXPERIENCE

Hinge

Staff Machine Learning Engineer, AI Platform

New York, NY, Apr 2024 – Current

Leading the platformization of Hinge's disparate AI/ML efforts.

- Re-architected [Top Photo](#) scoring pipeline to enable sub-hour reactive photo score updates and highly performant (hours) backfilling of photo scores for new model versions, reducing time-to-production for new photo-scoring models from months to days;
- Defined the technical direction of Hinge's model serving strategy. Conducted in-depth and methodical evaluation of two solutions (Ray Serve and [Databricks Model Serving](#))

Spotify

Senior Machine Learning Engineer

New York, NY, Mar 2021 – Mar 2024

Senior member of the ML Platform organization; contributed to and maintained Hendrix, Spotify's centralized machine learning platform. Primarily focused on managed multi-tenant compute infrastructure for Kubeflow and Ray in GCP. Later shifted focus to AI/ML governance and artifact/experiment management solutions.

- Evaluated Ray for adoption into the Hendrix platform in collaboration with the maintainer team at Anyscale; informed the development of Hendrix's initial centralized multi-tenant Ray infrastructure based on KubeRay;
- Spearheaded development of Backstage-based model governance solution in collaboration with non-technical stakeholders (Legal and Trust & Safety) in preparation for the EU Digital Services Act;
- Brought Hendrix's flagship Kubeflow-based orchestration product to general availability; defined multi-quarter engineering roadmap in collaboration with product and engineering stakeholders spanning infrastructural (SLOs, observability) and user-facing (documentation, API cleanliness) needs.

NVIDIA

Senior Systems Software Engineer, AI Infrastructure

New York, NY, Dec 2019 – Jan 2021

Worked on AI infrastructure for autonomous vehicle research and development.

- Implemented prototype for horizontally-scalable model-parallel training for multi-task learning using parameter server approach based on Ray and Horovod;
- Implement custom Kubernetes scheduler for preemptible workloads and resource reservations in an on-prem GPU cluster.

Twitter

Machine Learning Software Engineer

New York, NY, Aug 2018 – Dec 2019

Member of Cortex, Twitter's central ML platform organization. Worked on: workflow orchestration; experiment management/iteration; and overall ML engineering productivity.

- Enabled distributed training of TensorFlow models in Apache Mesos from an Airflow pipeline via Deepbird, Twitter's TensorFlow-based model training/evaluating/serving framework

EDUCATION

University of Chicago B.S. Computer Science, B.A Economics

2011 – 2015

SPEAKING

Ray Summit, "[Accelerating Time-to-Production for ML at Hinge](#)"

Sep 2024

TWIMLcon AI Platforms, "[How Spotify is Navigating an Evolving ML Landscape with Hendrix Platform](#)"

Oct 2022

MLconf, "[Empowering Traceable and Auditable ML in Production at Spotify with Hendrix](#)"

Mar 2022

KubeCon + CloudNativeCon, "[Scaling Kubeflow for Multi-tenancy at Spotify](#)"

Oct 2021

SKILLS

- **Machine Learning and Data:** [Ray](#), [PyTorch](#), [Apache Spark](#), [Apache Flink](#), [MLflow](#), [Apache Airflow](#), [Kubeflow](#), [TensorFlow](#)
- **Infrastructure and Distributed Systems:** Kubernetes, Terraform, Prometheus, gRPC, Docker
- **Enterprise Platforms:** Amazon Web Services (AWS), Google Cloud (GCP), [Databricks](#)
- **Programming Languages:** Python, Go, C++, Java

Last Updated: 2024-08-30. Also available at: <https://jonathanjin.in>