

# Jonathan Jin

Machine Learning Infrastructure Engineer

[jonathanjin.in](https://jonathanjin.in) [me@jonathanjin.in](mailto:me@jonathanjin.in) [linkedin.com/in/jinnovation](https://linkedin.com/in/jinnovation) [github.com/jinnovation](https://github.com/jinnovation)

## SELECT EXPERIENCE

Present 03/2021	<b>Spotify</b> SENIOR MACHINE LEARNING ENGINEER <span>New York</span> <i>Member of ML Platform. Working on <a href="#">Hendrix</a>, Spotify's centralized ML platform; multi-cluster, ML-centric cloud infrastructure with Kubernetes, KubeFlow, and Ray; and ML engineering SDKs for feature engineering, data processing, and model training, analysis, and deployment.</i> <ul style="list-style-type: none"><li>➤ Led multi-quarter promotion of ML Platform's flagship pipeline orchestration product to general availability; collaborated with product and engineering stakeholders across peer teams within ML Platform as well as "power user" teams – other applied ML teams at Spotify – to define engineering roadmap spanning three quarters and provide technical leadership/direction on prioritization, implementation, etc.; contributed to an increase in ML Platform adoption of 50% and increase in user satisfaction of 10%.</li><li>➤ Providing ongoing mentorship and support to three junior IC team members, i.e. half of the team's IC makeup; holding bi-weekly one-on-ones with each to: provide technical and career mentorship; help identify potential future ownership areas and unblock existing ones; and address ongoing concerns.</li><li>➤ Implemented formalized SLO tracking for multi-cluster, pipeline execution infrastructure; used Terraform to canonicalize SLO definitions for monitoring and violation reporting in Google Cloud Monitoring; implemented custom Kubernetes listener to implement nuanced and domain-specific SLIs.</li><li>➤ Bootstrapping early-stage development of centralized, managed Ray infrastructure based on KubeRay for high-performance ML prototyping and experimentation; collaborating with product and engineering leadership to concretize our platform's long-term Ray strategy;</li><li>➤ Driving platform unification initiative centered around ontologically-sound declarative configuration systems, based around Kubernetes Operators, for compute configuration, deployment of scheduled training pipelines, model serving, etc.;</li></ul> <div><a href="#">Ray</a> <a href="#">Kubernetes</a> <a href="#">GCP</a> <a href="#">Terraform</a> <a href="#">Helm</a> <a href="#">TensorFlow</a> <a href="#">TFX</a> <a href="#">KubeFlow</a> <a href="#">Prometheus</a> <a href="#">gRPC</a></div>
01/2021 12/2019	<b>NVIDIA</b> SENIOR SYSTEMS SOFTWARE ENGINEER, AI INFRASTRUCTURE <span>New York</span> <i>Member of AI Infrastructure. Contributor to <a href="#">MagLev</a>, NVIDIA's AI infrastructure for autonomous vehicle development. Also contributed to <a href="#">Modulus</a>, the deep learning SDK for autonomous vehicle R&amp;D.</i> <ul style="list-style-type: none"><li>➤ Initiated development of solution for "hybrid data/model parallelism" using a Ray-based parameter server design and Horovod to enable horizontally-scalable multi-task training;</li><li>➤ Co-delivered a Kubernetes-based scheduling mechanism to enable priority access to cluster resources for select use cases, e.g. prep for upcoming external demos, via virtual "resource shares";</li></ul> <div><a href="#">Ray</a> <a href="#">Horovod</a> <a href="#">TensorFlow</a> <a href="#">Kubernetes</a> <a href="#">Helm</a> <a href="#">gRPC</a> <a href="#">Bazel</a> <a href="#">SwiftStack</a></div>
12/2019 08/2018	<b>Twitter</b> MACHINE LEARNING SOFTWARE ENGINEER <span>New York</span> <i>Member of <a href="#">Cortex</a>, Twitter's central ML platform organization. Worked on: workflow orchestration; experiment management/iteration; and overall ML engineering productivity.</i> <ul style="list-style-type: none"><li>➤ Spearheaded initial integration of <a href="#">TensorFlow Extended (TFX)</a> with legacy <a href="#">Airflow</a>-based orchestration platform to increase agility of workflow development, iterative execution/experimentation, etc.</li><li>➤ Enabled distributed training of TensorFlow models in Apache Mesos from an Airflow pipeline via <a href="#">Deepbird</a>, Twitter's TensorFlow-based model training/evaluating/serving framework</li></ul> <div><a href="#">Apache Airflow</a> <a href="#">Apache Aurora</a> <a href="#">TensorFlow</a></div>

## SPEAKING

2022	<b>TWIMLcon AI Platforms 2022</b> , "How Spotify is Navigating an Evolving ML Landscape with Hendrix Platform"
2022	<b>MLconf</b> , "Empowering Traceable and Auditable ML in Production at Spotify with Hendrix"
2021	<b>KubeCon + CloudNativeCon</b> , "Scaling KubeFlow for Multi-tenancy at Spotify"

## SKILLS

Programming Languages	Python, Go, Bash, C++, Java
Machine Learning	Ray, KubeFlow, TensorFlow, TensorFlow Extended (TFX),
Distributed Systems	Kubernetes, gRPC, Docker
Infrastructure Tooling	Bazel, Prometheus, Grafana, M3, Cassandra, Apache Airflow
Cloud Infrastructure	Google Cloud Platform (GCP), Terraform

## EDUCATION

2015	<b>University of Chicago</b> , B.S. Computer Science, B.A. Economics
------	--