

Jonathan Jin

Machine Learning Infrastructure Engineer

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

SELECT EXPERIENCE

Present 03/2021	Spotify SENIOR MACHINE LEARNING ENGINEER New York <i>Member of ML Platform. Working on Hendrix, Spotify's centralized ML platform. My focuses revolve around: ML governance; cloud-native infra; and SDK development for MLOps.</i> <ul style="list-style-type: none">➤ Tech-leading a redesign of Spotify's home-grown ML experiment tracking solution. Steered initial vendor evaluation, coordinating vendor communications and technical collaboration. Post-eval, led team of four engineers to deliver an MVP based on MLflow in half a quarter. Responsible for defining the strategy and roadmap for experiment-tracking within Hendrix;➤ Spearheaded the development of the Hendrix Registry as part of company-wide AI/ML governance initiative; collaborated with stakeholders in Legal, Trust & Safety, etc. to deliver an AI system and model card solution based on Backstage and Kubernetes-based declarative infrastructure in under a quarter, resulting in over 100 models registered by 40 teams company-wide in less than a quarter;➤ Led multi-quarter promotion of ML Platform's flagship Kubeflow-based 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 multi-quarter engineering roadmap; provide technical and project leadership/direction. Contributed to an increase in ML Platform adoption of 50% and increase in user satisfaction of 10%.➤ 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.➤ 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; <div>Ray Kubernetes Go GCP Backstage Terraform Helm TensorFlow TFX Kubeflow Prometheus gRPC</div>
01/2021 12/2019	NVIDIA SENIOR SYSTEMS SOFTWARE ENGINEER, AI INFRASTRUCTURE New York <i>Member of AI Infrastructure. Contributor to MagLev, NVIDIA's AI infrastructure for autonomous vehicle development. Also contributed to Modulus, the deep learning SDK for autonomous vehicle R&D.</i> <ul style="list-style-type: none">➤ 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;➤ 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"; <div>Ray Horovod TensorFlow Kubernetes Helm gRPC Bazel SwiftStack</div>
12/2019 08/2018	Twitter MACHINE LEARNING SOFTWARE ENGINEER New York <i>Member of Cortex, 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">➤ Spearheaded initial integration of TensorFlow Extended (TFX) with legacy Airflow-based orchestration platform to increase agility of workflow development, iterative execution/experimentation, etc.➤ Enabled distributed training of TensorFlow models in Apache Mesos from an Airflow pipeline via Deeppred, Twitter's TensorFlow-based model training/evaluating/serving framework <div>Apache Airflow Apache Aurora TensorFlow</div>

SPEAKING

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

SKILLS

Programming Languages	Python, Go, C++, Java
Machine Learning and MLOps	Ray, Kubeflow, Apache Airflow, TensorFlow, TensorFlow Extended (TFX)
Infrastructure and Distributed Systems	Google Cloud (GCP), Backstage, Kubernetes, Terraform, Prometheus, gRPC, Docker

EDUCATION

2015	University of Chicago , B.S. Computer Science, B.A. Economics
------	----------------------------------------------------------------------