

Naomi Tanaka Senior ML Engineer

San Mateo, CA · naomi.tanaka@gmail.com · +1 650-555-0164 · linkedin.com/in/naomitanaka

PROFILE SUMMARY

- Senior ML Engineer with 7 years building recommendation and ranking systems at consumer gaming and AR scale, specializing in candidate generation, two-tower retrieval, and online experimentation.
- Hands-on across primary language (Python), training framework (PyTorch), distributed compute (Ray and Spark), vector search (FAISS and ScaNN), ML platform (Vertex AI and MLflow), and feature store (Feast), with strong fundamentals in two-tower and DLRM architectures.
- Deep expertise in candidate generation, online A/B test design, and training-serving skew prevention, using methodologies such as CUPED variance reduction and holdout audits to ship models whose offline gains survive the online test.
- Owns 3 production models end to end with Product, Backend, and Data Platform teams, from training pipeline and offline eval to online rollout, monitoring, and the on-call rotation for the homepage candidate generator.
- Emerging tech lead who mentors 3 mid-level engineers and runs the weekly experiment-review forum, with 4 RFCs adopted by the recommendations platform team, including the feature-store rollout.

TECHNICAL SKILLS

Languages:	Python, SQL, Go (services), Bash, C++ (basics)
Deep Learning:	PyTorch, JAX, Hugging Face Transformers, two-tower and DLRM models, sequence models
Retrieval & Ranking:	FAISS, ScaNN, candidate generation, hybrid retrieval, learning to rank, ranking pipelines
Distributed & Data:	Ray, Spark, Dataflow, Beam, BigQuery, Parquet, Feast feature store
ML Platform:	Vertex AI, Kubeflow Pipelines, MLflow, GitHub Actions, model registry, batch and online inference
Experimentation:	online A/B test design, CUPED, sequential testing, multi-armed bandits, holdout audits
Observability:	Prometheus, Grafana, model-drift monitoring, training-serving skew alerts, gRPC serving

EDUCATION

University of California, Los Angeles	B.S. in Computer Science	Los Angeles, CA · Sep 2014 - Jun 2018
---------------------------------------	--------------------------	---------------------------------------

WORK EXPERIENCE

Roblox	Senior ML Engineer	San Mateo, CA · Aug 2022 - Present
• Own the homepage candidate-generation model, a two-tower retrieval system over 40M+ experiences for 70M+ daily active users, with full responsibility for training, evaluation, rollout, and on-call. • Rewrote retrieval from a legacy matrix-factorization baseline to a two-tower model in PyTorch on Ray, lifting session relevance by 32% and homepage CTR by 11% in a 4-week A/B test. • Built a FAISS and ScaNN hybrid retrieval layer behind a gRPC service, serving 18k QPS at p99 under 35 ms with index refreshes every 30 minutes. • Cut training-serving skew incidents from 8 per quarter to 1 by adding training-time feature snapshots in Feast and a feature-consistency check in CI. • Design and ship 10 online experiments a quarter with Product, including 3 holdout audits, using CUPED to cut the time to a significant read by roughly a third.		
Niantic	ML Engineer	San Francisco, CA · Jul 2018 - Jul 2022
• Shipped 5 ranking models for the Pokémon GO sponsored-location surface, owning training pipelines on Kubeflow Pipelines with MLflow tracking, and 22 A/B tests behind a 14% revenue lift. • Owned the player-segmentation feature pipeline on Spark and BigQuery, processing 4B+ events a day for 12 downstream consumers with freshness alerts in Grafana. • Migrated 3 legacy TensorFlow models to PyTorch, cutting training time by 40% and unblocking the team's move to Ray distributed training.		