

Rohan Venkatesh

Staff Machine Learning Engineer | Senior Machine Learning Engineer | Software Engineer, Machine Learning

rohan.venkatesh@email.example • (415) 555-0172

PROFESSIONAL SUMMARY

Staff machine learning engineer with 12 years shipping ranking, speech and language model systems. Built a company-wide platform for fine-tuning, evaluating and serving large language models, and leads a group of ML and infrastructure engineers that product teams depend on.

PROFESSIONAL EXPERIENCE

Staff Machine Learning Engineer

Mar 2022 — Present

Bayline Support Software · San Francisco, CA

- Lead nine engineers building the internal LLM platform: LoRA fine-tuning pipelines, a shared evaluation harness and vLLM serving used by five product teams.
- Designed the retrieval layer behind the customer support assistant, combining hybrid search, reranking and citation checks so agents can verify every suggested answer.
- Created a release gate that blocks any model or prompt change failing regression suites on accuracy, refusal behavior and leakage of customer data.
- Lowered GPU serving spend by about \$40,000 a month through quantization, request batching and moving low-traffic adapters onto shared base models.

Senior Machine Learning Engineer

Jan 2018 — Feb 2022

Tidewater Audio Labs · Oakland, CA

- Trained and deployed on-device speech recognition in PyTorch, distilling a server model small enough to run offline on mid-range phones.
- Built distributed training on a Ray cluster with spot GPU instances and checkpoint recovery, shrinking full retrains from nine days to two.
- Mentored four engineers and set review standards for training code, including seeded runs, pinned dataset versions and reproducible configs.

Software Engineer, Machine Learning

Jul 2014 — Dec 2017

Marin Commerce Group · San Mateo, CA

- Shipped the first learned ranking model for product search as a C++ scoring library called directly from the Java search service.
- Built the impression and click logging pipeline that became the training data source for every later ranking model at the company.

EDUCATION

MS, Computer Science — Pacific Crest University

2014

BTech, Electronics Engineering — Deccan Plateau Institute of Engineering

2012

SKILLS

Technical: LLM fine-tuning (LoRA, DPO), Retrieval-augmented generation, Model evaluation, PyTorch, Distributed training (FSDP), vLLM, Ray, Kubernetes, GPU capacity planning, Python, C++, Technical leadership