

Venkatesh Arunachalam

[github/venkhatesh](https://github.com/venkhatesh) | [linkedin/venkhatesh](https://www.linkedin.com/in/venkhatesh) | venkhatarun@gmail.com | 1-812-778-5848

ACHIEVEMENTS

- Winner — NVIDIA–Atos–CDAC National AI Hackathon (1st place out of 500+ teams)
Built a Mask R-CNN system for automated rotten produce detection (>**95% accuracy**), recognized by NVIDIA & Atos for innovation.
- 2nd Place — HERE Maps Hackathon (out of 150 teams)
- Received the **Fynd Star Award (Quarterly Excellence)** for technical impact on data infrastructure
- SIES GST 2020 Technical Excellence of the Year Award (Batch 2016-2020)

EXPERIENCE

Machine Learning Engineer

June 2024 - present

Fidari Care

- Designed and deployed scalable backend APIs for retrieval, recommendations, and clinical search (p95 < 150ms, 1k QPS from 12k+ users), with automated Swagger/OpenAPI docs to streamline integration.
- Built real-time ingestion pipelines (Spark, Airflow, REST APIs) for EHR-FHIR and clinical data across oncology clinics, delivering high-throughput, reliable data flows for downstream search and analytics.
- Developed hybrid retrieval & ranking engine combining dense vector search with Elasticsearch BM25 re-ranking, improving relevance (**NDCG@10 +22%**).
- Optimized distributed LLM serving performance (FAISS, Redis, C++ optimized FAISS bindings, ONNX Runtime) with semantic caching, prompt deduplication, and token pruning, cutting latency and token cost by 40% while ensuring scalability across multi-GPU via LRU/TTL eviction policies.
- Currently developing an embedding-based recommendation engine for oncology marketplace services, leveraging vector similarity search to deliver personalized clinician & service recommendations, improving engagement and match relevance.

Research Software Engineer

January 2023 - June 2024

Indiana University Bloomington

- Led backend architecture and deployment of a multi-university mental health data platform enabled secure, scalable mobile/web data collection for 6,000+ users with role-based access and event tracking.
- Designed and owned a high-throughput streaming ETL pipeline using Kafka and Apache Flink, processed 245GB/day of behavioral sensor data with windowed aggregations, retries, and checkpointing for fault tolerance.
- Reduced inference latency of LLM-based virtual therapist and risk prediction services by profiling CUDA/PyTorch workloads and eliminating GPU kernel bottlenecks, increasing proactive outreach coverage by 30%.

Machine Learning Engineer Intern

June 2023 - December 2023

Fourkites Inc

- Optimized a distributed, high-throughput purchase order data pipeline (Java, Kafka, PostgreSQL) across 4+ microservices, improving data quality and latency for analytics and ML feature pipelines.
- Designed and deployed a distributed ETL sync service (Elasticsearch, Kafka) to improve feature freshness and consistency for downstream ML training and real-time inference, reducing CPU load by 60% and infra cost by 10%
- Built observability and monitoring dashboards (Prometheus, Grafana) to track data drift, pipeline health, and feature freshness, enabling faster root-cause analysis, reducing MTTR, and increasing uptime of ML data services.

Software Engineer

August 2020 - August 2022

Fynd Shopsense Retail Technologies

- Engineered high-performance data and API infrastructure (Java, Spring Boot, Kafka, Spark Streaming, DBT, Redshift) that powered real-time consumer analytics, improved API throughput and latency under peak loads, and reduced reporting latency from 30 minutes to 2 seconds, earning the Fynd Star Award for technical excellence.

PUBLICATIONS

- Data-Driven Football Scouting with Simulated Player Performance, IEEE ICMLA 2021 DOI: 10.1109/ICMLA52953.2021.00189

SKILLS SUMMARY

Languages: Python, C++, C, C#, Java, Go, Kotlin, JavaScript, TypeScript, Objective-C, Node

ML: PyTorch, TensorFlow, JAX, Hugging Face, ONNX Runtime, Spark, FAISS, Elasticsearch, Ray, Triton

Cloud & Databases: AWS, GCP, MongoDB, Redshift, PostgreSQL, MySQL

Distributed & Monitoring Systems: Kafka, Flink, Airflow, Kubernetes, gRPC, gdb, Perf, DVC

EDUCATION

M.S. Data Science, Indiana University Bloomington

Aug 2022 – Apr 2024

B.E. Computer Engineering, Mumbai University

Aug 2016 – May 2020