

Venkatesh P

Chennai, India · +91-91766-36234 · venkateshpvnky9@gmail.com

Portfolio · linkedin.com/in/venkatesh-ml · github.com/venkatesh-hyper · Open to relocation

Professional Summary

AI / ML Engineer experienced in building, fine-tuning, and deploying machine learning and GenAI systems. Proficient in Python, PyTorch and TensorFlow, with hands-on experience in LLM orchestration, RAG pipelines, vector databases, and scalable backend APIs. Skilled at cloud deployment (GCP, AWS), MLOps practices, and end-to-end productization of models. Seeking ML/AI engineer roles to build production-grade systems that deliver measurable business impact.

Technical Skills

Languages: Python, SQL, JavaScript, TypeScript (basic), C++

Data: Pandas, NumPy, Polars, PySpark

Visualization: Plotly, Matplotlib, Seaborn, Power BI, Tableau

ML / DL: PyTorch, TensorFlow, scikit-learn, HuggingFace Transformers, tokenization, embeddings

LLM / GenAI: LangChain, RAG, LoRA / QLoRA fine-tuning, vector DBs (Chroma)

MLOps & DevOps: Docker, Kubernetes, MLflow, DVC, Git

Deployment: FastAPI, Flask, Streamlit

Cloud: GCP (GCS, Vertex AI), AWS (EC2, S3), Azure (basics)

Experience

AI/ML Engineer Intern · NASO Technologies

Jan 2024 – Apr 2024

- Developed GenAI-driven health recommendation modules using LLMs, LangChain, and retrieval pipelines.
- Built production-grade FastAPI services containerized with Docker, improving latency and throughput.
- Designed prompt-engineering workflows for domain-specific expert reasoning.
- Integrated LLM inference with GCP-based backend systems and logging.

Machine Learning Researcher · RUSA Project, University of Madras

Apr 2024 – Jul 2024

- Performed genetic variation analysis (GWAS) and trained ensemble models for Multiple Sclerosis risk prediction.
- Built a Streamlit dashboard for polygenic risk scoring, SNP visualization, and interpretability.
- Deployed reproducible containerized workflows on GCP.

Selected Projects

Well-AI — LLM Health Assistant

GitHub

LangChain · RAG · FastAPI · HuggingFace · Docker

- Engineered a medical RAG pipeline with structured retrieval, chunking, and dense embeddings for accurate health recommendations.
- Integrated vector search with LLM reasoning (HuggingFace/transformers) to deliver explainable outputs.
- Deployed scalable FastAPI microservices; implemented session logging via Supabase to ensure data traceability.

Ticket Analysis Bot (RAG System)

GitHub

FastAPI · Streamlit · ChromaDB · Groq LLM · NLP

- Built a production-grade RAG system to analyze 70MB+ enterprise ticket logs with semantic search over ChromaDB.
- Implemented ingestion pipeline: PDF parsing, cleaning, chunking, and embedding using Sentence-Transformers.
- Integrated Groq Llama-4 Scout for low-latency contextual QA through FastAPI; exposed REST endpoints for real-time inference.
- Developed an interactive Streamlit dashboard enabling natural-language querying of enterprise ticket datasets.

Brain Tumor Classifier

[GitHub](#)

TensorFlow · CNN · Image Processing · Docker

- Developed a CNN achieving 95% accuracy on MRI brain tumor classification (4 classes: glioma, meningioma, pituitary, no tumor).
- Implemented preprocessing pipeline with normalization, augmentation, and automated labeling workflows.
- Built a clinical inference interface with model versioning and containerized deployment using Docker.

GWAS Risk Predictor for Multiple Sclerosis

[GitHub](#)

Bioinformatics · Ensemble ML · Genetic Data Analysis

- Engineered SNP-derived genomic features and trained ensemble ML models achieving $\sim 87.6\%$ predictive accuracy.
- Computed polygenic risk scores, allele-frequency correlations, and risk contribution metrics for biological interpretability.
- Designed interactive visualizations for gene-level impact analysis to support clinical decision-making.

Education

Master of Computer Applications (MCA) · University of Madras

2023 – 2025

GPA: 7.3 / 10

B.Sc. Biotechnology · RKM Vivekananda College

2020 – 2023

GPA: 7.2 / 10

Certifications

AWS Data Engineering Bootcamp · Udemy (2024)

Machine Learning A–Z · Udemy (2024)

Deep Learning Specialization · Coursera (In progress)

Microsoft Power BI Data Analyst Associate (In progress)

Achievements

- Winner, AI Front-End Hackathon 2025 (MCC Chennai).
- Finalist, National AI Hackathon (Top 5%).
- Published technical blogs on EDA, NLP, and Genomics ML.
- Mentored 150+ students in Physics and Chemistry.
- Vice President, Rotaract Club — led 5+ AI awareness and tech adoption events.