Samanvya Tripathi

AI Engineer · Boston, MA

samanvyat.work@gmail.com github.com/sacredvoid linkedin.com/in/samanvyatripathi

AI Engineer with 5+ years building production ML systems, including edge inference on Raspberry Pi, enterprise RAG pipelines, and multi-agent platforms. Shipped computer vision models and LLM services at Honeywell and two startups. MS Computer Science, Northeastern University.

EXPERIENCE

AI Engineer

Dec 2025 - Present

Nymbl · Boston, MA

- Architecting an enterprise Model Context Protocol (MCP) tool server with dynamic discovery, validation, workflow CRUD, and monitoring for a B2B AI platform
- Built a Retrieval-Augmented Generation (RAG) knowledge base platform with Qdrant vector search, Docling document chunking, gap analysis, and a Use Case Readiness scoring engine
- Developed real-time content moderation system with social media ingestion, severity classification, and compliance-aligned detection
- Delivering production features in Python (FastAPI, Qdrant, Docling) and TypeScript (React, Next.js, PostgreSQL) with Docker and AWS deployments

Machine Learning Engineer

Jan 2025 - Dec 2025

BulkMagic · Boston, MA

- Engineered a product summarization service using Llama 3.2 on FastAPI, deployed to AWS ECS with auto-scaling
- Built containerized microservices with 99.9% API uptime in production
- Created data validation pipelines processing thousands of product records with automated quality checks

Software Engineer Co-op

Jan 2024 - Aug 2024

Honeywell · Fort Mill, SC

- Optimized object detection models to 98.6% mAP for production computer vision systems in industrial settings
- Trained transformer-based image enhancement pipeline, 32% gain on target detection metrics
- Deployed edge AI inference models using TFLite and ONNX Runtime on embedded hardware

Graduate Teaching Assistant

Sep 2023 - Dec 2023

Khoury College of Computer Sciences · Boston, MA

- Supported 200+ students in Introduction to Data Science (CS3000) through weekly office hours, assignment grading, and curriculum development

Machine Learning Engineer

Jan 2020 - Sep 2021

Myelin Foundry · Bangalore, India

- Built super-resolution deep learning models optimized for on-device mobile inference
- Shipped real-time browser-based ML inference using TensorFlow.js and WebGL for consumer products
- Deployed anomaly detection systems on Raspberry Pi for industrial IoT monitoring in production
- Prototyped FPGA-based ML inference pipelines for low-latency edge computing applications

EDUCATION

MS Computer Science

Aug 2022 - Dec 2024

Northeastern University · Boston, MA

Focus: Machine Learning, Computer Vision, Natural Language Processing (NLP)

B.Tech Computer Science

2016 - 2020

SRM University · Chennai, India

Secretary, ACM SIGAI Chapter

SKILLS

LANGUAGES	PYTHON TYPESCRIPT JAVA SQL
FRAMEWORKS	PYTORCH TENSORFLOW FASTAPI NEXT.JS REACT LANGCHAIN
ML TOOLS	HUGGINGFACE TRANSFORMERS ONNX RUNTIME TFLITE VLLM WEIGHTS & BIASES
CLOUD	AWS (ECS, LAMBDA, S3, SAGEMAKER) GOOGLE CLOUD DOCKER KUBERNETES
DOMAINS	COMPUTER VISION NATURAL LANGUAGE PROCESSING (NLP) LLMS/RAG VOICE AI EDGE AI MULTI-AGENT SYSTEMS

HACKATHONS

1st Place MIT LLM Hackathon for Chemistry & Materials 2025

Best Use of Google Cloud HackHarvard 2022

Best AI/ML Hack HackUMass 2022