AMAN KUMAR

+91 9065939752 aman.kumar.dev24@gmail.com linkedin github Portfolio

Professional Summary

Machine Learning Engineer with 2+ years building end-to-end production ML systems, agentic AI pipelines, and LLM-powered automation. Strong in Python, deep learning, and MLOps — with hands-on experience deploying reliable, scalable ML services used by engineering teams in production.

Experience

SanDisk India (via Magnit Global) **Sept 2025 – Present**
Machine Learning Engineer (Contractor) *Bengaluru, India*

- Built **RTL Check Agent** — an agentic pipeline that automatically identifies and fixes lint issues in Verilog RTL code using LLM-based fix generation, RAG over lint rule documentation, and a designer review loop before applying fixes.
- Developed **EDA Agentic System** enabling AI agents to directly operate EDA tools (Fusion Compiler, PrimeTime) as tools — agents can invoke, interact with, and extract outputs from industry-standard EDA toolchains autonomously.
- Built **timing slack prediction model** (BiLSTM + Multi-Head Attention) trained on post-route timing data to predict slack values — reducing the need for running full STA sign-off flows to get timing estimates.
- Integrated **agentic AI in observability internal tools** — enabling better analysis of project status and data at different stages of the ASIC flow.
- Built **PyQt5-based desktop interface** for the RTL Check Agent — allowing designers to review, approve, or reject LLM-generated lint fixes before they are applied to the codebase.

Hyperbots **June 2025 – Sept 2025**
Applied ML Engineer *Bengaluru, India*

- Managed **full ML lifecycle in production** — Python microservices on AWS/GCP; **>99% uptime** with Docker, Kubernetes, Jenkins CI/CD, automated rollback, and drift detection.
- Built **Kafka async event streaming pipeline** — real-time ML inference signal routing, consumer orchestration, and closed-loop accuracy improvement with horizontal scaling architecture.
- Developed **document processing ML service** (OCR + FastAPI + Redis + human-in-the-loop) — async workflow, schema-validated outputs, automated quality gates; reduced processing time by **40%**.
- Built **Streamlit and Gradio dashboards** for model monitoring, inference visualization, and stakeholder interaction with ML pipelines.

Western Digital (SanDisk India) **Sep 2024 – June 2025**
AI-ML Intern *Bengaluru, India*

- Built **automated ETL pipelines** in Python processing IR drop and PrimeTime engineering reports — transforming raw structured outputs into curated, analysis-ready datasets for downstream verification workflows.
- Applied **statistical data validation** across multi-corner timing datasets; identified data quality issues and inconsistency drivers, improving pipeline output reliability by **55%**.

Technical Skills

Programming Languages: Python, SQL, Go, Tcl
Machine Learning & Deep Learning: PyTorch, TensorFlow, scikit-learn, BiLSTM, Attention Mechanisms, Hugging Face Transformers, statsmodels
Generative AI & Agentic Systems: LangChain, LangGraph, Model Context Protocol (MCP), Retrieval-Augmented Generation (RAG), FAISS, BM25, MMR, AWS Bedrock, ChromaDB, Pinecone
Backend Development & Data Engineering: FastAPI, Flask, PostgreSQL, Apache Kafka, REST APIs, Async Python
MLOps & DevOps: MLflow, Docker, Kubernetes, Jenkins, GitHub Actions, Git, CI/CD, Linux
Cloud Platforms: AWS, GCP
Data Analysis & Visualization: EDA, Power BI, Streamlit, Matplotlib

Projects

Talk to DB — NL to SQL with Streamlit Interface — *Python, FastAPI, Streamlit, PostgreSQL, LangChain, LangGraph*

- Built NL-to-SQL service — FastAPI backend, LangGraph agentic orchestration, FAISS + MMR retrieval, Redis caching, and **Streamlit interface** for interactive data querying and model monitoring.

Macro Regime Detection — *Python, hmmlearn, SciPy, NumPy, PostgreSQL, FRED API*

- Designed **end-to-end ML pipeline** — data ingestion (FRED API), feature engineering, Gaussian HMM training (EM optimization), evaluation, and walk-forward validation (241 months, zero look-ahead bias).
- Built **monitoring and reporting layer**: Sharpe 0.937 vs SPY 0.760, Max DD –21% vs –50.8% — automated backtesting with reproducible experiment tracking demonstrating full ML lifecycle rigor.

Education

Vellore Institute Of Technology, Vellore **Sep 2020 – Jul 2025**
MTech Integrated Data Science *CGPA: 8.1*
Relevant Coursework: Machine Learning, Deep Learning, Optimization Theory, Distributed Systems, Probability & Statistical Inference

Army Public School **2020**
CBSE XII *93.8%*

Certifications

Artificial Intelligence Analyst (IBM)
Introduction to Generative AI (Google)