

# Ayush Jain

Janak Park, New Delhi - 110064, India

+91-7000344937 | [by.ayushjain@gmail.com](mailto:by.ayushjain@gmail.com) | [in byayushjain](https://www.linkedin.com/in/byayushjain) | [aywhoosh](https://github.com/aywhoosh)

## SUMMARY

Python developer focused on backend engineering and distributed systems. Experienced in building scalable APIs and managing production cloud infrastructure on AWS and Kubernetes.

## PROJECTS

### • Uttar: Distributed GenAI Training & Inference System

Current

Tools: Ray, AWS EKS, PyTorch

- Architected a distributed inference cluster on AWS EKS using Ray Serve, enabling dynamic autoscaling based on GPU queue depth to handle bursty workloads.
- Optimized Retrieval-Augmented Generation (RAG) pipelines by implementing sharded vector indices (FAISS), reducing retrieval latency across distributed S3 storage.
- Reduced cold-start latency for model serving pods by configuring KubeRay efficient resource requests.

### • IRIS: Intelligent Retinal Imaging System

Feb. 2025

Tools: Python, FastAPI, AWS

[\[Link\]](#)

- Engineered and deployed a highly available inference API on AWS EC2 using Docker and FastAPI.
- Managed the system lifecycle to ensure high uptime and low-latency performance, achieving 2-3 second classification responses.

### • FinPlan: Internal FP&A Planning Tool

Dec. 2025

Python, FastAPI, SQLite, Streamlit, JWT

[\[Link\]](#)

- Built end-to-end internal finance application with budget planning, submit/approve workflow, and variance reporting.
- Implemented role-based access (Admin, Editor, Approver, Viewer) with JWT auth for secure multi-stakeholder collaboration.
- Designed CSV ingestion pipeline with validation and idempotent upserts for reliable actuals import from external systems.
- Developed Streamlit UI with interactive data editing, workflow controls, and Actual vs Plan analytics with charts.

## EXPERIENCE

### • IndianOil Corporation Ltd.

May 2025 - Aug. 2025

Machine Learning Intern

Gurgaon, Haryana

- Fine-tuned BERT-based architectures for aspect-based sentiment analysis; engineered a custom loss weighting strategy to handle class imbalance, boosting accuracy by 18% (60% → 78%).
- Optimized the feature extraction pipeline, achieving a 60% reduction in preprocessing latency via vectorized operations in scikit-learn.
- Deployed the inference engine into a production Python environment, handling real-time requests from the Anvaya system.

## EDUCATION

### • Manipal University Jaipur

Oct. 2022 - Exp. Apr. 2026

Bachelor of Technology - Computer Science & Engineering

Jaipur, Rajasthan

- GPA: 9.08/10.0

### • Delhi Public School, Mathura Road

April 2021 - May 2022

AISSCE (Class XII) - PCM with Computer Science

New Delhi, India

- Percentage: 90.4%

## SKILLS

**Programming Languages:** Python, C++, SQL, Java

**ML & Deep Learning:** PyTorch, Hugging Face, Transformers, Ray, TensorFlow, scikit-learn, FAISS, LangChain

**Cloud & Infrastructure:** AWS (EKS, S3, EC2), Docker, Kubernetes, CI/CD, Git

## CERTIFICATIONS

• [Natural Language Processing Specialization - DeepLearning.AI](#)

2025

• [Machine Learning Specialization - Stanford University & DeepLearning.AI](#)

2025

## HONORS AND AWARDS

### • Dean's List (Semester - I, II, IV, and VI)

2022 - 2025

Manipal University Jaipur

- Awarded for academic excellence in Fall 2022, Spring 2023, Spring 2024, and Spring 2025.