

# Aayush Kumar

+91-8077709912 | ayushkumarshivaliya@gmail.com | Moradabad, Uttar Pradesh  
GitHub | LinkedIn | Portfolio

## PROFILE

---

AI/ML & LLM Engineer focused on building scalable AI systems and optimizing Large Language Model (LLM) training. Experienced in LLM pipelines, system-level optimization, and real-world deployment.

## LLM & RESEARCH EXPERIENCE

---

- Built LLM systems using LangChain, LangGraph, and RAG pipelines
- Designed prompt engineering and evaluation workflows
- Worked with LLaMA, Gemini APIs, and agent-based architectures
- Focused on memory-efficient training and scalable deployment

## PROJECTS

---

### NeuroCache: Budget-Constrained Activation Offloading *PyTorch, CUDA, Systems*

- Proposed budget-controlled activation offloading using parameter  $k$  for LLM training
- Implemented via **PyTorch saved\_tensors\_hooks** with pinned CPU memory
- Achieved **~15% GPU memory reduction** with negligible throughput change
- Identified optimal tradeoff point at  $k \approx 5$  under real CUDA execution (RTX 2050)
- **Paper:** ResearchGate Link

### AI Interviewer Platform *LangChain, FastAPI, MongoDB*

- Built full-stack LLM-based interview system with automated evaluation pipeline
- Implemented question generation, response evaluation, and structured feedback
- **Live:** Project Link

### Trading Swarm *Python, Binance API, Railway*

- Built production-ready crypto trading bot using FALCON v1 strategy (EMA, MACD, RSI, volume)
- Implemented risk controls: ATR-based SL/TP, drawdown limits, and circuit breakers
- Designed modular architecture with position manager, trade logger, and alert system
- Deployed on Railway with optimized lightweight dependencies for fast builds
- **Code:** Project Link

## TECHNICAL SKILLS

---

- **Programming:** Python, SQL
- **AI/ML:** PyTorch, TensorFlow, Scikit-learn, NLP
- **LLM:** LangChain, LangGraph, RAG, Prompt Engineering
- **Tools:** FastAPI, Docker, MongoDB, Git

## EDUCATION

---

### B.Tech in Artificial Intelligence & Machine Learning 2023 – 2027

Moradabad Institute of Technology

CGPA: 8.75

### GATE (Data Science & AI) 2026

Score: 373    AIR: 8076

## CERTIFICATIONS

---

- **NPTEL Elite + Gold (Top 1%)** – AI: Concepts and Techniques
- Oracle OCI AI Foundations Associate