Manan Jain

Sunnyvale, California
☎ +1 (716) 709 1819 ✉ mananjain301@gmail.com linkedin.com/in/manan-jain30 github.com/MananJain2002
Portfolio: mananjain.dev

Experience

Daiichi Sankyo **June 2025 – December 2025**
AI Engineer (Global Development and Information Management Intern) *New York, USA*

- **Architected a scalable Hybrid Graph RAG Engine**, orchestrating **distributed ETL workflows** to synthesize unstructured medical corpora into **interconnected knowledge graphs** for complex **multi-hop reasoning**.
- Developed a **context-aware semantic chunking algorithm** and **fine-tuned 4-bit quantized** GPT-OSS-20B models using **QLoRA**, minimizing inference latency by **60%** and optimizing GPU compute costs while maintaining high-fidelity responses.
- Designed a high-throughput **RESTful microservice architecture** and automated **ETL pipelines** to ingest and sanitize diverse medical datasets, establishing a standardized API layer that cut data processing latency by **50%**.
- Automated **CI/CD pipelines** to orchestrate **cloud-agnostic Docker deployments** on **AWS** and **Google Cloud**, integrating multi-layer **AI Guardrails** to ensure **99.9% system reliability** and zero-trust privacy compliance.

University at Buffalo **November 2024 – June 2025**
Research Assistant *Buffalo, NY*

- Implemented algorithmic optimizations in **C++/Python** targeting energy-constrained deep learning workloads, utilizing **multiprocessing** and vectorization strategies that lowered energy consumption by **15%** across large-scale distributed clusters.
- Executed comprehensive **memory profiling** workflows using **Linux performance tools** like perf and Valgrind to identify and resolve critical resource bottlenecks, successfully maximizing hardware utilization efficiency for data-intensive tasks.
- Formulated **energy-performance benchmarking frameworks** to quantify compute overhead for high-load AI applications, using analysis to identify inefficiencies and decrease computational waste by **15%** across experimental simulations.

Experian **September 2023 – August 2024**
Software Development Engineer *Hyderabad, India*

- Engineered a scalable **Java Spring Boot microservices architecture** leveraging **Apache Kafka** for **asynchronous, high-throughput event streaming**, enabling the **fault-tolerant** processing of **millions of credit records daily**.
- Spearheaded core logic for the **CPD Waterfall** decision engine, implementing **parallel processing algorithms** and optimized **batch execution strategies**, cutting report generation time by **35%** for high-volume concurrent users.
- Optimized database throughput by refining **sharding strategies** and query execution plans for massive datasets, resolving **critical lock contention and IOPS bottlenecks**, improving system response latency by **25%** under peak loads.
- Constructed robust **RESTful API aggregators** to orchestrate secure data retrieval from diverse upstream services, creating a standardized integration layer that slashed latency by **20%** to support **real-time credit decisioning**.

Experian **September 2022 – September 2023**
Software Development Engineer (Intern) *Hyderabad, India*

- Orchestrated **asynchronous batch processing** using **Quartz Scheduler**, implementing **memory-efficient execution strategies** to process high-volume records, preventing system resource overloads and optimizing computational throughput.
- Resolved critical **N+1 query performance bottlenecks** in the CS-Access application by implementing **Hibernate caching strategies**, decreasing database load and user wait times by **25%**.
- Built a robust **CI/CD integration pipeline** using **Jenkins**, **Selenium**, and **Python (Robot Framework)**, automating regression suites that cut software release cycle overhead by **80%**.

Projects

Stock Trading RL Agent | *Python, PyTorch, RL, FinBERT* | Link **May 2024 – July 2024**

- Developed multiple **reinforcement learning agents** from scratch in **PyTorch**, integrating training pipelines and a **multi-signal state space** with **technical indicators, volatility metrics, and FinBERT sentiment analysis**.
- Conducted extensive backtesting and ablation studies, demonstrating **DDPG achieved highest profitability**, with sentiment features boosting Sharpe ratios and portfolio returns while significantly outperforming the **Dow Jones Index (DJI)**.

Technical Skills

Programming Languages: Python, Java, C, C++, JavaScript, SQL, Shell Scripting
AI/ML: PyTorch, TensorFlow, Transformers, Hugging Face, RAG, LangChain, LoRA/QLoRA, CUDA, FAISS, NetworkX, NLP, GAN, VAE, MLOps
Frameworks & Tools: Spring Boot, Hibernate, Apache Kafka, Flask, Docker, React.js, Node.js, Express.js, Jenkins, Quartz, Linux, Selenium, JUnit, Git
Databases & Cloud: MySQL, MongoDB, AWS (EC2, S3), Google Cloud Platform (GCP), Microservices, JDBC

Education

State University of New York at Buffalo **August 2024 - December 2025**
Master's in Artificial Intelligence (GPA 3.8/4.0) *New York, USA*

Keshav Memorial Institute of Technology **August 2019 - June 2023**
Bachelor's in Computer Science Engineering *Hyderabad, India*