Yash **Moar**

M.S. specializing in Machine Learning Systems · Former Software Engineer at HSBC

(+1) 540-824-9514 | yashmoar11@gmail.com | yashmoar11 | yash-moar

Education

Virginia Polytechnic Institute and State University (Virginia Tech)

Blacksburg, Virginia

M.S. in Computer Engineering • GPA: 3.79/4.00 — *Aug. 2025 - May 2027*

- **Coursework:** Advanced Machine Learning, Artificial Intelligence and Engineering Applications, Database Management Systems

Vellore Institute of Technology, Vellore

Tamil Nadu, India

B.Tech. in Electronics and Communication Engineering | specialization in Biomedical Engineering — *2018 - 2022*

Work Experience

Virginia Tech

Blacksburg, Virginia

Graduate Teaching Assistant (Spring '26) | Grader (Fall '25) — *Sep. 2025 - Present*

- Engineered a **HITL** pipeline using **Google Gemini Vision API** to programmatically generate alt-text for technical diagrams, achieving >**90% accessibility** and reducing **10+ hours/week** of manual workload.
- Manage **instructional support** for **70+ students**, conducting weekly **office hours**, grading technical assessments, and collaborating with Dr. Virgilio Centeno to refine **ECE 3304 - Introduction to Power Systems and Power Electronics curriculum**.

HSBC Technology

Pune, India

Software Engineer — *Aug. 2022 - Jul. 2025*

- Developed and optimized microservices based web applications to **automate home loan document generation systems** for banking staff and customer operations across global markets.
- Architected an address component using **5+ Higher Order Components (HOCs)** in React, adaptable to 6+ regional requirements, reducing **redundant code by 20%** and cutting regional implementation time by 30%
- Increased code coverage by 20% while **resolving 900+ Sonar and Checkmarx vulnerabilities**, reducing code duplication by 35% and ensuring zero critical or major issues in production.
- Engineered CI/CD pipelines using **Jenkins and Terraform**, reducing deployment time by 40 minutes
- Implemented **Promises, Redux, and AJAX** to streamline application flow, enhance state management, and improve asynchronous data handling in scalable, high performance applications.
- Engineered centralized configuration management using **AWS S3**, enabling real time parameter adjustments and reducing **deployment rollback by 20%** by ensuring consistent environments across development, QA, and production.
- **Mentored** new team members and **redesigned their training curriculum**, resulting in a **40% reduction in onboarding time**.

Key Projects

Real-Time Full-Stack Knowledge Graph System

FastAPI · Next.js · React · Neo4j · Docker Compose · Server-Sent Events · TypeScript · Pydantic — *Aug. 2025 - Present*

- Engineered a high-concurrency FastAPI backend with Server-Sent Events to stream AI responses token-by-token, decoupling inference from the request cycle and cutting time-to-first-response from 5s to <100ms.
- Built an interactive graph visualization in Next.js using react-force-graph-2d with useMemo caching, preventing re-renders during high-frequency stream updates and rendering live retrieval paths for users.
- Containerized the full polyglot stack (Python, Node.js, Neo4j) via Docker Compose with strict network rules, volume persistence, and end-to-end type safety (Pydantic + TypeScript), achieving 99.9% deployment consistency.

Distributed Real-Time Video Processing Pipeline

Apache Kafka · Ray Serve · Docker · Prometheus · Grafana · vLLM · Python — *Nov. 2025 – Present*

- Designed a decoupled microservices pipeline using Apache Kafka and Ray Serve, separating I/O-bound video ingestion from GPU-intensive processing to achieve sub-50ms end-to-end latency.
- Deployed a high-throughput LLM inference service (vLLM) with PagedAttention, increasing processing throughput by 4x whille eliminating GPU memory fragmentation under concurrent request batches.
- Built a non-blocking drift detection service using ResNet50 embeddings, exposing live data quality metrics via Prometheus and Grafana without impacting API response latency.

Skills

Languages	Python, JavaScript, Java, C++, Typescript, SQL, LaTeX, MATLAB
AI/ML & Agent Systems	PyTorch, LangGraph, Neo4j (Graph/Vector), RAG Pipelines, OpenAI API, Hugging Face, Scikit-learn, Pandas, NumPy
Full Stack Engineering	FastAPI, Next.js, React, Node.js, Server-Sent Events (SSE), REST APIs, React Force Graph, Tailwind CSS
Cloud & Data Infra	AWS (SageMaker, Lambda, EC2), Apache Kafka, PostgreSQL, Docker, Kubernetes, Jenkins, CI/CD, Git, Linux
Core Competencies	Data Structures & Algorithms, Object-Oriented Design (OOD), Database Design, Unit Testing, Agile Methodologies

Certifications & Awards

2024	**PAT on the Back Award**, at HSBC for exceptional performance and achievements
2023	**Pioneer of the quarter Award**, at HSBC for codebase optimization and vulnerability remediation
2020	**Algorithmic Toolbox**, UC San Diego | Coursera