

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

Autonomous Graph-Grounded Agentic RAG System (“Enterprise Brain”)

INDEPENDENT DEVELOPER

Aug. 2025 - Present

- Built a **Neuro-Symbolic AI** system for financial Q&A, reducing hallucinations by **40%** by grounding **LLM** reasoning in a **Neo4j Knowledge Graph** with auditable retrieval paths.
- Engineered a self-correcting **LangGraph** pipeline with **Google Gemini 1.5 Pro** and local **Ollama (Llama 3)**, autonomously grading document relevance and rewriting queries to meet accuracy thresholds.
- Developed a real-time “Thought Process” dashboard streaming agent reasoning via **Server-Sent Events**, improving system explainability and user trust in AI-generated responses.

Real-Time Multimodal Inference Pipeline

MLOPS ENGINEER

Nov. 2025 – Present

- Designed a decoupled streaming architecture using **Apache Kafka** and **Ray Serve**, achieving sub-50ms latency by isolating I/O-bound ingestion from GPU-bound inference tasks.
- Increased Vision-Language Model throughput by 4x by deploying **vLLM**, utilizing PagedAttention to solve GPU memory fragmentation issues during concurrent request batching.
- Implemented a non-blocking monitoring service using **Alibi Detect**, running statistical MMD tests on image embeddings in background threads to detect drift without degrading API response times.

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