

Vishal Srivastava

vishalgosrivastava@gmail.com | +1 949-795-8058 | LinkedIn | GitHub | Portfolio

EDUCATION

University of California, Irvine

MS in Data Science (GPA: 3.75)

Sep 2024– Dec 2025

Irvine, United States

SRM Institute of Science & Technology

Bachelor of Technology in Computer Science & Engineering (GPA: 3.9)

July 2017–June 2021

Chennai, India

SKILLS:

Programming Languages: Python, Java, C++, Bash, Shell

Developer Productivity & DevOps: CI/CD (Jenkins, GitLab), Docker, Kubernetes, Test Automation, Runbook Automation, Observability & Monitoring Dashboards (CloudWatch, custom anomaly detection), Git, Linux

AI & LLMs: LLMs, Multi-Agent Systems, AI Agents, LangChain, LangGraph, RAG, MCP, Prompt Engineering, PEFT/LoRA, HuggingFace, Vector Databases, NLP, Generative AI

ML Frameworks: PyTorch, TensorFlow, Scikit-Learn, MLflow, Weights & Biases

Backend & APIs: FastAPI, Flask, Django, REST, gRPC, GraphQL

Cloud & Data: AWS (Glue, Lambda, Step Functions, SQS, SNS, Kinesis, EventBridge, S3, DynamoDB, Bedrock, ECS, CloudFormation), PostgreSQL, MongoDB, Redis, Elasticsearch, Kafka, PySpark, Hadoop, ETL

EXPERIENCE

BARCLAYS | Software Developer

Mar 2022 – June 2024

- **Developed tool integration & validation services** spanning 40,000+ attributes with rule-based validation, schema enforcement, and exception routing via dead-letter queues, achieving 99.8% compliance while honoring RBAC and security boundaries.
- **Automated the release pipeline**, built CI/CD pipelines (Jenkins, GitLab) with automated testing gates and Runbook Automation for proactive incident response, shortening development cycles and enabling safe, repeatable deployments.
- **Built observability & monitoring dashboards** using CloudWatch, SNS, and custom anomaly detection that reduced system downtime by 30% and enabled data-driven operational decisions.
- **Architected an on-prem to AWS microservices** migration, designing fault-tolerant multi-step orchestration (AWS Glue, Step Functions, Lambda, SQS) that improved system uptime to 99.9%.
- **Designed an event-driven notification service** for 250,000+ customers (Kinesis, EventBridge, SQS, Lambda), handling 50M+ events/day at sub-second latency.

FORD MOTOR COMPANY | Analyst

Oct 2021 – Mar 2022

- **Built a reusable post-deployment health-check & test-automation framework** with automated smoke tests, integration validation, and rollback triggers, reducing production incidents by 25% and saving 7–8 person-days/month through automated diagnostics reporting.
- **Developed backend automation tooling** in Python and REST APIs that eliminated 90% of manual effort in operational workflows and accelerated bug resolution across high-reliability production deployments.

University of California, Irvine | Graduate Student Researcher

June 2025 – Sept 2025

UCI Environmental Health Lab | Advised by Prof. Jun Wu

- Improved object detection accuracy by 23% by implementing a customized CenterNet-ResNet50 architecture integrated with **CNN-based Super-Resolution (ESPCN)**, identifying trucks and containers in low-resolution satellite imagery of Southern California with 94% precision and 89% recall.
- Enhanced model interpretability and deployment efficiency by integrating **Residual Deblurring modules** and implementing automated data preprocessing pipelines using PyTorch, reducing inference time by 35% while maintaining model accuracy for environmental health research applications.

PROJECTS

FlashDB | Consistent Hashing, Synchronization, Multithreading, Replication

[GitHub](#)

- Implemented an LSM-tree storage engine with an in-memory MemTable and on-disk SSTables to improve write throughput and support efficient lookups.
- Implemented a background compaction process that merges SSTables in a merge-sort style and removes tombstones, running in parallel while the database continues to serve reads and writes.
- Designed a consistent hashing scheme with virtual nodes to distribute data across nodes and reduce reshuffling when the cluster scales.

Reddit Moderation Framework | Kafka, Flask, AI Agents, NLP, LoRA Finetuning

[GitHub](#)

- Architected a **Multi-Agent NLP System** utilizing **Mistral-7B** (Triage Agent) to normalize noisy text (slang/acronym expansion) and **Llama-3** (Repair Agent) to diagnose model failures via K-Means semantic clustering, enabling an Adapt Agent to automate **PEFT/LoRA fine-tuning** on subreddit-specific data.
- Engineered a **scalable real-time pipeline** featuring a **Flask interception proxy** that injects toxicity overlays directly into the DOM, supported by a **Kafka streaming architecture** for high-throughput data ingestion and a Streamlit dashboard for live monitoring and feedback collection.