DIVITH SELVAM

+91 9361230063 | divithselvam23@gmail.com | divithselvam.in | github.com/Divith123 | linkedin.com/in/divith-s

SUMMARY

2026 AI & Data Science graduate with software engineering experience across backend and distributed systems. Builds and evaluates machine learning and LLM systems in Python with PyTorch and Transformers, spanning model post-training, tool-use evaluation, inference, and agent workflows. Works with vLLM, Kubernetes, PostgreSQL, Redis, and OpenTelemetry to run reliable training, evaluation, and serving workflows.

SKILLS

Languages	Python, Go, SQL, TypeScript
Machine Learning	PyTorch
Deep Learning & LLMs	Transformers, PEFT, LoRA, QLoRA, TRL, vLLM
Evaluation & Agents	LLM Evaluation, Synthetic Data, Tool Calling, MCP
MLOps & Serving	FastAPI, Model Serving, Docker, Kubernetes, GitHub Actions, CI/CD
Data & Observability	PostgreSQL, Redis, OpenTelemetry, Prometheus, Grafana

PROJECTS

Tool-Use Model Post-Training & Evaluation Lab — *Mar 2025–Jun 2025*

Python · PyTorch · Transformers · PEFT · TRL

- Built deterministic tool-use environments and synthetic training trajectories covering multi-step tasks, retries, dependencies, malformed arguments, and incorrect tool selection.
- Fine-tuned models with QLoRA and compared model behavior across schema validity, tool selection, argument accuracy, task completion, held-out tool combinations, and ablations.

Distributed LLM Serving Control Plane — *Aug 2025–Dec 2025*

Go · Python · vLLM · Kubernetes

- Architected a multi-model vLLM control plane that routes streaming requests using tenant quotas, replica load, and KV-cache state, with Redis and PostgreSQL for coordination.
- Added bounded queues, backpressure, circuit breakers, and canary rollouts, then traced latency, cache behavior, and failover through OpenTelemetry.

Durable Agent Runtime for Computer Use & MCP — *Dec 2025–May 2026*

Python · Go · MCP · Playwright

- Engineered a computer-use runtime with event-sourced application state, checkpoints, retries, and crash recovery across long-running tool workflows.
- Added scoped credentials, permission checks, human approvals, task state, and isolated execution for tool actions that must remain safe across retries and recovery.

AI-Assisted Incident Triage Platform — *Jun 2025–Aug 2025*

Python · Go · Kubernetes · OpenTelemetry

- Instrumented a Kubernetes microservice environment with metrics, logs, traces, deployment events, service dependencies, and controlled fault injection.
- Built an AI-assisted triage pipeline that correlates telemetry, recent deployments, incident history, and service context to rank likely root causes and assess remediation decisions.

EXPERIENCE

CHN Technologies, *Software Engineer Intern* | Coimbatore, India — *Dec 2025–Present*

- Refactored Django and NestJS service boundaries and REST API flows across shared application workflows.
- Redesigned PostgreSQL schemas and composite indexes after reviewing query patterns and affected workloads.
- Automated Docker build and release workflows through CI/CD.

Wise Work, *Software Engineer Intern* | Bengaluru, India — *Sep 2024–Oct 2025*

- Engineered reusable Blazor/.NET components for shared application workflows.
- Profiled SQL joins, indexes, and response payloads to identify query and execution bottlenecks.
- Shaped authenticated API and data workflows, contributing to API contracts and service refactoring.

BEYOND ENGINEERING

Independent Design Work — *2026*

- Published two themes through Xiaomi's designer ecosystem, reaching **10M+ views and 2.3K followers** while refining interface and UX details.

Technical Mentorship — *Aug 2025*

- Led practical sessions for interns on Git/GitHub, APIs, authentication, deployment, documentation, and development workflows.

EDUCATION

SNS College of Engineering, *B.Tech — AI & Data Science* | Coimbatore, India — May 2026