Thanay Sisir Anabathula

anabathula.thanaysisir@gmail.com
github.com/thanay-sisir | LinkedIn | Portfolio
AI Engineering | Agent Systems and Sandboxing | Available full-time June 2027

TECHNICAL SKILLS

Programming languages: Python, C, C++
AI agents and evaluation: Agentic AI, tool calling, LangGraph, LangChain, LLM evaluation, Harbor, LangSmith observability
LLM training and inference: PyTorch, LoRA/PEFT, supervised fine-tuning (SFT), synthetic data generation, vLLM model serving
Backend and infrastructure: API development (FastAPI, Pydantic), Azure, Kubernetes, AI sandboxing (Docker, Apptainer), Slurm, SQLite
Engineering: Git, GitHub Actions CI, latency and throughput benchmarking, fault recovery

PROJECTS

SentinelSRE AI Incident Response Agent | GitHub Oct 2026

- Built a LangGraph agent to diagnose and remediate failures on a multi-service FastAPI demo platform; resolved an injected outage in approximately 90 seconds with 21 typed tools and end-to-end LangSmith traces.
- Controlled tool execution with 15 read-only tools and 6 mutating tools gated by deterministic policy checks and single-use HMAC tokens; bounded runs with model-turn and tool-call budgets, timeouts, retries, and output truncation.
- Built a Harbor evaluation harness with independent outcome verification and anti-cheating checks; earned 1.0 reward in one cloud sandbox trial, exposed 2 latent failure modes, and maintained 83 passing tests in Linux/Windows GitHub Actions CI.

Execution Verified Synthetic Data for Coding LLMs | GitHub Oct 2026

- Built a teacher-LLM pipeline that retained 1,894 Triton GPU kernels after AST screening and execution in an Apptainer GPU sandbox on A100s against PyTorch references, producing verified data for model fine-tuning.
- Fine-tuned Qwen3.8-27B with LoRA/PEFT on one A100 via Slurm; increased held-out kernel pass rate from 44.8% to 98.2% (1,178/1,200 tasks) across the evaluated operation families.
- Served base and LoRA models with vLLM and benchmarked inference, measuring 10.3-second p50 latency and 101.9 tokens/second in the tested configuration.

Orbital Federated Fine Tuning | GitHub Nov 2026

- Built a five-client federated LoRA system for Qwen3.8-27B on A100s under simulated orbital contacts and outages; a contact-window admission policy accepted 14 local training steps versus 8 for a fixed policy and reduced modeled traffic from 147.1 to 87.6 MiB in a selected outage scenario.
- Implemented SHA-256 chunk verification, idempotent receipts, and atomic SQLite checkpoints; recovered from 2 injected process exits with identical final adapters and validation loss. Evaluated scheduling across 216 paired CPU simulation cases with no reduction in accepted steps.

EXPERIENCE

Research Assistant | UVA Darden School of Business May 2026 - Present

- Built data pipelines that structured liability and joint-venture disclosures from SEC 10-K/10-Q filings for 250+ business development companies, delivering research-ready data for faculty, deans, and a Federal Reserve economist.
- Implemented fuzzy name matching of hedge fund managers across 7 TB of voter registration data to support cross-domain research.

Graduate Student Worker | UVA Engineering Link Lab Feb - May 2026

- Engineered battery-capacity features from EV fleet telematics to classify degradation tiers and validate partial-charge-based estimation as a key predictor for Prof. B. Brian Park's transportation research.

EDUCATION

University of Virginia | Master of Engineering in Systems Engineering
Jan 2026 - Expected May 2027 | GPA: 3.925/4.0

JNTU Hyderabad, India | B.Tech. in Computer Science, Minor in AI/ML | 2021 - 2025