Thanay Sisir Anabathula

GitHub | X | Portfolio | LinkedIn | anabathula.thanaysisir@gmail.com

EXPERIENCE

Research Assistant May 2026 - Present
UVA Darden School of Business | Full-Time Summer Intern until Aug 21st

- Built data engineering pipelines that turn complex, unstructured data into research-ready data for Darden faculty, Deans and a Federal Reserve economist, from extracting and structuring liability and joint-venture disclosures for 250+ BDCs out of SEC 10-K/10-Q filings to fuzzy name-matching hedge fund managers across 7 TB of voter registration data, supporting financial and cross-domain research.

Graduate Student Worker (Transportation Research Project—Prof. B. Brian Park) Feb 2026 - May 8 2026
UVA Engineering Link Lab Member

- Analyzed large-scale EV fleet telematics to engineer battery capacity features to classify vehicle degradation tiers and validate partial-charge-based estimation as the major predictor.

PROJECTS

Orbital Federated Fine-Tuning GitHub Nov 2026

- Built a five-client federated LoRA fine-tuning system for Qwen3.8-27B on A100s to study training over simulated orbital contacts and outages; designed a contact-window admission policy that delivered 14 accepted local training steps versus 8 for a fixed policy in a selected outage scenario, while reducing modeled traffic from 147.1 to 87.6 MiB.
- Engineered SHA-256-verified chunk transfers, idempotent receipts and atomic SQLite checkpoints to preserve training progress across failures; recovered from two injected process exits with identical final adapters and validation loss, and evaluated scheduling across 216 paired CPU simulation cases with no reduction in accepted training steps.

SentinelSRE - AI Agent Github Oct 2026

- Built an autonomous incident-response agent that investigates alerts and fixes failures on a multi-service FastAPI demo platform, giving the LLM 21 typed LangChain tools through LangGraph: 15 read-only tools to diagnose freely, and 6 mutating tools it can use only after a deterministic policy engine approves the action and issues a single-use HMAC approval token; resolved an injected outage in about 90s, with every step traced in LangSmith.
- Kept that autonomy predictable in cost and latency with hard per-run budgets (15 model turns, 35 tool calls), 60s LLM timeouts with retries, a 600s incident timeout and 50 KB tool-output truncation, then built a benchmark harness that logs wall-clock time, model turns and tool calls per trial (about 16s scripted, about 90s live).
- Proved the agent truly fixes incidents with a Harbor evaluation harness where an independent verifier grades real outcomes and catches and penalizes cheating; the agent scored a perfect 1.0 reward in a LangSmith cloud sandbox trial, the harness surfaced 2 latent failure modes that local testing missed, and 83 passing tests run in GitHub Actions CI on Linux and Windows.

Synthetic Data Generator for Coding LLM Agent Github Oct 2026

- Built an execution-verified synthetic data pipeline so the training data is proven correct: Grok 4.7 teacher-LLM workers write Triton GPU kernels, an AST screen checks each one, and an Apptainer GPU sandbox on A100s runs it against a PyTorch reference, keeping only the 1,894 kernels that match; served the base and LoRA Qwen3.8-27B models with vLLM to confirm the fine-tuned model is practical to deploy (p50 latency 10.3s, 101.9 tokens/s).
- Used that verified data to teach Qwen3.8-27B to write kernels for operation families: LoRA/PEFT fine-tuning on a 1 A100 via Slurm raised held-out pass rate from 44.8% to 98.2% (1,178 of 1,200 tasks), while ablations with a 0.5B LoRA model (12.3%) and a from-scratch 29.5M Transformer (0%) showed the gain comes from pairing verified data with a strong pretrained base.

SKILLS

- **AI/ML Core:** Model Training (LoRA/PEFT), Supervised Fine-Tuning (SFT), Post-Training Data Pipelines, Synthetic Data Generation, GPU Training, AI Agentic Workflows, Tool Calling, Evals
- **Inference & Infra:** LLM Inference (vLLM), Latency/Throughput Benchmarking, Docker / Apptainer (GPU sandboxing)
- **Libraries & Tools:** Python, PyTorch, LangChain, LangGraph, LangSmith (LLM observability), Harbor, FastAPI / Pydantic
- **Engineering & AI Coding:** Git / GitHub Actions CI, Azure, Cursor, Claude Code, Prime Agent

EDUCATION

Master of Engineering, Systems Engineering - University of Virginia | CGPA: 3.925/4.0 Jan 2026 – Exp. May 2027
Expected graduation: May 2027 | Available to start full-time: June 2027
B.Tech. in CS with minor in AI/ML - JNTU Hyderabad, India 2021 – 2025