DHARSHAN ATHIAMAN

+91 6374153073 | dharshanathi21@gmail.com | idharshan.com | linkedin.com/in/idharshan | github.com/iDharshan | Google Scholar

Skills

AI, LLMs & Agents: LangGraph, LangChain, Multi-Agent & ReAct Systems, Grounded Document AI, Conversational Analytics (NL-to-SQL), Prompt & Context Engineering, LLM Evaluation (Golden Datasets, MRR/nDCG), Multi-Provider Routing (OpenAI, Gemini, Anthropic)
Retrieval, ML & Computer Vision: Hybrid RAG (Dense + BM25 Sparse, RRF), Qdrant, ChromaDB, PyTorch, Vision Transformers (ViT), LoRA Fine-Tuning, LSTMs, Time-Series Modeling, ONNX, TensorRT, Edge AI (Jetson Orin Nano), NetworkX, scikit-learn, NumPy, pandas
Backend & Distributed Systems: Python, FastAPI, Pydantic, Server-Sent Events (SSE), Celery, RabbitMQ, Redis, PostgreSQL, MySQL, SQLAlchemy, Transactional Outbox Pattern, Advisory Locks, `sqlglot` AST Guardrails, React, TypeScript
Infrastructure, Security & Quality: Docker, Docker Compose, Nginx, AWS (S3), Linux, Git, pytest, Vitest, CI/CD, Multi-Tenant Isolation, RBAC, SSRF/DNS-Pinning Hardening, Root-Cause Analysis (RCA)

Experience

AI Engineer May 2025 – Present
SparxIT *Noida, IN (Remote)*

- Architected the core AI stack for an enterprise **NL-to-SQL & Conversational Analytics** platform across 300+ table ERP schemas; built multi-turn **LangGraph** query, Dashboard, and Report AI assistants with versioned state/revert, multi-dialect SQL generation, and standardized **SSE** streaming.
- Replaced noisy vector-rank table selection (7–14% precision, 12–56 irrelevant tables) with a **Schema Commitment Grounding Solver** (greedy set cover over typed requirement slots + proven join graphs) and a **Join Evidence Vector (JEV)** fallback pipeline; built golden-dataset evaluation harnesses to benchmark retrieval and SQL accuracy.
- Engineered a **13-agent LangGraph Governed Knowledge Pipeline** with Human-in-the-Loop (HITL) staging/publication gates and a **Celery/MySQL transactional vector outbox** syncing per-tenant **Qdrant** hybrid search (dense + BM25 RRF), cutting client onboarding from weeks to ∼**12 minutes**, reducing context noise by **30%**, and lifting table ranking accuracy by **+20.9%** (**0.950 MRR**).
- Built the greenfield **MetaKB** schema catalog sync subsystem using **Celery**, **PostgreSQL/MySQL advisory locks**, fencing tokens, and two-phase cancellation; enforced multi-tenant query governance via programmatic **sqlglot** AST verification, RBAC scoping, and read-only constraints.
- Designed and built a contract-first **Multi-Volume Tender Intelligence (Document AI)** platform (FastAPI, PostgreSQL-leased workers, React/TypeScript, `pdf.js`, Docker) where LLMs propose candidates while deterministic verifiers enforce page-level PDF citation spans, scope-aware conflict preservation, and pure-projection DOCX/PDF exports—cutting processing from **45–58 min** to ∼**7.2 min** (435s across **27,675 blocks** at **$0.36/run**).
- Engineered multi-provider BYOK LLM routing with token/cost accounting, circuit breakers, **SSRF/DNS-pinning** security hardening, and versioned Pydantic-to-TypeScript answer contracts supporting cohort rollout and instant rollback.

AI Research Intern July 2023 – May 2025
National Institute of Technology, Tiruchirappalli *Tamil Nadu, IN*

- Fine-tuned a Vision Transformer (**ViT-Base-Patch16-224**) with **LoRA** on **102,000** industrial images across 6 scrap categories, achieving **99.56% classification accuracy**; exported via **ONNX/TensorRT** for real-time edge inference (<**50ms** latency) on **Jetson Orin Nano Super**.
- Engineered an **LSTM-based soft-sensor** modeling non-linear biochemical dynamics in aquaponics ponds, achieving a **14% relative R^2 improvement** (**0.00074 MSE**) for real-time nitrate estimation without hardware probes.
- Developed an **LSTM** dynamical model for industrial sulfur recovery sensor time-series, achieving **74.78%** R^2 and **0.0061 MSE** for automated predictive refinery process control.

Projects

SIViT: Real-Time Edge Vision Inspection Microservice | *PyTorch, LoRA, ONNX, TensorRT, FastAPI, Docker* GitHub

- Productionized an end-to-end containerized FastAPI vision inference service deploying a TensorRT-optimized LoRA-ViT model across 6 European steel scrap classes (**99.56% accuracy**, <**50ms** latency on Jetson Orin Nano Super).
- Funded by a **Rs. 40,000 IEI R&D Grant** (`UG2025005`) and awarded **Track-Wise Best Paper (AI Domain)** at IEEE CISCON 2025.

Selected Open-Source ML & LLM Systems | *Python, PyTorch, NumPy, RAG, FastAPI* github.com/iDharshan

- **Tiny Transformer from Scratch:** Implemented multi-head self-attention, positional encoding, and backpropagation from first principles in pure NumPy and ported to PyTorch for numerical parity verification.
- **Vehicle Predictive Maintenance (41 Stars) & AI Research Analyzer:** Built real-time sensor failure risk prediction pipelines and a RAG-based academic PDF analyzer surfacing key findings and research gaps.

Publications & Grants

IEEE CISCON 2025 | *Track-Wise Best Paper Award (AI Domain)* 2025

- "**SIViT: Scrap Inspection Vision Transformer Fine-Tuned with LoRA for Industrial Scrap Classification**," A. Dharshan, R. Devadoss, R. F. Pravin, S. Ravimaran, and U. Srinivasulu Reddy, *XXII Control Instrumentation Systems Conference (CISCON)*, IEEE. [DOI]

IEEE INOCON 2024 | *6 Citations* 2024

- "**LSTM-based Soft-Sensor for Estimating Nitrate Concentration in Aquaponics Pond**," A. Dharshan, P. Kumar, S. Ravimaran, and U. Srinivasulu Reddy, *3rd International Conference for Innovation in Technology (INOCON)*, IEEE. [DOI]

IEI R&D Grant-in-Aid | *Project UG2025005* 2024 – 2025

- Recipient of a **Rs. 40,000 (INR) R&D Grant** from **The Institution of Engineers (India)** for completing hardware deployment of real-time deep-learning scrap inspection on **Jetson Orin Nano Super**.

Education

Saranathan College of Engineering Tamil Nadu, IN

B.Tech in Artificial Intelligence and Data Science, CGPA: 7.83 *Aug. 2021 – May 2025*