

Sanjay Kuppusamy Saravanan

Chennai, India

+91 9176861844 · sanjaysaravanan2317@gmail.com · LinkedIn · GitHub · Portfolio

EDUCATION

B.Tech in Computer Science Engineering (Specialization in AI & ML)

Chennai, India

Vellore Institute of Technology

2020 – 2024

PROFESSIONAL EXPERIENCE

AI/ML Engineer (Founding Member)

Chennai, India

SlateMate

Jun 2025 – Present

- Directing **Agentic AI** R&D for **real-time juvenile safety** and automated content monitoring. Architecting **Graph-RAG** frameworks for high-fidelity content moderation and user well-being.
- Engineering scalable, intelligent safety guardrails to protect next-gen users in digital spaces.

Machine Learning Engineer (Contract)

Remote, Netherlands

HyER Power

Oct 2025 – Present

- Designing **Energy Management System (EMS)** components utilizing high-precision forecasting and optimization to improve grid stability and load balancing.

Machine Learning Researcher

Chennai, India

Intern, National Institute of Wind Energy (Govt. of India)

Oct 2024 – Mar 2025

- Developed a **Temporal Fusion Transformer (TFT)** for AEP forecasting on live wind farm data, achieving a **0.040 MSE** and enabling site-specific accuracy.
- Engineered a **zero-shot generalization pipeline** for cross-site deployment without retraining, significantly reducing operational deployment overhead.
- Quantified long-term forecasting uncertainty using **Monte Carlo simulations**, in collaboration with **DTU Wind Energy Systems**.

PROJECTS

Gemma Alignment Paradigms

[Project Link]

- Developed a PyTorch framework to benchmark SFT, RL (PPO/GRPO/DPO), and Staged alignment on **Gemma 3 (270M)**.
- Engineered a **Hybrid Reward Model** with semantic scoring and safety heuristics to mitigate **reward hacking** and policy divergence.
- Optimized alignment stability via a **Staged SFT → RLHF** pipeline to maximize instruction adherence and training convergence.

GraphRAG Psychotherapist Agent

[Project Link]

- Architected a **Hybrid GraphRAG** engine fusing **Weaviate** and **Neo4j** to boost reasoning accuracy by 40%.
- Engineered a real-time streaming pipeline with **Redis** caching and safety guardrails to eliminate hallucinations.
- Deployed a scalable, containerized system with persistent memory for complex, stateful interactions.

Social Media ETL Pipeline

[Project Link]

- Developed a DuckDB-backed ETL pipeline achieving sub-10ms query latency for real-time social media analytics.
- Leveraged **Playwright** and rate-limiting to automate data extraction from complex SPAs, bypassing expensive API constraints.
- Automated text transformation and **sentiment labeling**, accelerating the generation of structured, research-ready datasets.

OPEN SOURCE & COMMUNITY

Member

Remote

Cohere Labs Open Science Community

Jun 2025 – Present

- A member of the global collective of researchers advancing **LLM interpretability**.
- Completed the intensive **ML Summer School**; engaged in expert-led technical deep-dives on **agentic workflows** and LLM reliability at Cohere Connect 2025.

PUBLICATIONS

- Automated Safety Compliance Monitoring with Autonomous Rover
- Securing Patient Health Records Using Blockchain
- Block-Chain Enhanced Patient Triage System

2024

2024

2024

TECHNICAL SKILLS

- Applied Machine Learning: Predictive Modeling, Time-Series Forecasting, Statistical Inference, Scikit-learn.
- Knowledge & Data Engineering: Neo4j (Graph), Weaviate/Milvus (Vector Search), Distributed Caching (Redis), ETL Pipelines, Pandas.
- Deep Learning & Vision: PyTorch, CNNs, Computer Vision, Model Optimization, Neural Architectures, NumPy.
- Generative AI & Agentic Systems: Graph-RAG, LangGraph, LangChain, Transformers, PEFT/LoRA.
- Engineering & MLOps: AWS (Bedrock, SageMaker), Docker & Docker Compose, CI/CD (GitHub Actions), FastAPI, Git.

CERTIFICATIONS

- Machine Learning with Python: from Linear Models to Deep Learning – MITx (Dec 2025)
- Model Context Protocol: Advanced Topics – Anthropic (Sep 2025)
- OCI Generative AI Certified Professional – Oracle (Jul 2024)
- Artificial Intelligence Analyst – IBM (Jul 2023)
- ML Summer School – Cohere Labs (Jul 2025)