

# Sanjay Kuppusamy Saravanan

Chennai, India

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

## EDUCATION

<b>B.Tech in Computer Science Engineering (Specialization in AI &amp; ML)</b> <i>Vellore Institute of Technology</i>	<b>Chennai, India</b> <i>2020 – 2024</i>
---	---

## PROFESSIONAL EXPERIENCE

<b>AI/ML Engineer (Founding Member)</b> <i>SlateMate</i>	<b>Chennai, India</b> <i>Jun 2025 – Present</i>
---	--

- 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.

<b>Machine Learning Engineer (Contract)</b> <i>HyER Power</i>	<b>Remote, Netherlands</b> <i>Oct 2025 – Present</i>
--	---

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

<b>Machine Learning Researcher Intern, National Institute of Wind Energy (Govt. of India)</b>	<b>Chennai, India</b> <i>Oct 2024 – Mar 2025</i>
---	---

- 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

<b>Gemma Alignment Paradigms</b>	<a href="#">[Project Link]</a>
----------------------------------	--------------------------------

- 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.

<b>GraphRAG Psychotherapist Agent</b>	<a href="#">[Project Link]</a>
---------------------------------------	--------------------------------

- 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.

<b>Social Media ETL Pipeline</b>	<a href="#">[Project Link]</a>
----------------------------------	--------------------------------

- 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

<b>Member</b> <i>Cohere Labs Open Science Community</i>	<b>Remote</b> <i>Jun 2025 – Present</i>
--	--

- 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

• <b>Automated Safety Compliance Monitoring with Autonomous Rover</b>	2024
• <b>Securing Patient Health Records Using Blockchain</b>	2024
• <b>Block-Chain Enhanced Patient Triage System</b>	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)