Nandhakumar Saravanan

Research Engineer in AI/ML

suriyank532002@gmail.com · +91 93440 02342 · GitHub · LinkedIn · Medium · Portfolio

SUMMARY

A research engineer with over three years building AI/ML applications for clinical trials. The research is the applied kind: I start from a problem, survey what exists or has just been published, test what holds up against real data, and ship what survives. Life sciences is where I work, not the limit of what I can work on.

EXPERIENCE

Saama — Research Engineer — Sep 2024 – Present

Agentic QAD — 2026 – Present

A multi-agent pipeline turns an SME's plain-language clinical data quality check description into a verified validation script. This reduced the time to add a new scenario from days to a few hours.

LangGraph · LLM routing · Vector DB · pytest

Query Anomaly Detection - QAD — 2024 – 2025

An LLM system that detects anomalies across a patient's full trial history, explains each flag with clinical reasoning, and drafts the query sent to sites. Over five versions, it went from fine-tuned open models to a hybrid harness that reached 97.5% accuracy, 96% precision and 95% recall in production.

Open source LLMs · LLM fine tuning · vLLM · LLM quantisation · GPUs

Saama — Machine Learning Engineer, Internship — Feb 2023 – Sep 2024

Subcat Classifier Model — 2023 – 2024

An ML engine to flag discrepancies in clinical trial EDC systems, trained entirely on synthetic data: 93% precision and 97% recall in validation, then about 50% precision once it met real data. The failure analysis is what shaped the LLM-based successor that followed.

XGBoost · MLP · Synthetic Data Generation

HACKATHONS

1st place — CDISC AI Innovation Challenge 2026 — August 2026

A team-of-four entry at CDISC's open challenge, built for Saama Technologies: a platform that carries a clinical trial protocol PDF through to a Statistical Analysis Plan and the R and SAS code that executes it, with statistician sign-off at every stage.

Best Technical Implementation Award — Saama AI CATALYST 2025 — April 2025

An internal hackathon build, four people over 36 hours: a pipeline that reads an unstructured clinical trial protocol PDF and emits it as CDISC USDM, the industry's machine-readable study definition standard.

PROJECTS & OPEN SOURCE

ClinParser (PyPI) — Structure-aware protocol parsing, fully offline. No LLM in the pipeline, so it runs on documents carrying PHI.

Programmatic Prompt Optimisation with DSPy — Compiling prompts automatically with DSPy instead of hand-tuning them. A 3B model, optimised, outscored an un-optimised 8B model on the same task.

mesa/mesa-examples — Merged fix restoring an example broken under Mesa 3.x; Prototyped contributor infrastructure — isolated CI, gallery generator, CODEOWNERS — for a GSoC 2026 proposal.

SKILLS

Languages & tooling Python · SQL · pandas · NumPy · Git · pytest

LLMs & agents Fine-tuning (LoRA, QLoRA) · Quantization-aware training · vLLM · DSPy · LangGraph · Vector databases

Machine learning XGBoost · scikit-learn · Synthetic data generation

Domain Clinical trial data (EDC) · CDISC USDM

EDUCATION

B.Tech, Information Technology — Kumaraguru College of Technology · CGPA 8.7 — 2021 – 2024