

# Edric Jeffrey Sam

+917305528953 | [edricjsam@gmail.com](mailto:edricjsam@gmail.com) | [LinkedIn](#) | [Github](#) | Coimbatore, India

## SUMMARY

AI/ML Engineer focused on Large Language Models, RAG systems, and AI reliability engineering. Experienced in fine-tuning and evaluating LLMs (QLoRA, LoRA) and building production AI systems integrating FastAPI, React, Docker, and cloud infrastructure. Developed end-to-end AI platforms including hallucination detection pipelines, adaptive learning engines using reinforcement learning, and multimodal GenAI applications. Passionate about building scalable and trustworthy AI systems for real-world deployment.

## SKILLS

- **Programming:** Python, R, JavaScript
- **Machine Learning & AI:** Computer Vision, Image Processing, OpenCV, Object Detection (YOLO), Natural Language Processing (NLP), BERT, Retrieval Augmented Generation (RAG), Generative AI, Reinforcement Learning (PPO), Bayesian Knowledge Tracing, LoRA/QLoRA Fine Tuning, Model Evaluation, Hallucination Detection.
- **Databases:** MongoDB, MySQL, PostgreSQL, Neo4j, FAISS
- **Backend & API:** FastAPI, Django, Node.js, Express.js, Rest APIs, WebSockets, ngrok
- **Distributed Systems & MLOps:** Docker, Celery, Redis, Asynchronous Processing, LLM Inference Pipelines, Containerization, Model Monitoring
- **Cloud & Deployment:** Render, Vercel, Azure
- **ML Frameworks & Libraries:** Langchain, Google Gen AI SDK, scikit-learn, Tensorflow, Pytorch, Hugging Face, Gymnasium, Stable-Baselines3, Transformers
- **Version Control:** Git

## WORK EXPERIENCE

Lysa Solutions | AI Intern | Coimbatore, India

Oct 2025 - Jan 2026

- Built a production-grade AI-powered grading system combining OCR, LLM-based evaluation, and result visualization.
- Designed a structured LLM grading pipeline with schema validation, criteria-wise scoring, and concept-level feedback.
- Implemented a centralized LLM Gateway for secure access, retries, token usage tracking, and cost monitoring; and systematic evaluation using DeepEval and LangSmith; migrated to Google GenAI stack.
- Developed asynchronous AI workflows using Celery, Redis, Docker, and PostgreSQL; deployed on Microsoft Azure with DevOps collaboration.

OneYes Infotech Solutions | Data Analyst Intern | Chennai, India

May 2025 - Jun 2025

- Built a Streamlit pipeline to automate K-Means, GMM, and Agglomerative clustering, saving time on data analysis.
- Visualized model performance with Silhouette and Davies-Bouldin scores in Power BI, enabling quick comparison of clustering methods and guiding model selection
- Enabled storage of user analysis history and generation of downloadable reports through MongoDB and FastAPI, allowing users to review past results and export findings

## EDUCATION

Kumaraguru College of Liberal Arts and Science | Bharathiyar University

Jul 2023 - May 2026

Bachelor of Science, Data Science (GPA: 8.2)

- **Achievements:** Perfect Score in Inferential Statistics, Mahatma Gandhi Merit Scholarship, CSR Project Chair - Rotaract Club of KCLAS (2024 - 25), Brigade Lead of Ignite '25

## PROJECTS

LLMOps Reliability Platform for Hallucination Detection & Monitoring

March 2026

- Fine-tuned Qwen3-8B using QLoRA (4-bit quantization) on ~8K samples to develop a hallucination detection model.
- Achieved 83.6% accuracy and 0.86 macro F1-score on a 2K-sample evaluation set using structured metrics (Precision, Recall, Confusion Matrix).
- Built an end-to-end LLM reliability platform integrating Groq inference APIs, FastAPI backend, and React frontend for real-time hallucination detection and explanation.

All-in-One RAG Assessment & Adaptive Learning Engine | *Best Final Year Project Award*

Dec 2025 - Feb 2026

- Built a RAG system (Gemini + LangChain + Gen AI SDK) to auto-generate question papers and answer keys in university-template DOCX with LaTeX math — reducing creation time from ~3 hrs to ~1 min.
- Developed LLM-powered rubric grading with Gemini Vision for handwritten PDFs and an analytics dashboard (score distributions, heatmaps, rankings) — reducing evaluation from ~15 min to ~1 min per paper.
- Engineered an adaptive learning engine with RL (PPO), Bayesian Knowledge Tracing, and a Neo4j knowledge graph for prerequisite-aware remediation — 30–33% better topic selection vs. baseline.
- Migrated MongoDB to PostgreSQL 16 (asyncpg), containerized full stack with Docker Compose (4 services, health checks, persistent volumes).

- Built a multimodal GenAI career intelligence platform using Gemini 3, TTS & STT to perform profile analysis, interview simulation, and automated asset generation.
- Designed a hybrid GenAI architecture with strict JSON schema enforcement, tool calling, and Google Search grounding for real-time skill gap analysis.
- Converted raw inputs (PDFs, GitHub repositories, audio) into deployable outputs including optimized resume bullets, cover letters, and HTML portfolios.

**Event Report Automated Generator (In House & Currently Active) | KCLAS****Mar 2025 - Apr 2025**

- Reduced university event report creation time from 1 hour to 5 mins using FastAPI & DOCX.

**Real Time Sign Language Detection and Translation using YOLO****Jan 2025 - Feb 2025**

- Real-time sign detection using YOLO (93% accuracy)
- Mediapipe and LSTM (Stacked LSTM - 97% accuracy); integrated WebSocket + FastAPI.

---

**CERTIFICATIONS**

- Python for Data Science (NPTEL)
- Digital Transformation & Business Intelligence using Excel and Power BI (Amypo)
- Web Development (IBM )
- Object Oriented Programming using Python (Infosys Springboard)
- AI Powered Business Analytics (National University of Singapore – International Immersive Program)

---

**PUBLICATIONS**

- **IEEE (Scopus Indexed)**

Emerging Trends & Analytical Perspectives in the Development of New Smartphones.

*Published in 2025 3rd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*

- **IJIRT (UGC)**

All-in-One RAG Assessment Engine: Dynamic Creation, Automated Evaluation, and University-Centric Output.

*Published on January 2026 | Volume: 12 | Issue: 8 | PageNo: 8308-8314*