

SAM SANKAR P

AI/ML Engineer & Applied AI Systems Engineer — LLM Infrastructure, Model Fine-Tuning & Agentic Systems

+91 95977 45327 • p.samsankar2005@gmail.com • Virudhunagar, Tamil Nadu, India

LinkedIn: linkedin.com/in/sam-sankar-p • GitHub: github.com/SS-360 • Hugging Face: huggingface.co/SamSankar • Kaggle: kaggle.com/samsankarp

PROFESSIONAL SUMMARY

Final-year Computer Science and Engineering student (CGPA 7.8) and AI/ML engineer targeting AI Infrastructure, MLOps, and LLM Deployment roles. Hands-on experience building three production-grade open-source systems spanning agentic tool infrastructure (143 typed MCP tools), large-scale LLM evaluation (1,090,163 examples across 38 datasets), and interpretable multimodal model finetuning (Gemma 4, GRPO/SFT). Proficient in Python, PyTorch, Hugging Face Transformers, FastAPI, Docker, Kubernetes, and CUDA, with practical exposure to model quantization, LoRA/QLoRA, vLLM-based inference, and CI/CD pipelines.

CORE SKILLS

Languages: Python, JavaScript, TypeScript, Bash, HTML, CSS, SQL

ML/DL: PyTorch, TensorFlow, Transformers, Hugging Face, LLMs, VLMs, Multimodal AI, Computer Vision, NLP, OpenCV, SentenceTransformers, BERT, DeBERTa-v3, NLI

Model Training & Optimization: SFT, LoRA, QLoRA, PEFT, GRPO, PPO, RLHF, Unsloth, BitsAndBytes, 4-bit Quantization, GGUF, Model Merging, CUDA, vLLM

Agentic AI & APIs: Model Context Protocol (MCP), LangChain, LangGraph, OpenAI API, Anthropic API, Tool Orchestration, Structured Outputs, Evaluation Pipelines

MLOps & Infrastructure: FastAPI, Docker, Docker Compose, Kubernetes, MLflow, Git, GitHub Actions, Pytest, CI/CD, Linux, GPU Compute Operations, Thread-Safe Services

EXPERIENCE

AI Infrastructure & MLOps Engineer — Independent R&D — Self-Directed

Mar 2026 – Jul 2026 | Python, PyTorch, FastAPI, Docker, Kubernetes, Hugging Face, MCP

- Designed, built, and deployed three production-grade open-source AI systems end-to-end (detailed under Selected Projects), covering agentic tool infrastructure, large-scale LLM evaluation, and multimodal model fine-tuning.
- Managed the complete lifecycle for each system: architecture design, model training/fine-tuning, containerized deployment (Docker/Hugging Face Spaces), and technical documentation.

Full-Stack Web Development Intern — IT Links, Palayamkottai, Tirunelveli

Feb 2025 | 10-day internship | HTML, CSS, JavaScript, React, Node.js, PHP

- Delivered a bulk certificate-generation web application that converted structured participant data into print-ready certificate outputs, replacing a fully manual generation workflow.
- Built responsive, reusable frontend layouts using React and vanilla JavaScript; implemented server-side template population, input validation, and file-generation logic using Node.js and PHP.
- Completed the full software delivery cycle — requirements gathering, implementation, testing, debugging, and handoff — within a 10-day sprint timeline.

SELECTED PROJECTS

MaterialPilot — Local-First MCP Bridge for Material Maker

Jul 2026 | Open Source | TypeScript, Node.js, MCP, GDScript, Godot 4.7, WebSocket JSON-RPC

- Designed and shipped an MCP-compatible platform exposing 143 typed tools and 392 generated node definitions, enabling AI agents to inspect, create, validate, and export procedural material graphs end-to-end.
- Engineered authenticated loopback communication between a TypeScript MCP server and a native GDScript bridge, preserving 100% local-first operation with zero external data transmission.
- Implemented revision-safe transactional editing with expected-revision checks, snapshot-first execution, atomic patch application, and full rollback/undo-redo support to guarantee data integrity.
- Added 4-tier permission modes (read-only, assisted, workspace-autonomous, developer) and CI-tested packages with crossplatform setup documentation for Windows, Linux, and macOS.

HallucinationGuard-Env v4.2.1 — LLM Grounding & Evaluation Environment

Mar 2026 – Apr 2026 | Open Source | Python, FastAPI, PyTorch, Hugging Face, Docker, OpenEnv

- Built an OpenEnv-compatible reinforcement learning environment integrating 1,090,163 examples across 38 datasets to train and evaluate LLMs on factual grounding, citation discipline, and refusal behavior.

- Designed a 9-component reward function combining factual correctness, source grounding, citation accuracy, confidence calibration, DeBERTa-NLI consistency, ROUGE, BERTScore, and AlignScore.
- Classified 8 hallucination categories across 5 severity levels and implemented curriculum-ordered task families with adaptive difficulty and session-isolated episode state.
- Deployed a full REST + MCP JSON-RPC service (reset, step, grading, batch evaluation, leaderboard endpoints) via Docker on Hugging Face Spaces with cached datasets and reproducible run scripts.

AEGIS — Agentic Deepfake Analysis with Gemma 4

May 2026 | [Gemma 4 Good Hackathon](#) | [Gemma 4 E2B](#), [PyTorch](#), [PEFT](#), [Unsloth](#), [DeBERTa NLI](#), [GGUF](#)

- Built a multi-agent, interpretable deepfake-detection pipeline using Gemma 4 E2B across 3 prompt-switched reasoning stages (Visionary, Inquisitor, Self-Critique) followed by a DeBERTa cross-encoder NLI judge.
- Reduced unsupported verdicts through contradiction filtering and conservative evidence aggregation, producing an inspectable reasoning trail instead of a single opaque score.
- Fine-tuned and published SFT and GRPO adapters via Unsloth, releasing a merged model and a 4-bit GGUF quantized variant to Hugging Face for lightweight local inference.
- Delivered an end-to-end Kaggle inference notebook, modular Python pipeline, and full technical documentation of benchmark results and limitations for hackathon submission.

Promptless Full-Stack Generator —

2026 | [Open Source](#) | [Python](#), [Database Schema Analysis](#), [HTML](#), [CSS](#), [JavaScript](#)

- Prototyped a schema-driven application generator that analyzes database structure and automatically produces a functional initial web interface without requiring a natural-language prompt.
- Implemented automated domain inference, schema extraction, and UI-mapping logic to convert raw database metadata into usable application scaffolding.

EDUCATION

Bachelor of Engineering (B.E.), Computer Science and Engineering — AAA College of Engineering and Technology (AACET), Sivakasi, Tamil Nadu

2023 – 2027 (Expected) | CGPA: 7.8/10

Relevant Coursework: Machine Learning, Deep Learning, Computer Vision, Data Structures and Algorithms, Operating Systems, Database Management Systems, Software Engineering, Probability and Statistics

CERTIFICATIONS

- Oracle Cloud Infrastructure 2025 Certified AI Foundations Associate
- Oracle Cloud Infrastructure 2025 Certified Foundations Associate
- GPU Architectures and Programming — NPTEL
- CPU Architecture — NPTEL
- Industry 4.0 and Industrial Internet of Things — NPTEL
- Google Analytics Certification
- Connecting to MongoDB in Node.js — MongoDB University

LEADERSHIP & LANGUAGES

Leadership: Treasurer, Designer, and Event Coordinator — Student Technical Association, AACET

Languages: English, Tamil

PROJECT & PORTFOLIO LINKS

GitHub: github.com/SS-360 | MaterialPilot: github.com/SS-360/materialpilot | HallucinationGuard: github.com/SS-360/hallucination-guardenv
HallucinationGuard Live Space: huggingface.co/spaces/SamSankar/hallucination-guard-env | AEGIS: github.com/SS-360/AEGIS-AgenticDeepfake-Detection-with-Gemma-4

Hugging Face: huggingface.co/SamSankar | Kaggle: kaggle.com/samsankarp