

KUMAR SATYAM

kumarsatyam3135@gmail.com — +91 7903844254 — linkedin.com/in/newton4th — github.com/krsatyam36

CAREER OBJECTIVE

AI/ML Embedded Software Engineer specializing in the intersection of **deep learning and hardware-aware inference**. Proven experience building **end-to-end MLOps pipelines**, deploying models across **NVIDIA Jetson, Hailo-8, and AMD Versal**, and optimizing computer vision systems for real-world edge environments. Passionate about transforming heavy ML architectures into high-throughput, low-latency production deployments.

SKILLS

- **Languages:** Python, C++, Java, Bash/Shell, SQL
- **ML / AI:** PyTorch, TensorFlow, scikit-learn, Hugging Face, YOLOv8, OpenCV, TFLite
- **MLOps / Deployment:** Docker, Podman, MLflow, ClearML, Model Registry, CI/CD
- **Edge AI / Accelerators:** TensorRT, NVIDIA Jetson, Hailo-8, AMD Versal ACAP
- **Embedded Platforms:** Raspberry Pi, ESP32, STM32, Pixhawk 6C
- **Robotics / UAV:** ROS2, MAVROS, MAVLink, PX4, ArduPilot, QGroundControl
- **Web Stack:** FastAPI, React, Vite.js
- **Cloud / Systems:** Linux, AWS (EC2, S3, SageMaker), Git, Confluence
- **AI Engineering Tools:** Cursor, Claude Code, Gemini CLI, GPT Codex, Claude Design, Lovable, Ollama

EXPERIENCE

- **AI / ML Embedded Software Engineer – Little Place Labs (LPL)** **Nov 2025 – Present**
 - **Edge AI Deployment & Model Porting:** Engineered automated conversion pipelines for PyTorch, TensorFlow, and ONNX models into optimized deployment formats including TensorRT and Hailo HEF. Executed proofs-of-concept involving complex internal architectures and YOLOv8 pipelines for Hailo-8 while maximizing throughput and hardware efficiency.
 - **Architecture Diagnostics & Testing Agents:** Built a model inspection platform supporting H5, BIN, ONNX, PyTorch, and TensorFlow formats for structural analysis and conversion debugging. Developed automated layer-testing agents with layer-break heuristics to identify compatibility failures, reducing retraining cycles and shortening DS team feedback loops.
 - **Dataset Intelligence Platform (Hackathon Winner):** Architected a full-stack internal dataset platform using FastAPI, React, and Vite featuring ML-readiness scoring, automated profiling, script generation, and chat-based EDA using local SLMs. Eliminated repetitive dataset downloads and manual scripting bottlenecks, significantly improving DS productivity. Awarded **3rd Prize** in internal engineering hackathon.
 - **MLOps Infrastructure & Standardization:** Established organization-wide Docker/Podman container naming and versioning standards. Integrated MLflow, ClearML, and model registry systems for experiment tracking and artifact lineage. Authored internal technical documentation covering Hailo workflows, secure Versal VE2302 access, and SSH recovery processes.
- **AI & Computer Vision Intern – PANTECH.AI (Remote)** **Apr 2024 – May 2024**
 - Delivered **7 end-to-end ML projects** including plant disease detection with 500+ classes and speech emotion recognition.
 - Achieved **85% accuracy** in speech emotion classification.
 - Worked across dataset preparation, training, optimization, and deployment pipelines.

PROJECTS

- **Edge-AI Enabled UAS for Search and Rescue (SAR)** **Jetson Orin, Pixhawk 6C, YOLOv8, ROS2**
 - Designed and integrated a heterogeneous two-drone swarm using Jetson companion computers with Pixhawk controllers via MAVLink/UART.
 - Built custom Python GNC stack with dynamic path planning and Offboard Control.
 - Deployed YOLOv8 for real-time human detection and GPS geotagging at **30 FPS**.
 - Built resilient WiFi / 4G telemetry bridge for BVLOS-style communication.
- **Personalized AI Shadow – LLM Fine-Tuning** **Mistral-7B, QLoRA, Hugging Face**
 - Built a personalized LLM assistant mirroring tone, reasoning style, and preferences.
 - Curated **3.4M-entry dataset (1.9GB)** of technical and reasoning data.
 - Fine-tuned Mistral-7B for **137+ hours**, achieving loss **0.35**.
 - Merged and quantized model to GGUF for offline CPU deployment.
- **VisionNet – Edge AI Drone Platform** **Jetson Nano, Raspberry Pi, TensorRT**
 - Built AI drone system for crack detection and solar panel inspection.
 - Achieved **99.3% detection accuracy** with **250 ms** median latency.
 - Applied TensorRT and INT8 quantization for real-time inference.

PUBLICATIONS

- **Design and Implementation of a Heterogeneous Edge-AI Drone Swarm for Autonomous Logistics** **2025**
 - First Author paper proposing decentralized drone swarm architecture with onboard AI inference.
 - Implemented perception stack using Jetson Orin Nano Super with TensorRT optimization.
 - Introduced secure BVLOS communication using 4G-LTE telemetry and Zero Trust networking.

ACHIEVEMENTS

- **Granted Design Patent:** AI-Based Modular Crack Detection Drone System (10 Nov 2025).
- Winner: Technothon IoT 2024 — Runner-up: AI Unleashed 2024 — SIH 2024 Institutional Finalist.
- Selected to pitch at IIT Kharagpur Incubation Centre.
- 3rd Prize, Internal LPL Engineering Hackathon.

EDUCATION

- **B.Tech – Computer Science (AI/ML), Techno India University** **Kolkata, 2025**
 - **CGPA:** 8.0 / 10