

PROFESSIONAL SUMMARY

AI Engineer working across generative AI, computer vision and robotics. Built a production generative design platform end to end, from LoRA fine-tuning to web and iOS release, and deployed transformer-based drone surveillance for Indian Army defence projects. Two-plus years shipping ML systems at scale, with merged open-source contributions to Google DeepMind TORAX.

TECHNICAL SKILLS

Generative AI: Diffusion models (Stable Diffusion/SDXL, FLUX), LoRA fine-tuning, Diffusers, ComfyUI, inpainting and image-to-image, style transfer, image-to-3D

LLMs & Multimodal: Gemini, OpenAI and Claude APIs; vision-language models; prompt engineering; RAG; AI agents and tool use (MCP); LLM evaluation

ML & Deep Learning: Python, PyTorch, JAX, TensorFlow, Hugging Face Transformers, Scikit-learn, XGBoost, LightGBM; computer vision (DETR, YOLO, DeepSORT, OpenCV)

Robotics & Drones: ROS, ArduPilot, MAVLink, ground control stations, ESP32 and embedded C++, IMU sensing

MLOps & Infrastructure: Docker, Kubernetes, AWS, GCP (Vertex AI, BigQuery), MLflow, Kubeflow, Airflow; CUDA, mixed precision, quantization

Engineering: TypeScript, Node.js/Express, React, React Flow, Go, C++, Java; REST/GraphQL, WebSockets; Spark, Kafka; Git, CI/CD

PROFESSIONAL EXPERIENCE

FAM, AI Engineer

Oct 2025 – Present

- Built a node-based generative design platform end to end (React, React Flow) with 25+ composable nodes for interior and architectural visualization, released on web and iOS (Capacitor).
- Designed a Node.js/Express API unifying Gemini/Imagen, OpenAI, Claude, FLUX Pro, Stability AI and Meshy behind one interface, with request queueing, retry/backoff and SSRF protection.
- Shipped sketch-to-render, virtual staging, multi-style transfer, region-based editing, section/elevation-to-render and image-to-3D pipelines.
- Fine-tuned custom LoRA adapters on diffusion models from curated interior imagery, building the dataset pipeline (curation, auto-captioning, trigger tagging) and tuning rank and learning rate against overfitting; served them through a ComfyUI-based backend so designers could stack brand styles without retraining.

Whoopmasters India, AI & Robotics Intern (*Indian Army defence projects*)

Jul 2026 – Present

- Built an AI aerial-surveillance system on DETR (Detection Transformer) for real-time object detection, deployed and flown on operational drones.
- Engineered a custom ground control station integrating ArduPilot over MAVLink, unifying mission planning, live telemetry, vehicle control and the AI detection feed in one operator interface.
- Developing ROS-based control software for an 18-servo hexapod robot, covering gait generation, IMU attitude sensing and failsafe disarm.

Foundever, Machine Learning Engineer

Jul 2023 – Aug 2025

- Built and deployed scalable ML models for recommendation and classification, integrating them into Android apps with TensorFlow Lite and cutting on-device inference latency by 40%.
- Architected an end-to-end ML pipeline for large-scale data ingestion, preprocessing and model deployment.
- Integrated DeepSORT tracking with YOLO for real-time multi-camera person tracking; applied transfer and self-supervised learning to improve feature extraction.

PROJECTS

Deeplense Gravitational Images ([LINK](#))

Designed physics-inspired CNN architectures that embed astrophysical symmetries for gravitational-lens classification, with an evaluation framework benchmarking against ResNet and ensemble baselines.

Multi-Camera Real-Time Tracking System ([LINK](#))

Implemented YOLOv5 and DeepSORT to track subjects across multiple camera feeds, automatically closing inactive feeds; reduced processing time by 30%.

Quantitative Stock Analysis Platform ([LINK](#))

Built a Flask API serving LightGBM predictions from RSI, MACD and Bollinger Band features; optimized inference with Joblib, reducing compute load by 30%.

OPEN SOURCE CONTRIBUTIONS ([LINK](#))

Google DeepMind TORAX: Added absorption-fraction heat-source modelling and blocked double-counted radiation losses (merged).

Google Gemini SDKs: Added Application Default Credentials, streaming callbacks and custom fetch support. **Neuroinformatics**

Movement: Implemented straightness index for trajectory analysis. **LangChain:** Authored integration examples. Also contributed to Ray, MuJoCo, IBM Terratorch, OSS-Fuzz, Supabase and Elasticsearch.

EDUCATION

San Jose State University: B.Sc. in Computer Science & Applied Mathematics, CGPA 3.7/4.0

Relevant coursework: Deep Learning, NLP, Statistical ML, Algorithms, Database Systems