

Amrithesh Kakkoth

+91 88483 67638 | amrithesh.dev | amritheshkakkothh.mec@gmail.com | linkedin.com/in/amrithesh-kakkoth | github.com/Amrithesh-Kakkoth

EDUCATION

Government Model Engineering College
Bachelor of Technology in Computer Science, CGPA: 7.5

Kochi, Kerala
Sept. 2022 – May 2026

EXPERIENCE

Evals Intern Aug 2025
HUD (YC W25) Remote

- Built and created test for RL environments using python, Docker and Shell.

SDE (ML) Intern Jan 2026 – Present
Ente.io On-Site (Bangalore)

- Fine-tuned a YOLOv5-Face model with custom hyperparameters to detect pets, improving detection accuracy by **7–8%** based on IoU and confidence threshold metrics.
- Designed and implemented an end-to-end pet detection pipeline, optimizing preprocessing and inference for robust real-world performance.
- Built custom architecture for embedding-based model for pet identification to enable instance-level recognition across images.

SDE (ML) Intern May 2025 – July 2025
Mecha Systems Remote

- Architected production CLI agent translating natural language to shell commands using fine-tuned Qwen 2.5 0.6B with LoRA (Unsloth), achieving 64% execution reliability.
- Implemented 8-bit quantization reducing memory footprint by 60%; deployed via llama.cpp for efficient local inference on Linux devices. Trained on 10k+ real-world command examples.

PROJECTS

VectraDB | *Rust, Tokio, gRPC (Tonic), REST (Axum), Sled, Protocol Buffers* Solo Project

- Engineered three indexing algorithms: HNSW (sub-millisecond search), LSH (hash-based), and Product Quantization for optimized similarity search trade-offs.
- Built dual API architecture with gRPC and REST using Protocol Buffers; implemented persistent storage with Sled for crash-resistant recovery. Containerized with Docker; includes Python client with sync/async support.

CodeAnalyzer Pro | *Python, LangChain, ChromaDB, AST, Groq, Docker* Solo Project

- Engineered code-aware RAG system with AST parsing for deep structure analysis, entity extraction (functions, classes, variables), and relationship mapping across codebases.
- Integrated GitHub repo analysis, security scanning, performance bottleneck detection, and multi-language support (Python, JavaScript, TypeScript) with a natural language CLI interface.

SKILLS & INTERESTS

Languages & Systems: Rust, Python, C++, C, Bash, SQL, Linux

Infrastructure & APIs: Docker, gRPC (Tonic), REST (Axum), Protocol Buffers, Tokio, Git

Databases: PostgreSQL, Sled, ChromaDB, Vector DBs (FAISS, Chroma)

ML/AI Stack: PyTorch, Keras, TensorFlow, LangChain, Hugging Face, RAG Patterns, LLM Fine-tuning

Interests: Agentic AI, Systems Programming, Distributed Systems, RAG Architecture, LLM Inference