

Lasantha Kulasooriya

Machine Learning Engineer | AI Agents & LLM Systems | Physical AI & Edge AI

lasantha-ai.pages.dev

+94705356893 | lkulasooriya97@gmail.com | linkedin.com/in/lasantha-kulasooriya | github.com/LasaK97

SUMMARY

I translate rules a human used to apply by judgment into decision logic a system can execute correctly, on its own, at scale, whether that's insurance risk assessment, a robot's mission behavior, or an LLM agent's tool-use. **Senior Machine Learning Engineer** with production experience across AI agent orchestration, RAG systems, voice AI, and physical robotics: BehaviorTree.CPP mission logic running unattended in the field, a cloud-edge voice AI platform integrating four LLM/speech providers behind one runtime, and a self-hosted LangGraph multi-agent platform built and shipped independently. Comfortable owning a system end to end, from SME requirements through production deployment and measured evaluation.

TECHNICAL PROFICIENCY

AI Agents & Orchestration: BehaviorTree.CPP, LangGraph, LangChain, Multi-Agent System Design, Autonomous Decision Logic, Rule-Based Decision Systems

LLM & Generative AI: OpenAI, Anthropic, Mistral, Cerebras, Groq, Hugging Face Transformers, LiteLLM, RAG, RAGAS (Evaluation), Langfuse (Observability), Prompt Engineering, Vector Databases (Qdrant, FAISS, ChromaDB)

Voice & Realtime AI: Deepgram, ElevenLabs, LiveKit

NVIDIA Edge AI & GPU Computing: DeepStream, VPI, TensorRT / TensorRT-LLM, CUDA/CuPy, Jetson AGX Orin

Cloud & Backend: Microsoft Azure, Oracle Cloud, FastAPI, REST APIs, MQTT, WebSockets, PostgreSQL, MongoDB, SQL/NoSQL, Docker

Languages & DevOps: Python, C++, JavaScript/TypeScript, ROS 2, CI/CD, Git

EXPERIENCE

Senior Machine Learning Engineer

Feb 2026 – Present

Hype Insight Global

Colombo, Sri Lanka

- Designed and built a C++/ROS 2 agent orchestration layer (BehaviorTree.CPP) for the Manriix robot that translates mission-level operational rules into autonomous decision logic, coordinating safety, navigation recovery, task sessions, and device state across multiple subsystems in production.
- Designed a cloud-edge voice AI platform for Manriix integrating LiveKit, Deepgram, OpenAI, and ElevenLabs behind one agent runtime, with MQTT-based deployment sync so cloud-managed configuration pushes live to the edge without a runtime rebuild.

Data Scientist

Oct 2024 – Feb 2026

Hype Insight Global

Colombo, Sri Lanka

- Built the multi-camera perception and navigation-decision pipeline for Manriix, sustaining 19fps detection across three ZED X cameras at 20–30% GPU utilization via direct SDK integration, CUDA-accelerated clustering, and an 8-level recovery architecture.
- Built ROS 2 service integrations connecting perception, camera, and navigation subsystems into one coordinated automation workflow.

Associate Data Science and AI Engineer

Jul 2024 – Oct 2024

Janashakthi Insurance (JXG)

Colombo, Sri Lanka

- Translated insurance business rules into a machine-learning-based personalized recommendation system, working directly with SMEs to convert portfolio/policy data into decision logic.
- Built a deep-learning claim-prediction model converting policyholder onboarding data into risk-decision logic, achieving 65%+ evaluation accuracy.

Intern Data Science and AI Engineer

Jan 2024 – Jun 2024

Janashakthi Insurance (JXG)

Colombo, Sri Lanka

- Built a real-time KPI pipeline and decision-logic-driven leaderboard for a company-wide sales competition, streaming ranked results into a business-facing web app.
- Contributed to a real-time, cross-department insurance KPI dashboard unifying multiple data sources into one operational reporting system.

KEY PROJECTS

Manriix – Autonomous Event Robot

manriix.com

C++17 · ROS 2 · BehaviorTree.CPP · Voice AI · TensorRT · Jetson

Hype Insight Global

- Built the C++ ROS 2 BehaviorTree.CPP mission orchestrator coordinating safety, voice interruption, navigation recovery, and docking through one reactive behavior tree, with atomic crash-safe mission-state persistence and an automated GTest contract-test suite.
- Designed the cloud-edge voice AI platform using LiveKit, Deepgram, OpenAI, and ElevenLabs behind one agent runtime, with MQTT-based deployment sync so cloud-managed roles and configuration can be pushed live to the edge without rebuilding the runtime.
- Built the TensorRT-accelerated voice-guided photo capture system using YOLO11-pose, MediaPipe, and RealSense depth, targeting sub-15 ms pose inference within a 10 Hz vision loop and controlling Canon camera and DJI gimbal capture.

Virtual Me – Production RAG Chat Backend

github.com/LasaK97/VirtualMe

TypeScript · Cloudflare Pages Functions · Qdrant · RAGAS · Langfuse

Public

- Built and deployed a production RAG chatbot with hybrid dense and sparse retrieval, three-provider LLM failover across Cerebras, Groq, and Mistral, and semantic caching, entirely on free-tier infrastructure.
- Cut cache-hit latency by 50% from 4.36s to 2.19s and eliminated false refusals from 2/34 to 0/34 through Langfuse-traced, measurement-driven optimisation; evaluated retrieval and answer quality with RAGAS, achieving 0.95 context precision and 0.81 faithfulness.

The Small Council – Self-Hosted Multi-Agent Platform

github.com/LasaK97

Python · LangGraph · FastAPI · Next.js · PostgreSQL · Qdrant

Personal Project

- Architected a self-hosted multi-agent platform as a modular monolith so adding a new agent requires a new module rather than a re-architecture; built and shipped Grand Maester Aemon, a LangGraph-orchestrated tutor agent using FSRS spaced-repetition scheduling and rubric-graded partial-credit feedback through Telegram.
- Deployed the platform on Oracle Cloud with PostgreSQL using a per-agent schema and a Qdrant hybrid storage layer; shipped one of four planned council agents into real daily use.

EDUCATION

University of Peradeniya

Peradeniya, Sri Lanka

BSc (Hons) in Data Science, Faculty of Science — GPA: 3.65

2019 – 2024

CERTIFICATIONS

Large Language Models (Level 1, Level 2, Level 3) - H2O.ai

Issued Jun 2024

Deep Learning Specialization (5 certificates) - Coursera

Issued Aug 2025

Applied Data Science Module - WorldQuant University

Issued Jul 2024

AI/ML Engineer (Stage 2) - SLIIT

Issued Aug 2023

Oracle Cloud Infrastructure 2024 AI Foundations Associate - Oracle

Issued 2024

I hereby certify that the above information is true and correct to the best of my knowledge.

29.07.2026



Lasantha Kulasooriya