

# Rusiru Erandaka *AI & ML Engineer*

✉ contact.erandaka@gmail.com 📞 +94710438198 📍 Colombo, Sri Lanka 🌐 <https://erandaka.dev/>

🌐 <https://lk.linkedin.com/in/rerandaka> 🌐 <https://github.com/rusiru-erandaka>

## Profile

AI/ML Engineer with **hands-on experience building and deploying production-grade AI systems** across Computer Vision and LLM pipelines. Specialized in **real-time inference, edge deployment, and agentic AI systems**, with proven impact in optimizing performance under hardware constraints. Demonstrated ability to **design end-to-end AI pipelines** from data engineering to deployment backed by measurable improvements in latency, efficiency, and accuracy.

## Skills

### AI / ML Engineering

PyTorch, TensorFlow, Scikit-learn, Transformers

### LLMs Systems & Agentic AI

RAG, CRAG, LoRA / QLoRA, Prompt Optimization, LangChain, LangGraph

### Backend & Development

FastAPI, Flask, Docker, GitHub Actions, MLflow

### Databases

MongoDB, PostgreSQL, SQLite

### Computer Vision & Edge AI

YOLOv8, Vision Transformers (ViT), TensorRT, NVIDIA Jetson

### Programming Languages & Frameworks

Python, Java, JavaScript, Spring Boot Next.js, React.js

### Cloud Platforms

AWS (Lambda, EC2, S3 bucket), Vercel

## Work Experience

2025/09– 2026/04

Moratuwa, Sri Lanka.  
(Onsite)

### Intern AI/ML Research Engineer (Computer Vision)

*Arthur C. Clarke Institute for Modern Technologies*

#### Train-Elephant collision prevention system (Edge AI + Real-Time CV)

- Designed and deployed a **thermal vision-based real-time detection system** using YOLOv8 and ViT for low-visibility environments.
- Achieved **85.66% detection accuracy** under challenging conditions (night / fog / thermal noise).
- Optimized inference pipeline on **NVIDIA Jetson TX2**, reducing GPU utilization from **23% → 9% (~60% improvement)** by applying **TensorRT optimization**.
- Engineered a **real-time computer vision pipeline** for obstacle and track monitoring with a **custom TCP/IP-based gimbal control system** by reverse engineering device communication protocols.
- Delivered a **deployable edge AI system**, balancing accuracy, latency, and hardware limitations.

## Education

2022/08 – Present

### BSc. (Hons) In Electronics & Computer Science.

University of Kelaniya, Sri Lanka.

CGPA: 3.42/4.00

## Publications

ICATC 2025 (IEEE)

Erandaka, R., et al.

“Multimodal Browser-Based System for Online Child Safety”

ICATC 2025 (IEEE) | DOI: 10.1109/ICATC68823.2025.11407778

## Projects

---


2026/04 – Present

### Autonomous Synthetic Data Factory for Agentic Reward Model Training

Python | LLMs | RLHF | GitHub Actions | Hugging Face | SQLite

- Architected an end-to-end **autonomous pipeline** using **5 LLMs** to generate high-quality synthetic datasets for **RLHF and agentic model** training.
- Designed multi-stage orchestration pipeline. (Task Generation → Execution → Dual Labeling → Validation → Uploading)
- Automated daily dataset publishing to Hugging Face via GitHub Actions. (zero manual intervention)
- Developed real-world data ingestion engine (GitHub, Stack Overflow, research papers) to generate executable ReAct-style tasks across 6 domains.
- Implemented a **constitutional labeling framework** to generate automated reward signals and inter-labeler agreement metrics using LLM-as-a-judge.
- Generated structured agent traces with tool usage logs and failure classifications (hallucinations, safety violations) for reward model training.

**Dataset:** [Rusiru-erandaka/Agent\\_Supervisor\\_Training\\_dataset](#) 

**GitHub:** [rusiru-erandaka/Autonomous\\_Agentic\\_Data\\_Factory](#) 

2026/03 – 2026/04


### Automated Market Intelligence Pipeline for UpToDate Sri Lankan Food Price Gathering

(140+ Downloads from 1 month)

Python | ETL | GitHub Actions | Hugging Face | PDF Parsing

- Built an automated ETL pipeline to ingest and structure daily CBSL commodity price reports for 24 products across 5 markets using a custom PDFplumber parser.
- Engineered parser resilience across 3 years of inconsistent PDF schemas and handled missing commodities via graceful NaN degradation instead of pipeline failures
- Recovered 1,200+ days of historical data by replacing unreliable scraping with deterministic URL-based data extraction
- Designed production-ready dataset pipeline with automated publishing to Hugging Face and schema validation for data consistency

**Dataset:** [Rusiru-erandaka/UpToDate\\_Srilanka-Food-prices](#) 

**GitHub:** [rusiru-erandaka/Food\\_dataset\\_preparation\\_Pipeline](#) 

2025/01 – 2025/09

### Multimodal NSFW Content Detection System (Research Project)

Python | YOLOv8 | BigBird | PyTorch | FastAPI | AWS

- Built multimodal content detection system combining BigBird-RoBERTa (97.3% text accuracy) and YOLOv8m (0.899 mAP@50) for image classification
- Designed a hybrid deployment pipeline with on-device visual inference and low-latency text analysis via AWS-hosted FastAPI
- Engineered a custom **multithreaded web browser** architecture to enable real-time filtering with <250ms response time while reducing overall system overhead.
- Curated and annotated 27K+ multimodal dataset (text + images) and applied class balancing and augmentation techniques for improved model performance.

**Trained Model:** [Rusiru-erandaka/Child\\_safety\\_bigbird](#) 

2025/08 – 2025/10

### RAG Based Natural Language Agentic Business Chatbot

Gemini API | LangChain | FAISS | Pinecone | PyPDF

- Built an agentic RAG chatbot with contextual memory using LangChain
- Enabled multilingual queries (Sinhala & English)
- Integrated FAISS and Pinecone for efficient vector retrieval
- Implemented session-based memory for improved conversational continuity

## Certificates

---

- AWS Cloud Quest Cloud Practitioner 
- Docker Training (KodeKloud) 
- Feature Engineering Techniques Kaggle 