

Thathsarani Sandarekha

Machine Learning Engineer | Data Scientist

LinkedIn , GitHub , Medium

Email : thathsaranisandarekha@gmail.com

Mobile : +(94) 76 842 3305

TECHNICAL SKILLS

Programming Languages: Python, Java, R, SQL, JavaScript

Frameworks & Libraries: PyTorch, TensorFlow, Scikit-learn, Keras, Huggingface, OpenCV, Pandas, NumPy, Matplotlib, Seaborn

Machine Learning & AI: Deep Learning (CNN, RNN, LSTM, Transformers), NLP (BERT, LLaMA, RAG, LangChain), Computer Vision (YOLOv8, PaddleOCR, SAM, CLIP), Time Series Forecasting

MLOps & Data Engineering: PySpark, Databricks, Docker, Git, AWS, Azure, MLflow, FastAPI, Streamlit

Data Visualization & Analytics: Power BI (DAX, Power Query, custom KPIs), Excel (Power Query, PivotTables, XLOOKUP), SQL (relational queries, joins, aggregations), React.js

PROFESSIONAL EXPERIENCE

Intern Data Scientist

Colombo, Sri Lanka

OCTAVE – John Keells Group, Colombo, Sri Lanka

Jul 2023 – Jul 2024

- **Data Engineering:** Built scalable data pipelines on **Azure Databricks** using **PySpark**, enabling automated ingestion and preprocessing of large marketing datasets. Integrated with **Power BI** to deliver real-time dashboards for brand insights.
- **Natural Language Interfaces:** Developed a **RAG-based assistant** with **LangChain** and **Azure OpenAI** to allow non-technical business users to query sales and finance data using natural language, returning both textual and visual outputs.
- **Applied ML/DL Research:** Collaborated with data scientists, engineers, and finance teams at **Cinnamon Hotels** to design AI-driven solutions for budgeting, revenue optimization, and energy savings.
- **MLOps Practices:** Strengthened deployment skills through **model versioning**, **cloud training**, and **monitoring pipelines**, ensuring robust performance in production environments.
- **Data Accuracy:** Utilized **Excel (Power Query, PivotTables, XLOOKUP)** for data cleaning, reconciliation, and KPI validation prior to dashboard refreshes, improving trust in analytics across business teams.

PROJECTS

Air Quality Data Assistant – End-to-End ML Agent System

Designed and deployed an AI agent to analyze multi-room air quality sensor data. Enabled natural language queries over raw NDJSON files with dynamic chart/table generation. Built schema normalization, code execution, and visualization pipeline using FastAPI + React. Delivered a scalable system simulating real-world IoT analytics without a database.

Tech Stack: Python, FastAPI, Matplotlib, Pandas, SentenceTransformers, Groq LLaMA3, React.js **Demo Video**

MetaFaces – Research on Consistent Multi-Character Image Generation

Built a memory-guided diffusion pipeline (SDXL) with attention injection and anchor-based consistency. Integrated Grounding DINO + SAM for spatial segmentation and improved character preservation. Published at FTC 2025, Munich (Springer LNNS) with benchmark results (CLIP Score: 0.3619, DreamSim: 0.1753).

Advanced state of the art in character consistency across visual storytelling tasks.

Tech Stack: Python, Transformers, Diffusion Models, OpenCV, Grounding DINO, SAM

ResuMatch AI – Resume Parser & JD Matching Platform

Developed an LLM-based system to automatically parse resumes and score alignment against job descriptions. Implemented semantic similarity search with FAISS and SentenceTransformers to return ranked results. Provided recruiters with detailed subscores (skills, tech stack, relevance, communication) and explanations. Reduced manual screening time, improving recruiter efficiency.

Tech Stack: Python, Streamlit, Groq (LLaMA3), Sentence Transformers, FAISS, LangChain **Demo Video**

ChatGroq – Multi-Document Conversational QA System

Built a chatbot that supports Q&A across multiple document formats (PDF, CSV, Excel). Implemented semantic chunking + FAISS retrieval to deliver accurate, source-backed responses. Delivered an intuitive Streamlit interface for interactive querying and document traceability. Enabled knowledge workers to query large document sets without manual searching.

Tech Stack: Python, Streamlit, LangChain, FAISS, Hugging Face, LLaMA3, CLIP, LLaVA

ITSM SLA Compliance Dashboard – Power BI

Developed and maintained an ITSM SLA Compliance Dashboard in Power BI to monitor Incident, Problem, Change, and Request processes across 17 months. Automated KPI tracking for SLA compliance %, ticket volumes, resolution times, and violations using DAX and Power Query, reducing manual reporting effort. Enabled managers to drill down by priority, department, and region, improving visibility into SLA breaches and driving timely corrective actions.

Tech Stack: Power BI, Power Query, DAX, Excel

Additional Projects (GitHub portfolio)

- Brain Tumor Segmentation and Classification (U-Net + CNN)
- Edge AI Parking System (YOLOv8 + PaddleOCR + Flutter)
- Time Series Forecasting (XGBoost)
- Sentiment Analysis (BERT)
- Person Re-ID (YOLOv8 + Siamese Network)

EDUCATION

BSc (Hons) Artificial Intelligence and Data Science

2022 – 2025

Robert Gordon University

Scotland, UK (Delivered in Sri Lanka)

G.C.E. Advanced Level - Physical Science Stream

2020

Pushpadana Girls' College, Kandy

Sri Lanka

PUBLICATIONS

- *MetaFaces: Dynamic Memory Networks for Consistent Character Representation in Multi-Image Visual Storytelling* — Accepted at the **FTC 2025**, Munich, Germany — Springer LNNS.
- *Systematic Review of Consistent Multi-Character Image Generation* — Accepted at the **ICMVA 2025**, Melbourne, Australia — SPIE Proceedings.
- *Comparative Analysis of Smoke and Fire Detection Systems* — Accepted at the **ICIVC 2024**, Suzhou, China.