

KAVINDU LAKSHAN PERERA

AI Engineer | Machine Learning Engineer | Generative AI Developer

Colombo, Sri Lanka | +94 77 456 7890 | kavindu.perera@email.com | linkedin.com/in/kavinduperera-ai | github.com/kavinduperera

PROFESSIONAL SUMMARY

Detail-oriented and innovation-focused AI Engineer with 4+ years of experience designing, training, deploying, and maintaining machine learning, deep learning, computer vision, NLP, and generative AI solutions. Experienced in building production-ready AI applications for fintech, HR-tech, retail analytics, document automation, and customer service domains. Skilled in Python, PyTorch, TensorFlow, Scikit-learn, FastAPI, LangChain, OpenAI APIs, Azure AI, AWS, Docker, SQL, and MLOps practices. Strong ability to translate business requirements into scalable AI systems that improve efficiency, decision-making, and user experience.

CORE COMPETENCIES

AI/ML	Supervised learning, unsupervised learning, classification, regression, clustering, recommendation systems, anomaly detection, time-series forecasting
Deep Learning	CNN, RNN, LSTM, GRU, Transformers, transfer learning, model fine-tuning, object detection, image classification
Generative AI	LLM integration, RAG pipelines, embeddings, semantic search, prompt engineering, agent workflows, chatbot development
NLP	Text classification, sentiment analysis, named entity recognition, summarization, document Q&A, information extraction
Computer Vision	OCR automation, image preprocessing, object detection, OpenCV, YOLO, document image analysis
Deployment/MLOps	FastAPI, Flask, Docker, CI/CD basics, model versioning, API deployment, monitoring, cloud services

TECHNICAL SKILLS

Languages	Python, SQL, JavaScript, TypeScript, Bash
Libraries	PyTorch, TensorFlow, Keras, Scikit-learn, Pandas, NumPy, OpenCV, Hugging Face, LangChain, Matplotlib
Cloud	Microsoft Azure AI, AWS S3/Lambda/SageMaker basics, Google Cloud basics
Databases	PostgreSQL, MySQL, MongoDB, Firebase, Redis, FAISS, ChromaDB, Pinecone
Tools	Git, GitHub Actions, Docker, Postman, Jupyter, VS Code, Linux, Jira, Confluence
Testing/Quality	Unit testing, API testing, data validation, model evaluation, bias checks, error analysis

PROFESSIONAL EXPERIENCE

AI Engineer - NexaLabs Technologies (Pvt) Ltd, Colombo, Sri Lanka

January 2023 - Present

- Develop and deploy AI-powered software modules for document processing, customer support automation, predictive analytics, and business workflow optimization.
- Built Retrieval-Augmented Generation pipelines using LangChain, OpenAI APIs, vector databases, and enterprise knowledge sources to support internal search and automated Q&A.
- Designed RESTful AI model-serving APIs using FastAPI and Docker, enabling software engineering teams to integrate ML outputs into production applications.
- Improved OCR-based invoice extraction accuracy from 82% to 94% through image preprocessing, field-level validation rules, and model-assisted extraction.
- Collaborated with product managers, QA engineers, and DevOps teams to define acceptance criteria, deployment workflows, and monitoring requirements for AI features.
- Prepared technical documentation covering datasets, model assumptions, API contracts, evaluation metrics, and operational limitations.

Machine Learning Engineer - Lanka DataWorks Solutions, Colombo, Sri Lanka

August 2020 - December 2022

- Developed predictive models for customer churn, credit risk scoring, sales forecasting, and user behavior analytics using Scikit-learn and TensorFlow.
- Performed data cleaning, exploratory data analysis, feature engineering, model selection, hyperparameter tuning, and evaluation using accuracy, precision, recall, F1-score, AUC, RMSE, and MAE.

- Created automated dashboards and model output reports for business stakeholders using Python, SQL, and BI-friendly exports.
- Reduced manual report preparation time by approximately 60% by automating data preparation and model result generation workflows.
- Worked with relational databases and ETL-style scripts to prepare structured datasets for ML model training and validation.

Data Science Intern - Ceylon Analytics Hub, Colombo, Sri Lanka

January 2020 - July 2020

- Supported senior data scientists in developing machine learning prototypes for retail sales prediction and campaign segmentation.
- Prepared Python notebooks for EDA, missing value treatment, visualization, feature correlation analysis, and baseline model comparison.
- Assisted with documentation of experiment results, dataset limitations, and model performance summaries.

SELECTED PROJECTS

AI Document Extraction Platform

- Built a document intelligence solution to extract invoice numbers, supplier details, dates, line items, tax amounts, and totals from scanned PDFs and images.
- Used OCR preprocessing, validation rules, LLM-assisted extraction, and human review workflows to improve reliability for finance teams.
- Technologies: Python, OpenCV, FastAPI, PostgreSQL, Docker, Azure Blob Storage, OpenAI API.

Enterprise Chatbot with RAG

- Developed an internal AI assistant that answers employee questions from policy documents, onboarding materials, and technical guides.
- Implemented embeddings, vector search, source-aware answers, context ranking, prompt templates, and response guardrails.
- Technologies: LangChain, ChromaDB, FastAPI, React, OpenAI embeddings, Azure App Service.

Computer Vision Attendance Verification Prototype

- Created a prototype for attendance verification using face detection, image quality checks, and duplicate submission detection.
- Focused on privacy-aware processing, false-positive reduction, and audit logging for HR operations.
- Technologies: Python, OpenCV, PyTorch, PostgreSQL.

Customer Churn Prediction Model

- Designed a churn prediction model using customer activity, billing behavior, support history, and engagement metrics.
- Delivered explainable insights to help marketing teams prioritize retention campaigns.
- Technologies: Pandas, Scikit-learn, XGBoost, SHAP, SQL.

EDUCATION

BSc (Hons) in Computer Science - University of Colombo School of Computing, Sri Lanka

2016 - 2020 | Second Class Upper Division

- Final Year Research: Sinhala-English document classification and summarization using transformer-based NLP models.
- Relevant Modules: Artificial Intelligence, Machine Learning, Data Mining, Database Systems, Software Engineering, Algorithms, Statistics.

CERTIFICATIONS

- Microsoft Certified: Azure AI Engineer Associate - 2024
- Deep Learning Specialization - Coursera / DeepLearning.AI - 2023
- Machine Learning with Python - IBM Skills Network - 2022
- Docker for Developers - Udemy - 2022
- Prompt Engineering for Developers - OpenAI Learning Path - 2024

ACHIEVEMENTS

- Won 2nd place at Sri Lanka AI Innovation Challenge 2023 for an AI-based document automation prototype.
- Published a technical blog series on building RAG applications for enterprise knowledge bases.
- Mentored 5 junior developers and interns on Python, model evaluation, API integration, and AI project documentation.
- Recognized internally for delivering an AI automation module two weeks ahead of schedule.

PROFESSIONAL STRENGTHS

- Strong analytical thinking with the ability to connect AI techniques to measurable business outcomes.
- Clear communication with technical and non-technical stakeholders.
- Practical understanding of model limitations, data quality issues, privacy concerns, and production reliability.
- Comfortable working in Agile teams using Jira, sprint planning, code reviews, and cross-functional collaboration.

LANGUAGES

Sinhala	Native proficiency
English	Professional working proficiency
Tamil	Basic conversational proficiency

REFERENCES

Available upon request.