Kashmith Samaranayaka

Software Engineer · AI/ML

kashmithnisakya@gmail.com • +94 77 271 5351 • kashmith-nisakya • kashmithnisakya

Professional Summary

Software Engineer with 2.5+ years building production **AI/ML and backend systems** in Python, Jac, and Go, specializing in LLM-powered and agentic applications, MLOps pipelines, and scalable infrastructure. Shipped AI products used by real customers, including the backend of **Tobu** (**1,000+ users**) and live apps on Google Play and the App Store. **Open-source core contributor to the Jaseci stack** (**70+ merged PRs**, **950+ GitHub stars**). Co-author of **two IEEE-published papers** on deep learning. Graduate of the **University of Moratuwa**.

Technical Skills

Languages: Python, Go, TypeScript, Jac (jaclang), SQL, Bash
AI/ML & GenAI: LLMs (OpenAI GPT-4o), RAG, Agentic AI (LangGraph), Fine-tuning (Unsloth), NLP, Deep Learning, Veo 3
ML Frameworks: PyTorch, TensorFlow/Keras, scikit-learn, Hugging Face
MLOps & Infra: Docker, Kubernetes, MLflow, Apache Airflow, CI/CD (GitHub Actions), Google Cloud, AWS
Backend & Data: FastAPI, Go, Jaseci, PostgreSQL, PySpark, Kafka, Redis Streams, WebSockets, Vector DBs (Chroma)

Experience

AI/ML Engineer April 2024 - Present
Jaseci Lab, Ann Arbor, MI, USA *Remote*

- Core contributor to the open-source **Jaseci stack** (**950+ GitHub stars**) with **70+ merged PRs**, as a core developer of the **jac-scale** Kubernetes deployment and scaling features and contributor to the **jaclang** language.
- Developed and maintained backend infrastructure for Tobu, an AI-powered personal memory platform with integrated LLMs serving **1,000+ users**; took the system to production fast enough to support the company's **$1.5M raise**.
- Shipped ML features for image analysis and conversation processing that turn unstructured photos and chats into structured, searchable information, and built scalable PySpark pipelines for user analysis and feature optimization.
- Lead developer of **Flowline**, an AI-powered project management platform deployed on Kubernetes, and built the **Veo 3** image-to-video pipeline behind **TLite**, live on the App Store (see Projects).

Data Scientist April 2024 - June 2024
Utech Technologies (Pvt) Limited, Wattala, Sri Lanka *Remote*

- Designed and deployed an end-to-end energy forecasting pipeline with MLflow experiment tracking and model versioning for reproducible, production-ready forecasts.
- Automated ML workflows with Apache Airflow for data ingestion, model retraining, and deployment.
- Built production API services with FastAPI and Docker for real-time energy prediction and dashboard visualization.

Application Engineer Intern January 2023 - July 2023
Synopsys Lanka (Pvt) Ltd, Colombo, Sri Lanka *On-site*

- Developed an NLP-based search engine with LSTM models for automated email classification.
- Automated large-file compilation and notification systems using Python, Bash, and CSH scripting.
- Managed 17 processor designs through custom automation scripts for daily performance assessment and resource optimization.

Education

B.Sc. Engineering (Hons.) in Electrical Engineering Jan. 2020 - July 2024
University of Moratuwa *Moratuwa, Sri Lanka*

- **Overall GPA: 3.31 (Second Upper Class)** · Washington Accord–accredited program
- Dean's List – 2nd semester

Open-Source Contributions

Jaseci Stack: Core Contributor (jac-scale, jaclang)

- **70+ merged PRs** into the open-source Jaseci project (**950+ GitHub stars**), spanning the jac-scale and jaclang components.
- Core developer of **jac-scale**, Jac's Kubernetes deployment and scaling layer that takes identical local Jac code to production, auto-provisioning databases, Redis caching, and authentication with zero configuration.
- Contributor to **jaclang**, an AI-native language featuring graph-based Object-Spatial Programming with native Python, JavaScript, and C interoperability.

Projects

Industry & AI Projects

Flowline - AI-Powered Project Management Platform (Live)

- Lead developer of a Kubernetes-deployed platform that syncs with GitHub and keeps the team board current as issues and pull requests change.
- Built an AI assistant that answers questions about the team's work, writes daily standup notes, and turns plain-language requests into tasks; building an **AI code review agent** with cost limits and an evaluation set to measure review quality.

TLite (Tobu Lite) - AI Image-to-Video Generation (Live)

- Built the image-to-video feature, a generative media pipeline integrating Google's **Veo 3** model that lets users create videos from their photos using selectable templates.

Einstein Pro - AI Patent Generation & Document Management Platform

- Backend developer for an AI platform that generates patents and manages documents, including a RAG-based chatbot over a vector database for fast, grounded answers.

Deep Research Assistant - Agentic AI Application

- Built a production-grade agentic app for analytical question answering using LangGraph for multi-step reasoning, with web search and Chroma vector retrieval, deployed via Docker with a Streamlit interface.

Spenzo - Personal Money Manager (Live)

- Lead backend developer of the Go API behind the mobile and web apps, with an OpenAPI contract, contract tests, and CI/CD from GitHub Actions to Google Cloud Run and Cloud SQL.

Self-Adaptive NILM using Deep Learning (Research)

- Designed WaveNet/CNN models with ensemble methods for appliance-level power prediction; applied pseudo-labelling and transfer learning for self-adaptation to appliance ageing without retraining. Published at the **2024 IEEE World AI IoT Congress**.

Publications

M. S. K. Nisakya, S. M. L. Arampola, W. A. Yasodya, S. Kumarawadu, V. Logeeshan, C. Wanigasekara. **"Self-Adaptive Non-Intrusive Load Monitoring Using Deep Learning."** *2024 IEEE World AI IoT Congress*, 2024.

W. A. Yasodya, S. M. L. Arampola, **M. S. K. Nisakya**, V. Logeeshan, S. Kumarawadu, C. Wanigasekara. **"Self-Adaptive Deep Learning Framework for Non-Intrusive Load Monitoring: Addressing Aging Appliance Challenges With Transfer Learning and Pseudo Labeling."** *IEEE Access*, vol. 13, pp. 106524–106539, 2025.

Professional Certifications

Machine Learning Specialization
DeepLearning.AI & Stanford University

TensorFlow: Advanced Techniques Specialization
DeepLearning.AI

Certified Kubernetes Administrator (CKA) *(In Progress)*
Udemy

Machine Learning Engineering for Production (MLOps) Specialization
DeepLearning.AI

Data Structures and Algorithms Specialization
University of California San Diego

AWS Solutions Architect – Professional *(In Progress)*
Udemy