HYSTON KAYANGE

+82-10-2783-1710 | hylukayange@gmail.com | AI Resume Assistant | Portfolio | GitHub | LinkedIn | Seoul, South Korea

Professional Summary

AI Research Engineer with three years of academic research experience and over five months of industry experience building discriminative models, generative models, and LLM-based systems, including recommendation systems and Retrieval-Augmented Generation (RAG). Experienced in full-stack development using React, Next.js, and the MERN stack. Focused on deploying AI solutions in industrial environments to solve real-world problems and drive business growth.

Skills

- **AI & Machine Learning:** Discriminative AI, Generative AI, Recommendation Systems, Retrieval-Augmented Generation (RAG).
- **Programming & Frameworks:** Python (FastAPI, LangChain, PyTorch, Pandas, NumPy, Scikit-learn), TypeScript, Node.js, ML Pipelines, REST APIs.
- **Databases:** MySQL, SQLite, Firebase, Supabase, Redis, Qdrant, Milvus (Vector Databases).
- **Cloud & DevOps:** AWS (S3, EC2, SES, OpenSearch, SageMaker), Docker, BullMQ.
- **Tools & Platforms:** Jupyter Notebook, Google Colab, VS Code, GitHub, Hugging Face, Postman.
- **Research & Analysis:** Data Analysis, Data Cleaning, Data Visualization, Model Training & Fine-tuning, Model Evaluation, Literature Review.
- **Web Development:** React, Next.js, TypeScript, Tailwind CSS, Express.js, HTML, CSS, JavaScript.

Experience

- **Full-Stack AI Engineer** — Seoul, South Korea
  CY Co., Ltd [🌐] — July 2025 – Present
  - **Project: MI-NOW — Enterprise Business Intelligence Platform** [GitHub]
  - Deployed an enterprise Business Intelligence platform delivering ETL and RAG APIs for Korean manufacturing and global industrial data.
  - Designed a scalable ETL pipeline ingesting Korean and global manufacturing web content and company-specific documents, with OCR support.
  - Implemented a Hybrid RAG architecture combining semantic search, keyword matching, and file-based retrieval, optimized for Korean enterprise data.
  - Evaluated RAG system performance, achieving 97.7% precision, 70.0% recall, 82.3% F1 score, and an average of 12.4% hallucination rate on internal enterprise evaluation datasets.
  - Architected a secure dual-provider AI inference system, using on-premise Ollama models as the primary LLM and CLOVA Studio as fallback, enabling automated health checks and zero-downtime failover.
  - Implemented multi-tenant isolation and role-based access control for secure enterprise usage.
  - Developed an admin dashboard for system monitoring and platform management.
  - **Project: SRT Voice of Customer (VoC) — AI Customer Service Platform**
  - Contributed to the development of an AI-powered VoC platform for Korea's SRT (Super Rapid Train), automatically classifying and responding to customer complaints across multiple departments using Tree-RAG (T-RAG).
  - Built the T-RAG pipeline, including entity detection, Milvus vector search, LLM reranking, and LLM response generation, achieving 95% precision in retrieval and response selection.
  - Implemented a multi-node LLM proxy with GPU/CPU load balancing to support scalable and reliable inference.

- **Assistant Researcher** — Seoul, South Korea
  System Software Lab, Soongsil University [🌐] — Sept 2022 – Feb 2025
  - Worked collaboratively with a Korean team member on the *"XR Twin-based Rehabilitation Training Content Technology Development"* project (IITP/MSIT-funded, Project No. 2022-0-00218), as part of the Digital Twin research team. Focused on AI-driven rehabilitation technologies. Responsibilities included performing data analysis and developing a hybrid heart rate prediction model to support AI-driven custom coaching through personalized fitness recommendations. [GitHub]
  - Led research on probabilistic and adaptive feature selection (ProAdaFs) for deep recommender systems (DeepFM, DCN, Wide & Deep), improving AUC to 0.8088.
  - Authored 3 peer-reviewed papers (1 journal, 2 conferences) on personalized fitness recommendations and feature selection.
- **ICT Manager** — Mzuzu, Malawi
  United Civil Servant SACCO - Head Office [🌐] — Sept 2021 – Aug 2022
  - Managed Fintech systems and network infrastructure, ensuring 99.9% uptime across banking operations.
  - Maintained databases for financial operations and reporting.
  - Handled hardware maintenance and software troubleshooting in the head office and branch locations.

Projects

- **AI Resume Assistant (Personal Portfolio Chatbot, May 2025)**
  - Built and deployed an AI Resume Assistant using Retrieval-Augmented Generation (RAG), enabling recruiters to interactively query my CV,expertise, and experience [Live Demo].
  - Integrated Qdrant vector store with LangChain for semantic search across uploaded resumes and articles.
  - Developed full-stack features including a secure admin dashboard and user interface. [GitHub]
  - **Tools:** Next.js, TypeScript, LangChain, OpenAI, Qdrant, Firebase, Vercel.
- **Feature Selection Tool (Dec 2023)**
  - Developed a Python-based Feature Selector Tool for automated data preprocessing, feature importance analysis, and visualization, supporting classification and regression tasks. [GitHub]
  - **Tools:** Python, Scikit-Learn, Feature-Engine, Seaborn, Matplotlib.

Education

- **MSc in Computer Science and Engineering** — Seoul, South Korea, GPA: 4.14/4.50
  Soongsil University — Sept 2022 – Feb 2025
  - Thesis: "A Multi-Model Machine Learning Framework for Personalized Fitness Recommendations Using DBNs and LSTMs".
- **BSc in Information Communication Technology** — Lilongwe, Malawi, GPA: 3.30/4.0
  Daeyang University — Sept 2017 – Sept 2021

Licenses & Certification

- **Generative AI Fundamentals** — View Credential
  Databricks — May 2025 – May 2027
- **Introduction to Retrieval-Augmented Generation (RAG)** — View Credential
  Duke University (Coursera) — May 2025
- **Fundamentals of LLMs (The LLM Course)** — View Credential
  Hugging Face — May 2025
- **React Foundations for Next.js** — View Credential
  Vercel — May 2025
- **Next.js App Router Fundamentals** — View Credential
  Vercel — May 2025

Publications

C.1 H. Kayange et al. (2024). "ProAdaFs: Probabilistic and Adaptive Feature Selection in Deep Recommendation Systems." *ICOIN Conference*, Vietnam. DOI

C.2 H. Kayange et al. (2023). "Deep Adaptive Feature Selection in Deep Recommender Systems." *Korean Society of Information Science*, Jeju Island. DOI

J.1 H. Kayange et al. (2024). "A Hybrid Approach to Modeling Heart Rate Response for Personalized Fitness Recommendations." *Electronics*, Vol. 13, Issue 19. DOI

- **Google Scholar Profile:** Google Scholar

Languages

- **English:** Fluent
- **Korean:** Beginner