

HIN JIAN HENG

AI ENGINEER

+60 16-738 9583 | jianhenghin0831@gmail.com | [linkedin.com/in/jian-heng-hin](https://www.linkedin.com/in/jian-heng-hin) | <https://github.com/JianHengHin0831>

SUMMARY

First Class Honours Computer Science graduate with experience in building and deploying end-to-end AI systems. Experienced in enhancing model efficiency to enable real-time medical AI applications. Skilled in PyTorch, GCP, and DevOps. Commencing Master's in AI at the University of Malaya (Oct 2025).

EDUCATION

- University of Malaya (#60 QS 2025)**
Master of Artificial Intelligence (Ranked #60 in QS World University Rankings 2025)
Expected Start: Oct 2025
- University of Southampton, Malaysia Campus (#80 QS 2025)**
Bachelor of Science in Computer Science
Oct 2022 - Jun 2025
 - Achievement:** First Class Honours (Percentage: 72.92%)
 - Relevant Coursework:** Machine Learning, Natural Language Processing, Cloud & Website Security
- Professional Certifications**
 - IBM AI Engineering Professional Certificate
March 2025

TECHNICAL SKILLS

- Languages:** Python, SQL, Java, JavaScript/TypeScript
- AI / ML:** PyTorch, TensorFlow, LangChain.js, Hugging Face Transformers (.js & Python), Pandas, NumPy, Scikit-learn
- Concepts:** MLOps, RAG, XAI, NLP, Vector Embeddings, Computer Vision
- Cloud & DevOps:** GCP, Docker, Vercel, Supabase, pg_vector, Git, Tailwind CSS, Nuxt.js

PROJECT

- Interpretable ML for Medical Analysis** | *Python, PyTorch, TabNet* [\[GitHub\]](#)
 - Designed a distilled surrogate model to provide real-time, interpretable predictions for a TabNet model in a medical CAD setting. Boosted inference speed from **10s to 50ms** and predictive accuracy from **94% to 97%**.
- Secure Cloud-Based Group Messaging Platform** | *GCP, DevOps* [\[GitHub\]](#)
 - Architected and deployed a scalable, real-time messaging application on Google Cloud Platform (GCP).
 - Implementing core DevOps practices including automated monitoring and auto-scaling. built a strong foundation in cloud infrastructure and CI/CD, skills directly transferable to building production-level MLOps pipelines.
- AI-Powered Semantic Search Engine** | *Nuxt.js, Hugging Face, LangChain.js, Supabase* [\[GitHub\]](#)
 - Architected and developed a full-stack AI knowledge base with a RAG pipeline for high-accuracy semantic search
 - Deployed on Vercel, running open-source models locally via Transformers.js, integrated with Supabase and custom SQL for efficient vector retrieval, achieving zero API cost and high performance.
- Generative AI & Computer Vision Projects** | *Python, PyTorch, Hugging Face* [\[GitHub\]](#)
 - Trained generative models including DCGANs and explored modern architectures like Diffusion models.
 - Fine-tuned a Hugging Face Transformer for a text classification task, achieving competitive accuracy.
- Full-Stack Library Management System** | *JavaScript, Node.js* [\[GitHub\]](#)
 - Led the full-stack development of a web-based library management system, responsible for backend, frontend, and user authentication, implementing a basic NLP-powered chatbot using cosine similarity for query matching.

INTERNSHIP EXPERIENCE

- Software Engineer Intern | Feedme POS | June 2024 – Oct 2024**
 - Developed a dynamic, component-based website builder in Nuxt.js to empower non-technical staff to manage corporate website content and layouts.
 - The system was deployed to power the live corporate site ([Feedme.ai](#)), significantly improving content update efficiency and reducing developer workload.