Kwang Yang Chia - Singaporean Citizen

90180736 | chia.kwang.yang@gmail.com | linkedin.com/in/kwang-yang-chia | github.com/kwangyy | kwangyyinc.com

Education

National University of Singapore   **Aug 2022 - May 2025**
Bachelor of Science (Honours) in Data Science & Analytics
- **GPA:** 4.30/5
- **Relevant Modules:** Data Visualisation (DSA2101) - A+, Artificial Intelligence: Technology and Impact (IT1244) - A, Honours Project in Data Science (DSA4288) - A, Data Science In Practice (DSA3101) - A-

Working Experience

Software Engineer Intern, SimpleAI   **Jul 2024 - Nov 2024**
- Overhauled prompt retrieval stack by re-architecting vector database implementation to increase **relevant document retrieval accuracy by 20%** via higher-dimensional embeddings & enriched metadata indexing
- Developed production-grade RAG (Retrieval-Augmented Generation) workflows integrating MongoDB, React, and TypeScript, **reducing document fetch latency by 15%**
- Fine-tuned prompts and decoding constraints using OpenAI's structured output format, increasing extraction accuracy from noisy OCR accounting invoices

AI Engineer Intern, AgentScale AI   **May 2024 - Jul 2024**
- Designed a RAG workflow with Query Expansion & Query Rewriting for Atlas, a Large Language Model-based browser assistant Chrome extension, **cutting input token costs by 40%**
- Optimised LLM methods with model refinement & efficient caching methods to **cut client-side inference time by 69.84% from 12.6 to 3.8 seconds**
- Developed automated scripts to clean HTML files for LLM-enabled code generation to scrape websites efficiently, **reducing token cost by approximately 60%**
- Built an LLM-powered semantic knowledge graph system for real-time sales prospecting with contextual candidate links and relevance scoring
- Formulated new context attribution rate & context miss rate evaluation metrics with LLM-as-a-judge responses

Artificial Intelligence & Data Science Teaching Assistant, NUS   **Aug 2023 - Present**
- Led core learning modules for **150+ undergraduates** in AI and Data Science courses, achieving a **40% full-score midterm rate** in DSA2101 (Data Visualisation) and **85.7% average** in IT1244 (Intro to AI)
- Assessed and provided actionable feedback on **28 ML projects** across NLP, vision, and tabular domains, demonstrating expert-level proficiency in cross-domain ML evaluation
- Designed and deployed hands-on scikit-learn and NumPy assignments adopted by other TAs and students for exam preparation and project bootstrapping

Research

Combatting Game Toxicity With Large Language Models (with Prof Tong Xin), NUS   **Jan 2025 - May 2025**
- **Designed and implemented RAG-Keyword**, a novel retrieval mechanism combining strengths of RAG-Token and RAG-Sequence to optimize efficiency, reduce noise, and prioritize relevance of keywords
- **Achieved 0.93 recall and 0.74 F1-score on toxic class** (14.7% prevalence) using a hybrid RAG pipeline with Llama-3.1–8B
- Reduced **irrelevant context retrieval by 58%** using a novel RAG-Keyword mechanism combining token-weighted and keyword-priority scoring
- Generated interpretable explanations with semantic justification chains using Pinecone vector segments aligned by region-specific slang
- Decreased **false negative rate by 36%** while maintaining macro precision, demonstrating consistent superiority over non-RAG LLM baselines in ablation studies

Data Science Research Intern (with Prof Linda Sellou), NUS   **Mar 2023 - Aug 2023**
- Confirmed instrument consistency with **Cronbach's Alpha ≥ 0.85** & visualized Likert response skew with diverging bar plots for internal reporting
- Verified statistically significant shifts in conceptual confidence via paired t-tests **on 52 survey responses**

Projects

AI-Powered Business Insight Generator - NUS – Guru Network AI Innovation Challenge 2025　　　**Mar 2025**
- Built and deployed a Next.js-powered LLM dashboard integrating Pinecone and OpenAI APIs, generating real-time financial insights from structured business data
- Tuned vector-indexed prompt flows across SQL and MongoDB backends to support generalization across multiple industries

AI-Powered Knowledge Graphs - A- in DSA3101　　　**Oct 2024**
- Achieved **>90% precision in entity-linking** using HuggingFace Transformers deployed via async API and Neo4j to visual React frontend
- Containerized deployment pipeline using Docker Compose for reproducibility & scalability

Accessible Personal Finance for Aging Population - HackSingapore 2024　　　**May 2024**
- Developed a CPF-focused financial literacy assistant for aging users by scraping structured data from DBS, DollarsAndSense, and integrating it into an accessibility-first UI powered by OpenAI APIs

Credit Card Approval Classification - Top 5 Project in IT1244 Cohort　　　**Apr 2023**
- Tuned XGBoost ensemble models to outperform random forests and logistic regression by 10% AUC

Activities

Media Director, NUS Raffles Hall　　　**Aug 2024 - Apr 2025**
- Engineered and deployed scalable systems with comprehensive unit testing and user-centric UIs to support administrative and ad hoc workflows **across 80 CCAs**
- Led **7 CCA executive committees** overseeing media operations for **250+ events**, ranging from small-scale productions to large-scale, multi-day hall-wide events at Raffles Hall

Student Lead, NUS Raffles Hall Developers　　　**Sep 2022 - Apr 2024**
- Shipped & maintained Telegram bots, web apps & feedback tools used by **671 residents** with **>95% uptime**
- Eliminated manual deployment steps by containerizing all services and implementing one-click deployments via DigitalOcean scripts, **reducing loading speeds by 90%**

Workshop Director, NUS Statistics & Data Science Society　　　**Aug 2022 - Apr 2025**
- Delivered **14 workshops to 300+ participants** on LLMs, NLP, SQL, & data cleaning with collaborations with student clubs such as NUS Hackers, NUS Fintech Society & NUS Business Analytics Consulting Team
- Increased **reusability of training materials by 3x** by standardizing Jupyter Notebooks and slide formats for distribution

Technical Skills

- Languages: Python, SQL, JavaScript, TypeScript, Bash
- Frameworks: React, Flask, Docker, Next.js
- AI Tooling: OpenAI API, Pinecone, LangChain, Retrieval-Augmented Generation (RAG), HuggingFace, Transformers
- Data & Infra: MongoDB, PostgreSQL, MySQL, Neo4j, DigitalOcean, AWS
- Dev Tools: Git, Jupyter, Unix, Anaconda, CI/CD (GitHub Actions), Streamlit