Caleb Kan

+44 0782 8210751 | calebkan1106@gmail.com | linkedin.com/in/caleb-kan | github.com/caleb-kan | calebkan.com

Education

Imperial College London — London, United Kingdom
Degree: Bachelor of Engineering – BEng, Computing — Sep 2023 – Jun 2026
Grade: First-Class Honours (Predicted)

Experience

Software Engineer Intern — Jul 2025 – Sep 2025
HubSpot — London, United Kingdom
Software Engineer Intern — Aug 2024 – Oct 2024
Midas Advisory — London, United Kingdom

- Automated non-operating expense data processing for top 15 U.S. banks using open-source LLMs, web scraping, APIs, and parallelisation with Ray, cutting processing time from hours to minutes and boosting accuracy to 97%.
- Implemented a Milvus vector database with high-dimensional embeddings using the BAAI/bge-m3 model. Built a scalable architecture for future ML model and data source integration, enhancing efficiency.
- Developed a hybrid search system combining Milvus's RRFRanker and dual LLMs: a small LLM achieving 93% filtering accuracy and Hermes-3-8B-GGUF for schema-based reporting. Improved company market intelligence and analytics.

Research Intern — Jul 2023 – Jul 2023
Department of Computer Science, City University of Hong Kong — Hong Kong

- Achieved 92.5% accuracy in GDPR compliance detection across 400+ websites by developing a GDPR compliance checker extension with a distinguished professor, utilising BeautifulSoup4, LangChain, and OpenAI's LLM for automated analysis.
- Increased daily active users by 45% and improved user-reported compliance understanding from 3.2 to 4.7 on a 5-point scale by designing a user-friendly interface displaying website GDPR compliance status, enhancing data privacy awareness.
- Boosted successful data erasure request rate by 150%, processing 100+ requests in the first month by creating a one-click feature for automated data erasure requests, streamlining the exercise of GDPR rights.

Projects

WACC (Compiler Project) | Scala — Jan 2025 – Mar 2025

- Developed a WACC compiler frontend, building a lexer and parser for syntax analysis, creating an abstract syntax tree representation, implementing a symbol table for semantic checks, and designing descriptive error reporting to aid debugging.
- Implemented the backend using TAC intermediate representation with ARM32/AArch64 support, architecture-specific dependencies, code optimisations (constant propagation, folding, control flow analysis), and a standard math library.

PintOS (Operating Systems Project) | C — Oct 2024 – Dec 2024

- Enhanced OS kernel functionality by implementing timer-based thread synchronisation, advanced priority scheduling with priority donation, and the BSD scheduler, ensuring efficient multitasking and thread management.
- Developed a virtual memory subsystem, including paging, frame management with second-chance eviction, and supplementary page tables, enabling support for user programs, memory-mapped files, and stack growth.

ARMv8 Emulator, Assembler, and Visualiser | C — May 2024 – Jun 2024

- Engineered a cycle-accurate ARMv8 emulator and two-pass assembler in C, implementing precise instruction management and robust error handling in compliance with ARMv8 specifications.
- Developed a SDL2-based GUI for real-time emulation visualisation, featuring drag-and-drop assembly parsing and dynamic rendering of CPU states, registers, memory maps, and ALU.

AI Research Agent | Python, NoSQL — Aug 2023 – Sep 2023

- Developed a Python AI research agent using LangChain, OpenAI's LLM, Serper API, BeautifulSoup4, and X API for automated search, web scraping, cited summaries, and content posting. Integrated MongoDB Atlas for geolocation tracking.
- Implemented a Streamlit web app enabling anyone to use the AI research agent, integrating a geospatial visualisation pipeline to map worldwide usage using MongoDB Atlas geolocation data. Analysed engagement patterns of 70+ global users.

Honours and Awards

1st Place – G-Research London Coding Challenge 2024
Issued by G-Research

- Achieved First Place among 40+ contestants in the G-Research London Coding Challenge, a 3-hour programming contest where participants compete to achieve the highest score on a given problem.

Skills

Languages: **Native**: English · Mandarin · Cantonese
Programming Languages: C · Haskell · HTML/CSS · Java · Kotlin · NoSQL · Prolog · Python · Scala · SQL
Frameworks: Bootstrap · Django · FastAPI · Firebase · LangChain · Milvus · MongoDB · MySQL · Streamlit
Libraries: BeautifulSoup4 · Matplotlib · NLTK · NumPy · OpenCV · Pandas · Playwright · PyTorch · Ray · SciPy · SymPy
Developer Tools: CI/CD · Docker · GDB · Git · Google Colab · Hugging Face · JetBrains · Ollama · Unix · Valgrind · VS Code