Caleb Kan

+44 07828210751 | 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 — Aug. 2024 – Present
Midas Advisory | Python — London, United Kingdom

- Engineered a modular app for processing non-operating expense data of the top 15 U.S. banks by integrating open-source LLMs, web scraping tools (Playwright, BeautifulSoup4, Selenium), and APIs (Requests). Reduced data collection time from hours to minutes and increasing accuracy from 85% to 97%.
- Decreased data retrieval time from minutes to seconds by implementing a vector database (Milvus) with embeddings; engineered a dual-model system using a small LLM for 93% accurate data filtering and a large LLM (LLama 3.1) for schema-based reporting.
- Developed an adaptable architecture that seamlessly integrates new models, enhancing market intelligence and banking sector analysis capabilities to support informed decision-making across departments.

Research Intern — Jul. 2023 – Jul. 2023
Department of Computer Science, City University of Hong Kong | Python — Kowloon, 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 over 100 requests in the first month by creating a one-click feature for automated data erasure requests, streamlining the exercise of GDPR rights.

Projects

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

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

Chess AI | Kotlin — Dec. 2023 – Jan. 2024

- Developed a Kotlin AI for a pawn race chess game, implementing complete game logic, move validation, and an ASCII terminal interface. Optimised AI decision-making using a depth-limited minimax algorithm with alpha-beta pruning.

AI Research Agent | Python, NoSQL — Jul. 2023 – Aug. 2023

- Built a Python AI research agent using LangChain and OpenAI's LLM for cited summaries, integrating web scraping, search, automated content posting on X, and geolocation tracking via BeautifulSoup4, Serper API, X API, and MongoDB Atlas.
- Created a Streamlit geospatial visualisation pipeline to map global AI agent usage, leveraging geolocation data from MongoDB Atlas to analyse user engagement patterns.

Skills

Languages: **Native**: English · Mandarin · Cantonese
Programming Languages: C · Haskell · HTML/CSS · Java · Kotlin · NoSQL · Python · SQL
Frameworks: Bootstrap · Django · FastAPI · Firebase · Flask · LangChain · Milvus · MongoDB · MySQL · PostgreSQL · SQLite · Streamlit
Developer Tools: Android Studio · Atom · BitBucket · CLion · Git · GitHub · GitLab · GNU Debugger · Google Colab · Hugging Face · IntelliJ · LaTeX · Ollama · PyCharm · Unix · Valgrind · Visual Studio · VS Code · Xcode