Lim Chee Kin
Kuala Lumpur, Malaysia • +6011 2127 6330 • limcheekin@vobject.com
linkedin.com/in/limcheekin • github.com/limcheekin

Professional Summary

Senior Tech Lead, AI Engineer, and Solution Architect with over 25 years of experience, including a recent intensive focus on AI engineering. Proven ability in building and leading multinational teams to deliver advance software solutions. Expertise in architecting and implementing open-source AI, local AI infrastructure, and data sovereignty solutions, particularly for mobile-first web applications. Committed to standardizing processes and driving innovation through privacy-preserving and democratized AI.

Skills

- **AI & Machine Learning:**
  - Local AI/Edge Computing Strategy & Implementation, Retrieval Augmented Generation (RAG) Systems, Advanced Text Embeddings, LLM Integration & Deployment (Local & Cloud), AI Ethics & Data Privacy, AI Agents
  - **Tools & Frameworks:** Ollama, LocalAI, llama.cpp, Whisper.cpp, Piper (TTS), FastWhisperAPI, MLC LLM, FastKoko, Gemini API, Groq API, Open Text Embeddings, FastAPI, Sentence Transformers, Modal.com, HuggingFace (SmolAgents, Transformers, FastRTC), Open WebUI, Telegram AI Chatbot
  - **Vector Databases:** Qdrant, PostgreSQL, SurrealDB
- **Proficient:** Python, Java, Spring Boot, Grails Framework, Micronaut Framework, Flutter, MongoDB, Oracle DB, MySQL DB, GraphQL API, OData API, Amazon Web Services (AWS), Docker, JUnit, Design Patterns, Application Architecture, Git, Subversion
- **Moderate:** DevOps, Jenkins, Solidity, Rust, Substrate
- **Methodologies:** Agile, CMMI Level 3
- **Languages:** English (Fluent), Chinese

Professional Experience

Founder, AI Engineer & Solution Architect | RAG.WTF & AI Initiatives | 2021 – Present

- Championed and developed open-source AI solutions focusing on data ownership, privacy, and efficient local AI processing; architecting and building RAG.WTF (to be open-sourced) as a platform for private, personalized knowledge discovery and management.

- Engineered [open-text-embeddings](), a Python library providing an OpenAI-compatible API for diverse open-source sentence transformer models (e.g., BGE, E5, all-MiniLM-L6-v2), enabling significant cost reduction (estimated 70-90% vs. proprietary APIs) and eliminating vendor lock-in for embedding generation within RAG.WTF.
- Developed [Talk To AI]() application, integrating HuggingFace's FastRTC for low-latency (<300ms) real-time voice AI interactions; engineered flexible backend supporting local STT (LocalAI/Whisper.cpp, FastWhisperAPI), LLM (LocalAI/Llama.cpp, MLC LLM), and TTS (LocalAI/Piper, FastKoko) alongside cloud-based STT (Groq), LLM (Groq Llama3), and TTS (Microsoft Edge) APIs.
- Pioneered and operationalized a self-hosted AI infrastructure using a dedicated local AI server (Coolify, Docker), successfully running multiple local LLMs (Llama 3.2, DeepSeek-R1 distilled models, Qwen3, Phi4, Gemma3, etc. via Ollama, LocalAI, llama.cpp), embeddings, and STT/TTS to ensure complete data privacy, operational control, and resilience for AI workloads.
- Advocated for and implemented robust data sovereignty in AI applications by integrating embedded database solutions like [SurrealDB.wasm]() with RAG.WTF, empowering users with complete ownership and control of their AI-generated data and knowledge bases.
- Strategically utilized Modal.com for scalable, serverless execution of demanding AI/ML workloads, such as batch processing and experimental model fine-tuning, complementing local AI capabilities with cloud-based flexibility.
- Previously initiated [Talking Book]() YouTube channel (AI-powered summaries and Telegram bot) and [Fluwix]() (Flutter showcases), foundational explorations into applied AI and developer tooling.

Senior Java Tech Lead | DXC Technology | 2013 – 2020

- Led a 6-developer international team (China, Europe, Malaysia) to deliver the Analytic Console for the Agile Process Automation (APA) platform (VueJS, Spring Boot, GoLang, MongoDB, AWS), contributing millions to company revenue.
- Headed a 5-developer international team (US, Malaysia) to create the EPIC Configurator for internal Enterprise Invoice Processing using JQuery UI, Grails Framework, and Oracle DB, significantly streamlining configuration processes and reducing errors.
- Implemented a Divestiture and Acquisition (DnA) dashboard (AngularJS, Jersey, Spring, MongoDB, AWS), providing critical, real-time data insights that directly supported strategic executive decision-making during the HP split and Micro Focus acquisition.
- Prototyped an Intelligent Voice Agent (Conversational AI) leveraging Amazon

Connect, AWS Lambda, Amazon Lex, Dialogflow, and Micronaut Framework, demonstrating potential for substantial service enhancement and operational cost reduction.

Senior Java Tech Lead | Zurich Technology Services Malaysia | 2012 – 2013

- Led a 5-member production support and development team for motor and property claims systems (IBM WebSphere, FileNet, DB2, Spring, GWT, Hibernate/JPA).
- Resolved over 60 critical production issues within weeks, dramatically improving system stability and user satisfaction, evidenced by maintaining ticket counts at a single-digit level.

Independent Software Professional | Self-Employed | 2010 – 2012

- Developed and maintained impactful open-source Grails Framework projects:
  - **Grails Activiti Plugin:** BPM workflow system; grew the forum to 100+ members, addressed 100+ queries.
  - **Grails Form Builder Plugin:** Empowered non-programmers to create online forms.
  - **JQuery Validation UI Plugin:** Enhanced developer productivity with client-side validation.
  - Explore all 10+ projects at limcheekin.blogspot.com/p/my-grails-plugins.html

Senior Solution Consultant | Penril Datability (M) | 2008 – 2010

- Secured a RM6 million Internet Banking Systems tender from Kuwait Finance House (Malaysia) Berhad.
- Secured a RM2 million eCustody System tender from Malayan Banking Berhad.
- Led a 4-member team in developing the customer web portal for Kuwait Finance House (Malaysia) Berhad.

Technical Lead | CMG Online | 2005 – 2007

- Led an 8-member international team to deliver a RM2 million Integrated Hospital Inventory System for government hospitals.
- Defined CMMI Level 3 compliant organizational standards, increasing staff productivity.

Previous Roles (1999 – 2005)

- **Software Consultant** | Media Trend | 2004 – 2005
- **Software Engineer** | Siemens Multimedia | 2002 – 2004
- **System Engineer** | Online One Corporation | 2001 – 2002
- **System Engineer** | Tee Yam Holding | 1999 – 2000

Education & Certifications

- BSc (Hons) Degree in Computing (Multimedia Technology), 2nd Upper
  Staffordshire University, UK | 2000 – 2001
- Higher Diploma in Software Engineering, Credit
  A.P.I.I.T | 1998 – 1999
- Diploma in Computing and IT, Credit
  A.P.I.I.T | 1997 – 1998
- **Polkadot DevCamp 2022**
- **HuggingFace Agents Course** (Enrolled, In Progress)