

陳智賢 Kade Chen — LLM 應用工程師履歷

陳智賢 Kade Chen

LLM 應用工程師 / ML 工程師 應徵：台灣大哥大 [I108] LLM 應用平台架構師（可遠端工作）

 welovetosharetogether@gmail.com |  Taipei (UTC+8)  box1401.github.io | 
github.com/box1401 |  linkedin.com/in/kade-chen-rs  Live demo：compass-web-aujzogkiva-de.a.run.app

概要

ML / LLM 應用工程師，國立中央大學遙測科技碩士（2026.03 畢業），自 2024 年 9 月起累積近兩年深度學習實戰經驗，研究主題為 SAR 影像深度學習物件偵測。已 ship 一個 production 級多代理人 LLM 應用 — **Compass Equity**（FastAPI + Next.js 14 + LangGraph + 5 agent + Hybrid RAG + 部署於 GCP Cloud Run），原始碼公開、live demo 可立即驗證。同時為 PyTorch 地理空間庫 `torchgeo` 兩個 PR 的作者（5 maintainer review、100% test coverage）。

可量化指標

指標	數值
Compass Equity 端到端 production 部署	GCP Cloud Run × 3 service，月成本 ~\$0.20
自動化測試	17/17 unit tests pass 、20 題 RAGAS golden set
RAG 檢索品質提升	top-1 score：dense-only 0.78 → hybrid+rerank 0.99 （blog #2 公開實測）
Infra-as-Code	Terraform 11 個 .tf、CI/CD 全自動部署、Workload Identity Federation 無 SA JSON
多 LLM provider	<code>LLM_PROVIDER=gemini\ ollama</code> 一行 env 切換
開源貢獻	torchgeo PR #3655（28 tests / 100% coverage、5 maintainer review）+ #3662
公開技術寫作	3 篇英文技術部落格 ship 於 GitHub Pages

對應職缺需求

必備技能

JD 條目	落地實作
1. 深度學習 / ML	碩論：SAR 斑點雜訊抑制 vs DL 船舶偵測效能評估 (PyTorch、Lee / Refined Lee / Frost)；bge-reranker-v2-m3 cross-encoder 部署於 Cloud Run
2. NLP / LLM / Transformer	Gemini 2.5 Flash + function calling、Ollama Qwen2.5 本地推論；LangGraph state machine 編排 5 個 agent
3. Python / FastAPI / LangChain (LangGraph) / React	後端 FastAPI + Pydantic v2 + SQLAlchemy 2.0 async；前端 Next.js 14 (App Router) + shadcn/ui + Tailwind
4. SQL / Pandas / NumPy	PostgreSQL + pgvector + asyncpg；技術指標計算 (MA / RSI / 乖離率 / PE 百分位) 以 NumPy 實作
5. 大規模分散式系統	Cloud Run × 3 service (api / web / reranker) + Cloud Tasks 佇列 + 外部 Supabase Postgres；scale-to-zero、Terraform 管理
6. 公有雲 (GCP / AWS / Azure)	GCP：Cloud Run、Artifact Registry、Secret Manager、Cloud Storage、Cloud Tasks、Workload Identity Federation
7. RAG 技術實作	Hybrid RAG：BM25 (Postgres tsvector) + Dense (pgvector cosine) + RRF fusion + cross-encoder rerank
8. 2 年以上 DL 實戰經驗	2024.09 – 至今近 2 年：碩士論文 + 2 個 OSS PR + 5 個個人作品 (v1 stock-rag 至 v2 Compass Equity 演進)

加分條件

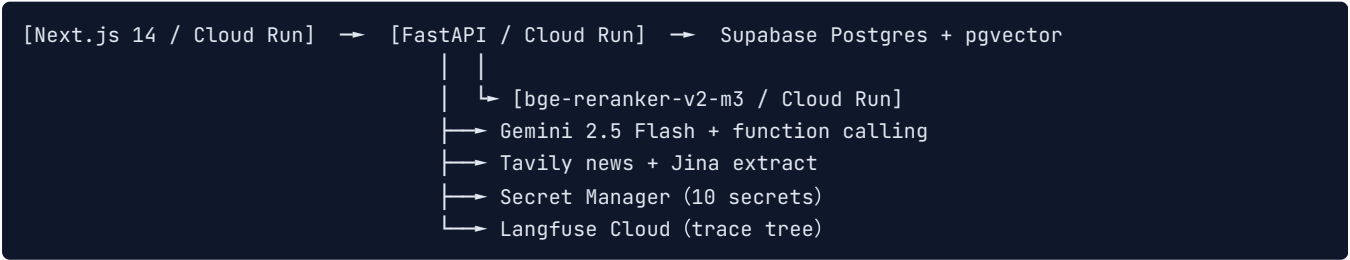
JD 條目	落地實作
PyTorch / TensorFlow	PyTorch (碩論模型訓練、torchgeo Kornia transform、bge-reranker、ChatTTS GPU 推論)
生成式 AI 應用 (智能客服 / 知識管理)	Compass Equity 為投研領域知識管理：thesis 文件 ingestion → 語意檢索 → 引用追蹤 → 自然語言問答
Agent Flow 設計與優化	LangGraph 條件邊 (should_revise) 控制 RiskAgent self-critique 迴圈；單次 /analyze 觀察值：9-14 個 trace events
Agentic AI 應用	AnalystAgent 透過 Gemini function calling 動態選擇 rag_search / news_search / read_url 工具，非 ReAct prompt parsing
新技術研發 / OSS	torchgeo PR #3655 Lee filter as Kornia transform + PR #3662 Refined Lee
跨部門 / 跨領域協作	碩論研究與指導教授每週同步；torchgeo 與遙測 / ML 雙領域 maintainer 多輪 code review
創新解題	在 production 部署過程中診斷 Cloud Run secret version 與 Terraform ignore_changes 交互、IAM 最終一致性延遲、Artifact Registry 與 docker push 順序等多層問題，整理為公開 retrospective
善用 AI coding tools	整個 Compass Equity 透過 Claude Code 協作完成；blog 第三篇將開發流程寫成 case study
英文聽說讀寫	全部 portfolio README、code comment、3 篇技術部落格皆英文撰寫；TOEIC 可依需求安排測驗

主要專案

Compass Equity — 多代理人台股投研平台 | 2026.04 – 2026.05

🔗 連結 - Live demo：<https://compass-web-aujzogkiva-de.a.run.app> - Source：<https://github.com/box1401/stock-rag-agentic> - 部落格 3 篇：[總覽](#) · [Hybrid RAG 數據](#) · [Cloud Run 部署 retro](#)

架構



5 個代理人 + Self-critique 迴圈

Supervisor 編排 → Data 抓 7 個 TWSE / FinMind / Yahoo endpoint → Analyst 透過 function calling 自主選用 rag_search / news_search / read_url → Risk 審查 (hallucination / 單面論述 / 模糊修飾語檢測) → 未通過則回 Revise (max 2 revisions) → Reporter 渲染 markdown + citations。

核心數據

- Top-1 retrieval score : dense-only 0.78 → hybrid + cross-encoder rerank **0.99**
- 端到端 production 延遲：10–30 秒（單次審查） / 90–120 秒（觸發 2 次 revise）
- 三 service Cloud Run scale-to-zero，月成本 ~\$0.20（Artifact Registry only）
- 實測 trace（2026-05-05 prod）：14 events、6 citations、RiskAgent 觸發 3 次 review

Stack

層	技術
前端	Next.js 14 App Router · TypeScript · shadcn/ui · Tailwind · TradingView Lightweight Charts · i18next (中英)
後端	FastAPI · Pydantic v2 · SQLAlchemy 2.0 async · Alembic · LangGraph
LLM / Embedding / Rerank	Gemini 2.5 Flash · gemini-embedding-001 (768d) · bge-reranker-v2-m3
RAG	Postgres tsvector BM25 · pgvector dense · RRF fusion · cross-encoder rerank
資料	Supabase Postgres · GCS · Cloud Tasks
觀測	Langfuse Cloud · 自製 SSE 即時 trace UI
部署	Cloud Run × 3 · Terraform · Workload Identity Federation · GitHub Actions · Docker

Stock RAG v1 — 本機 Ollama 投研報告 | 2026.01 – 2026.03

Compass Equity 的前身，專注「local LLM、無 API 月費」設計。

- 整合 7 個 TWSE 官方 endpoint + FinMind API（法人買賣超、融資融券、技術指標、財報）
- 多執行緒平行爬蟲（3 worker 新聞搜尋 + 4 worker Jina 文章擷取）
- 本機 Ollama + Qwen2.5:14b 推論，產出 21 檔台股每日 / 每週 markdown 報告
- 累計訊號 CSV（23 欄位）供回測分析
- Repo：<https://github.com/box1401/stock-rag>

智慧冰箱管理系統 | 2025

LINE Bot 拍照即辨識食品 → Google Gemini Flash 視覺模型輸出結構化 JSON → Google Sheets 後端分頁存儲 → 每日保存期限檢查與過期推播。Docker 容器化部署。Repo：<https://github.com/box1401/refrigerater>

NinjaPause — 手勢控制音樂播放器 | 2024

MediaPipe Hands 21 個手部關鍵點偵測，自製手指伸展計數 + 捏合距離演算法。雙模式：Spotify Web API 精確音量控制 + 系統媒體鍵通用控制。cooldown 機制防誤觸。Repo：<https://github.com/box1401/ninjapause>

Streamlit web app ; EPUB 解析 (TOC / HTML 清洗 / 智慧章節擷取) + 三 TTS 引擎切換 (Edge TTS / Google Cloud WaveNet / ChatTTS + PyTorch GPU) ; 智慧文本分段 + 音訊片段串接。Repo : <https://github.com/box1401/ebook-reader>

開源貢獻 — torchgeo (PyTorch 地理空間影像庫)

- [PR #3655](#) — `LeeFilter` 包裝為 Kornia `IntensityAugmentationBase2D` 變換：+319 行、28 tests、100% coverage、5 maintainer review、lead maintainer LGTM
- [PR #3662](#) — `RefinedLeeFilter` (SNAP / GEE-style edge magnitude hybrid)：+237 行、+23 tests、stack on #3655、目前在審

學歷

國立中央大學 — 遙測科技碩士學位學程 | 2024.09 – 2026.03 - 論文：合成孔徑雷達影像中斑點雜訊抑制前處理對深度學習船舶偵測效能之影響評估 - 指導教授：任玄 博士 - 研究領域：SAR 影像前處理、Lee / Refined Lee / Frost 濾波器比較、深度學習物件偵測

國立成功大學 — 地球科學教育學系 (學士)

技能

類別	技能
程式語言	Python · TypeScript · JavaScript · SQL · HCL (Terraform)
AI / ML	PyTorch · LangGraph · LangChain · Hugging Face · Sentence-Transformers · OpenCV · MediaPipe · pgvector
LLM 應用	Gemini API · Ollama (Qwen2.5) · Function calling · Hybrid RAG · Agentic flow · Langfuse trace · RAGAS eval
Web	FastAPI · Pydantic v2 · SQLAlchemy 2.0 async · Next.js 14 · shadcn/ui · Tailwind · SSE
資料	PostgreSQL + pgvector · Supabase · Pandas · NumPy · Alembic
雲端	GCP (Cloud Run · Artifact Registry · Secret Manager · Cloud Storage · Cloud Tasks · Workload Identity Federation)
DevOps	Terraform · GitHub Actions · Docker · CI/CD · pytest · ruff · mypy
語言	繁體中文 (母語) · English (讀寫流暢，技術文件 / 部落格 / OSS code review 全英)

公開技術寫作 — box1401.github.io

1. [Building Compass Equity — multi-agent equity research copilot](#) — 8 min
 2. [Hybrid RAG with RRF + cross-encoder — what the numbers look like](#) — 10 min
 3. [Shipping a LangGraph stack to Cloud Run with keyless OIDC](#) — 11 min
 4. [Building a Local-LLM RAG Pipeline for Taiwan Stocks](#)
 5. [Implementing the Lee Filter as a Kornia Transform](#)
-

工作型態

- 居家上班 / 遠端：UTC+8 時區、自備獨立工作環境
 - Onsite（台北）：彈性配合
 - 可開始日期：論文口試完成（2026.03）後立即可全職入職
-

履歷所述全部公開可驗證。Live demo、source、blog、OSS PR 連結點選即可。