聯絡

hengweibin1898@gmail.com

www.linkedin.com/in/wayne-hengweibin-ai (LinkedIn)
waynemyportfolio.netlify.app/ (Portfolio)

熱門技能

Agentic AI Development
軟體設計
人工智慧

Languages

中文 (Native or Bilingual)
英文 (Limited Working)

Certifications

Gemini Certified Educator

王偉斌

Generative AI Engineer | Agentic RAG & LLM Optimization | Python, LangChain | Building Scalable AI Agents
台灣 臺北市 台北

簡介

專注於 Agentic RAG 架構、工業級 LLM 服務部署與前沿 Context/Harness Engineering 的 AI 系統架構師。擁有豐富的 AI 代理人/智能體 （AI Agent）開發與微調經驗，擅長設計專用工具（Function Calling/Tools），協助 AI 精確理解並處理真實半導體與製造業產線的複雜結構化/非結構化資料。具備端到端（End-to-End）全棧 AI 服務的研發與維護能力，涵蓋後端 API、向量與關聯式混合資料庫（VectorDB/SQLDB）架構設計，以及高可用 Docker 容器化服務維運。致力於在嚴格保障企業資料隱私的前提下，打破真實工業數據的壁壘，構建高度可靠、低幻覺的企業級智慧決策引擎。

經歷

ASE Global
工程師
2024 年 11 月 - Present (2 年)
台湾

主導 Agentic RAG 架構設計與企業級 AI 服務部署
主導 Agentic RAG 系統架構，專為真實工業與產線複雜資料設計專用 AI 工具鏈（Tool Calling）；深度導入 Prompt / Context / Harness Engineering，突破傳統檢索瓶頸，顯著提升 LLM 對特定領域知識的理解與推論精準度。

全棧 AI 服務整合與混合檢索系統維護
維護端到端（End-to-End）前後端及 AI 服務，建構 VectorDB 與 SQLDB 混合檢索架構，兼顧結構化與語意資料管理，確保生產環境系統具備高可用性與低延遲維運品質。

模型微調與地端容器化架構運維
執行領域專用 LLM 微調（Fine-tuning）流程，並透過 Docker 容器化進行地端（On-premises）服務部署、資源監控與故障排除，在嚴格保護企業資料隱私前提下極大化模型應用效能。

Hihealth
演算法實習生

2022 年 2 月 - 2022 年 6 月 (5 個月)
台灣 臺北市

在「您好健康」實習期間，我擔任的是演算法實習生，除了負責 Android 應用的穩定性維護之外，我更開發了高效的數據收集架構，確保能精準收集到研發所需要的感測器資料。

實習期間，我與 Research 實習生緊密寫作，尋找前瞻研究並轉化為實際功能。我主導了基於單一慣性感測單元（IMU）的動作便是專案，利用 Python、Java、Matlab 與機器學習技術，在僅有單組傳感器的資料限制下，剋服了動作細節辨識的挑戰，並成功將模型整合至手機上進行實時（Real-time）效果測試。

這段經歷磨練了我與研究團隊的交流，合作並尋找最佳的技術解決方案。

學歷

國立臺北科技大學
碩士, 資訊工程系 · (2022 年 9 月 - 2024 年 8 月)

台北科技大学
學士, 資訊工程系 · (2018 年 9 月 - 2022 年 7 月)