

郭東輝

milk333445@gmail.com | github.com/LLMSystems | 台北, 台灣

個人摘要

具備兩年機器學習與 LLM 微調與應用開發經驗，熟悉 Python、SQL、RAG、AI Agent、模型微調與企業內部資料整合。曾於富邦金控主導越南市場研究 Agent、法規 QA 小模型幻覺降低、新聞檢索 MCP 工具與 LLM 部署管理平台，能從資料收集、清理、標註、模型訓練到企業系統整合端到端落地 AI 解決方案

工作經歷

資料科學家 儲備幹部, 富邦金控

台北 2024/07 – 2026/04

- 主導設計與實作富邦投信越南市場研究 Agent 系統，整合總經、股票、新聞與內部研究文件資料來源，支援投信分析師以自然語言完成市場查詢、事件追蹤與報告生成，整體研究效率提升 20% 至 30%
- 針對金融業內部客服 QA 高正確性與雲端推論成本問題，主導設計地端 LLM 行為對齊方案，建置 Gemma 3 12B 的 LoRA + DPO + GRPO 微調與評估流程，搭配高風險 QA benchmark、LLM-as-a-Judge 與人工抽查，使幻覺率相對下降 65%、回覆品質提升 21% (一作投稿 regtech award 獲獎)
- 使用千萬筆 Oracle 資料建立富邦信貸轉換預測模型，進行資料清理、特徵工程與模型訓練，將目標客戶辨識效率提升 3 倍
- 建置基於 Azure 的新聞檢索 MCP 工具，結合 SearxNG 與結構化解析，支援跨子公司使用
- 建置內部 LLM 部署與管理平台，提供 10 至 15 位同仁使用，具備模型路由、監控 Dashboard 與負載感知排程功能

研發工程師, Nexuni

新竹 2023/10 – 2024/06

- 參與邊緣 AI 應用開發，負責物件偵測與姿態估計模型訓練部署，並透過 TensorRT 部署於 Jetson Nano，達成 30+ FPS 即時推論效能
- 參與即時人流計數與車牌辨識數位孿生系統開發，整合電腦視覺模型、後端服務與邊緣裝置部署，落地新竹停車場

後端工程師實習, 富邦銀行

台北 2023/07 – 2023/09

- 協助交易室主導開發金融市場智能分析 Prototype，整合股價、新聞與財報資料，並導入 OpenAI + LangChain 架構，支援金融事件查詢、資料彙整與初步分析

專案經歷

Vinabot: 越南市場研究 Agent (富邦金控)

- 建置 LLM 驅動研究 Agent 系統，整合股票、總經、新聞與文件解析等多來源資料，讓投信分析師可透過自然語言完成查詢、趨勢分析與報告生成
- 設計自動化市場監控流程，每日蒐集股票新聞與價格資訊，並結合異常偵測與新聞分析自動產生事件報告
- 開發 RAG + Agent 架構 (FastAPI / Next.js / PostgreSQL / vLLM)，支援多工具協作與端到端分析流程
- 支援 PDF、Excel 與研究報告等多模態資料處理與查詢，大幅降低人工整理數據資料成本。提升投信分析師研究資料整理與報告產出效率 20% 至 30%。

金融業內部客服 QA 地端 LLM 對齊-降低幻覺 (富邦金控)

- 針對金融業內部客服 QA 高正確性需求，設計地端 LLM 行為對齊與微調方案，以降低長期依賴 GPT-4o 的推論成本，並提升客服知識問答在高風險場景下的可靠性
- 以 Gemma 3 12B 為 base model，建置 LoRA + DPO + GRPO 微調流程，將幻覺問題定義為不確定性下的決策問題，訓練模型在證據不足時拒答、在有依據時正確回答。
- 在內部 benchmark 上，最終幻覺率相對下降 65%，回覆品質提升 21%，同時保持通用能力

學歷

資訊管理與財務金融學系碩士

2022/09 – 2024/06

國立陽明交通大學

碩士論文：開發基於 Transformer 的交易模型，將多維技術指標轉換為影像表示，利用注意力機制捕捉市場動態並提升交易績效

財務金融學系

2018/09 – 2022/06

輔仁大學

競賽與研究經歷

2026 RegTech Initiative Award- Taiwan-Asian Banking & Financial Fintech Awards

基於行為對齊強化學習的 LLM 幻覺抑制研究 - 一作

- 將 LLM 幻覺重新定義為不確定性下的決策問題，學習何時回答、何時拒答與如何表達信心
- 設計 Behavior-Oriented Reinforcement Learning Framework (DPO + GRPO)
- 以 Gemma 3 12B 作為 base model，微調後相較未微調版本，幻覺率相對下降 65%、回覆品質提升 21%、通用能力維持不變

2023 中信 AI 智慧創新應用競賽—冠軍

- 開發金融市場事件分析系統，整合股市、新聞與 Wikipedia 等異質資料來源
- 建立多來源資料整合流程，包含網頁資料、Yahoo Finance 與財報資料庫
- 利用 OpenAI LLM 分析金融事件、股價與新聞間的潛在因果關係
- 建立金融 Dashboard 與股價視覺化工具，提升分析與決策效率

AI CUP 2023 玉山銀行信用卡盜刷偵測競賽

- 於 394 隊中取得 Private Leaderboard 前 5 名
- 使用 XGBoost 與 LightGBM 建立 Ensemble 模型，進行大量 Feature Engineering，包含交易行為、卡片統計與歷史詐欺特徵
- 提出 Seen / Unseen Cards 分群建模策略，導入 historical fraud flag 特徵，將 F1-score 提升至 0.85+、Precision 提升至 0.90+

SIDE PROJECT | 個人開源 AI 系統開發

GitHub: github.com/LLMSystems

- 建立完整 LLM 應用技術棧，涵蓋 RAG、AI Agent、文件解析、模型推論、路由與自動化評估系統
- 開發多用戶 Web Browse Agent、file2md、Tiny-RAGFlow、LLM 路由服務、TensorRT 服務與 LLM 評估等模組，實作從資料處理到推論部署與評估的端到端 GenAI 系統能力

語言能力

英文：TOEIC 860

日文：JLPT N2

證照

CFA (特許金融分析師) Level I

2023/01

CFA Institute

Deep Learning Specialization

2025/05

DeepLearning.AI

Machine Learning in Production

2025/05

DeepLearning.AI

Generative AI with Large Language Models

2025/05

DeepLearning.AI