



謝尚霖

Data Science

持續學習

跨領域

AWS

AI

“「資料科學家 / AI 工程師 | 具金融 + 物理 + 資料科學背景 | 熟 Python / LLM / RAG / DEVOPS / ML / DL」”

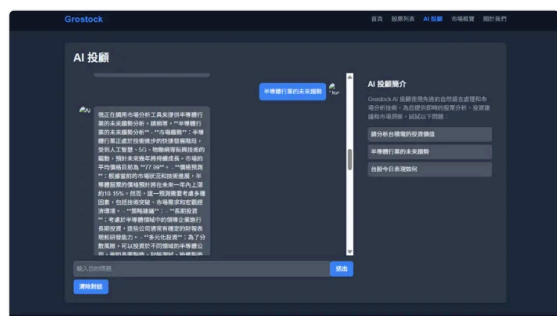
輔仁大學 | 金融與國際企業學系金融研究所 碩士畢業

希望職稱：資料科學家、資料分析師、AI 工程師

我具備金融背景與資料科學實作經驗，熟悉 Python、SQL、機器學習與深度學習，曾開發投資分析平台與 LLM 應用。參與大數據訓練計畫，掌握 ETL、RAG、MLflow、Docker、AI Agent 等技術，持續深化 MLOps 與 AI 工程 實戰，期望以 資料驅動 創造商業價值。

個人資料 男、30歲、役畢(2018/4)
就業狀態 待業中
主要手機 0975-601-655
E-mail shanglin8471@gmail.com
通訊地址 新北市板橋區文化路***
英文姓名 Shang Lin, Hsieh
聯絡電話 (02)2253-5755
聯絡方式 0975601655
駕駛執照 普通重型機車駕照、普通小型車駕照

專案成就



Grostock — 智能財經分析與投資平台

2024/12~2025/5

角色：全端開發者 / AI工程師 / 資料科學家

情境 (Situation)

投資決策往往需要結合技術指標、財報數據與市場消息，傳統人工處理效率低且容易出錯。

任務 (Task)

設計一個自動化 AI 平台，整合金融資料，並提供可解釋的投資建議。

行動 (Action)

- 建立 ETL 流程，自動擷取 MACD/RSI 指標，整合 PostgreSQL + MongoDB。
- 訓練 LSTM、GAN、Stacking 模型（時序預測），提升預測穩定性。
- 使用 Backtrader 進行策略回測，並整合 GraphRAG + Neo4j 構建知識圖譜。
- 建立多 Agent 協作架構，處理投資決策任務，並以 React + D3.js 製作互動儀表板。
- 部署於 FastAPI + Docker + K8s，利用 Dagster 排程與 AWS 雲端，確保系統高可用性。

成果 (Result)

- 每日處理 **1000+ 策略數據**，自動化率達 **80%**，顯著減少人工處理。
- 預測模型準確率 **85%**，支援即時投資判斷。
- 提升決策一致性，讓投資流程透明化並降低錯誤風險。

[前往查看 >](#)

LegalDocGen — AI 律師相關文件生成系統

2025/9~仍在進行

角色：AI 工程師

情境 (Situation)

律師需反覆撰寫或修改標準合約，耗時且容易遺漏法規更新。

任務 (Task)

開發一個 AI 平台，協助律師自動生成與檢查法律文件，提升工作效率與正確性。

行動 (Action)

- 利用 BeautifulSoup + spaCy 進行法規抽取與命名實體識別。
- 使用 Chroma 向量資料庫與 NetworkX 知識圖譜，建構多 Agent 架構。
- 導入 GraphRAG 提升文件檢索準確度，並在 RTX 4060 GPU 上以 TinyLLaMA + LoRA 微調。
- 整合 MLflow + GitHub Actions，建構 CI/CD 自動化流程。
- 前端以 Streamlit + Plotly 提供互動式回饋介面。

成果 (Result)

- 法律文件生成時間從 **數小時縮短至 15 分鐘內**。
- 文件生成 BLEU 分數提升 **20%**，更貼近律師書寫習慣。
- 初步驗證 AI 法務助理可應用於 **中小型律所**，具備商業化潛力。



美股漲跌預測平台

2024/5~2024/7

角色：資料科學家

情境 (Situation)

投資人需要一個多模型的股價預測平台，避免單一模型失效風險。

任務 (Task)

建立 GPU 加速的多模型預測平台，提升預測準確率與效率。

行動 (Action)

- 建立資料管道：用 Python 抓取 Yahoo Finance，存入 MySQL。
- 加入 SMA、MACD、RSI、Hurst 等技術指標，強化特徵工程。
- 訓練 RNN、SVM、MLP 並以 Stacking + Logistic Regression 做集成學習。
- 使用 Bayesian Optimization 自動調參，並以 cuML/cuDF GPU 加速訓練。

成果 (Result)

- 股價趨勢預測準確率達 **80%**，科技股預測率 **74%**。
- 訓練速度提升 **4 倍以上**，能快速更新模型。
- 成功部署為 **可視化投資決策工具**，支援實務操作。

情境 (Situation)

內部財務文件龐大，員工需要一個快速的 AI 問答助理，減少人工查詢時間。

任務 (Task)

建置可檢索多格式文件 (PDF, FAQ)，並提供即時問答的 AI 系統。

行動 (Action)

- 使用 FAISS + LangChain + FastAPI 建立知識檢索管道。
- 整合 PDF OCR，自動解析財務報表與 FAQ。
- 建立多 Agent 協作系統，分工處理文件檢索、答案生成、結果驗證。
- 前端以 Streamlit 提供文件預覽與查詢介面。

成果 (Result)

- 回答正確率提升 **30%**，平均回覆時間縮短 **25%**。
- 驗證 LLM 在企業內部知識檢索的應用可行性。
- 參賽「AI CUP 2024 玉山LLM挑戰賽」，展現落地潛力。

[前往查看 >](#)

NeuroQuery

2025/7~2025/9

情境 (Situation)

商業分析師在進行資料查詢時，經常需要 IT 支援撰寫 SQL，造成溝通延誤。

任務 (Task)

建立一個 LLM SQL 自動生成平台，讓使用者可用自然語言查詢資料。

行動 (Action)

- 部署 Ollama + Phi4 14B 模型，專注 SQL 語句生成。
- 建立語意解析模組，支援多表 JOIN、條件查詢。
- 前端使用 Streamlit 提供互動式輸入，並由 AI Agent 負責句法解析與結果解釋。

成果 (Result)

- SQL 生成準確率達 **88%**。
- 減少業務單位 **80% 的查詢依賴 IT 部門**，顯著提升決策效率。
- 展示 LLM 在 BI 報表與即時查詢上的應用價值。

[前往查看 >](#)

角色：AI 工程師/資料科學家

情境 (Situation)

公司/客戶需要快速掌握大量中文新聞重點，但來源格式雜、含簡繁混用，人工整理耗時且難以持續評估品質。

任務 (Task)

打造一套**給非技術人也能用的**「新聞分類+摘要」系統：自動處理資料、產生主題與摘要，並具備**可重現的量化評估**與後續模型迭代能力。

行動 (Action)

- 規劃並實作**8 階段 Pipeline**：資料夾重整 → 簡轉繁 (S2T) → 文字清理 → 統計 → **摘要生成** → **BERTScore/ROUGE 評估** → 資料集切分 → (可選) 訓練。
- 建立 models/llm_model.py **模型抽象層**，同一介面支援**雲端 LLM**與**本地 Transformers**，以參數化 CLI (--src/--dst/--model/--split) 一鍵切換。
- 以 **Kaggle—Fudan Categorised News** 為示範語料，制定**資料夾與實驗規範** (raw → clean → summaries → dataset)，輸出統計報表與評估結果，利於版本化與比較。
- 完成對外文件 (README) 與 **MIT License**，確保外部讀者能快速落地並遵循授權規範。

成果 (Result)

- 建立**端到端自動化**流程，非技術使用者可將新聞投入即取得**主題標籤+精簡摘要**；技術人員可依評估報表快速迭代模型。
- **評估標準化**：以 BERTScore/ROUGE 建立可重現的基準；分類任務預留 Accuracy/F1 介面，方便後續擴充。
- **導入成本降低、上手更快**：新成員可依 CLI 與資料規範在短時間內復現結果並比較多組實驗設定。

[前往查看 >](#)

專長

程式語言與開發基礎

- Python (主力)：熟練用於資料處理、建模、API 開發與自動化腳本。
- Java：具備物件導向與演算法基礎，理解後端服務架構。
- JavaScript (含 React)：掌握 React 框架撰寫互動式前端頁面，串接後端資料動態顯示。
- HTML / CSS：能設計基本介面排版與樣式設定。

#Python #Java #CSS #HTML #ReactJS

機器學習與深度學習

- Scikit-learn / XGBoost / LightGBM：實作分類、回歸、集成模型與超參優化。
- PyTorch：實作 CNN、RNN、Transformer 模型，用於 NLP 與時序預測。
- 模型融合與優化：使用 Stacking、Bagging、Boosting 與貝葉斯優化提升效能。
- GPU 加速技術：熟悉 cuDF、cuML、cuPY 等 NVIDIA 工具加速模型訓練。
- LLM 應用：整合 HuggingFace、RAG、LangChain、PhiData 架構，開發 AI 法律助手與金融投資推薦系統。

#機器學習 #深度學習 #scikit-learn #LLM #NLP #AI #人工智慧 #PyTorch

MLOps 與資料 workflow 管理

- Dagster：設計資料處理與模型訓練 workflow，支援排程與容錯。
- MLflow：進行模型版本控管、實驗記錄與部署流程追蹤。
- ELK (Elasticsearch, Logstash, Kibana)：監控資料處理流程與應用端狀態。
- Grafana + Prometheus：實作效能監控與異常告警系統，應用於 AI 模型部署監控。
- Pydantic + Argparse：撰寫模組化 CLI 與參數化任務流程。

#MLOps

網頁開發與應用部署

- FastAPI：熟悉 RESTful API 設計、參數驗證與模型服務部署。
- Streamlit / HTML / CSS：開發互動式 AI 應用前端，結合模型與資料展示。
- React.js：撰寫前端表單與互動模組，配合後端資料動態更新。
- Git / GitHub / GitLab：熟練分支管理、合併、CI/CD 自動化流程。
- Docker / Kubernetes (K8s)：進行容器化部署與自動化調度，應用於多服務協作與模型上線。

#JavaScript #HTML #CSS #Github #Git #ReactJS

資料處理與資料視覺化

- SQL (PostgreSQL / MySQL)：熟悉資料清洗、索引優化、Window Function、CTE 等進階應用。
- MongoDB (NoSQL)：能處理非結構化資料，撰寫 Aggregation Pipeline 進行即時分析。
- ETL 流程設計：規劃資料收集、轉換與儲存流程，應用於模型前處理與商業報表製作。
- 資料視覺化：熟悉使用 Power BI、Tableau、Matplotlib、Plotly 等工具進行可視化分析與 Dashboard 建置。

#ETL #PostgreSQL #MySQL #RDBMS #資料庫程式設計 #Power BI #Tableau

大數據與分散式系統處理

- Hadoop / Hive / Spark：掌握 HDFS、MapReduce 與 Spark SQL，應對大規模資料分析與 ETL 任務。
- 資料儲存與調度架構設計：整合 MongoDB / PostgreSQL + Spark 作為資料處理管線主架構。
- 應用場景經驗：如 AI 投資平台、法律文書生成系統，實作完整 Data Pipeline + 模型訓練 + Web API + 前端互動界面。

#hadoop #Hive #資料庫程式設計


資展國際(前資策會教研所) -Big Data 巨量資料分析班(進修)

2024/2~2024/7

- 系統性學習資料科學與 AI 技術，包含：
 - 程式語言與資料處理：Python、Java、SQL、Pandas、Numpy
 - 資料工程與平台：Hadoop、Spark、MongoDB、PostgreSQL、PowerBI
 - 機器學習 / 深度學習：Scikit-learn、TensorFlow、PyTorch、Keras、LLM 應用實作
 - 網頁與應用開發：HTML、JavaScript、爬蟲技術 (requests, BeautifulSoup, Selenium)
 - API 與部署：FastAPI、Flask、Docker、簡易後端整合與服務部署
- 專題實作：「美股預測平台」 | 結合 LSTM、SVC、MLP 進行多模型預測與資料視覺化呈現
- 每週進行實作演練與技術檢討，強化跨模組整合與專案開發能力

工作經驗

總年資 4~5年工作經歷



風險管理人員
聯邦商業銀行股份有限公司 (銀行業 500人以上)
風險管理人員 | 台北市松山區

2023/11~2024/2
4個月

- 1.每日監控銀行借貸金額，確保操作符合風險控管規範，並執行數據檢核與報表產出。
- 2.分銀行層級執行信用風險分析，整合多部門業務資料以支持策略制定。
- 3.協助測試與優化機器學習風險預測模型 (含模型效能監控)，提出 retrain 與調參建議，確保模型穩定與準確度。
- 4.與資訊部門合作，設計跨部門資料庫邏輯架構，提升查詢效率與正確性。

#風險管理 #Machine Learning #金融風險分析 #SQL #模型驗證 #跨部門協作

輔

教學助理

輔仁大學金融國企系（大專校院教育事業）

助教 | 新北市泰山區

2022/9~2023/11

1年3個月

- 1.設計課堂問卷並分析學生反饋，優化教學設計流程。
- 2.規劃機器學習競賽，包含資料清理、評分邏輯與實施細節，提升學生實戰經驗。
- 3.管理金融科技實驗室，負責硬體設備維護與軟體環境部署。
- 4.擔任營隊講師，教授 XGBoost 應用於 P2P 信用風險預測，兼顧理論講解與實務演練。
- 5.協助學弟妹進行競賽輔導，包含數據前處理、模型選型與技術指導。
- 6.協助教師專案與課務行政，確保課程與研究任務如期執行。

#Machine Learning #Python #資料分析 #問卷設計 #TA 經驗

廣

倉儲管理與財務管理

廣明藥局（其他醫療保健服務業）

藥師助理 | 新北市板橋區

2018/4~2021/8

3年5個月

- 1.負責藥品進銷存管理，整理每日庫存數據與調貨明細。
- 2.協助客戶諮詢與 LINE 官方帳號經營。
- 3.協助藥局每日帳務與月底結算流程。

#庫存管理 #家族企業 #行政管理

中

研究助理

中央研究院物理研究所（自然科學研發業）

研究助理 | 台北市南港區

2014/1~2017/11

3年11個月

- 1.參與有機朗肯循環（ORC）發電機實驗，負責設備設計、數據收集與實驗分析。
- 2.與博士後研究員及跨團隊密切合作，執行實驗進度與數據校正。
- 3.運用資料分析方法驗證實驗模型與假設，提升研究成果可靠性。
- 4.支援研究報告撰寫與實驗室設備維護。

#數據分析 #ORC #實驗設計 #研究支援 #資料驗證 #跨團隊協作

學歷

輔仁大學

金融與國際企業學系金融研究所 | 碩士畢業

2021/9~2023/6

東吳大學

物理學系 | 大學畢業

2013/9~2017/6

求職條件

希望性質 全職工作

上班時段 日班

可上班日 錄取後隨時可上班

希望待遇 面議

希望地點 台北市、新北市

遠端工作 對遠端工作有意願

希望職稱 資料科學家、資料分析師、AI 工程師

希望職類 資料科學家、AI工程師

希望產業	軟體及網路相關業、電信及通訊相關業、金融機構及其相關業、投資理財相關業、其他專業／科學及技術業
工作內容	我期望進入結合 LLM 應用與資料科學的團隊，擔任 AI 工程師或資料科學家職位。近年專注於大型語言模型的應用開發，包含 RAG、GraphRAG 架構設計、語意檢索優化與 AI Agent 任務鏈建構，並具備實務開發經驗。希望未來工作能讓我參與生成式 AI 系統整合，提升問答準確性與商業落地性，並深入 MLOps 與 API 工程領域。期望在實作中累積技術深度，也有機會參與產品決策或解決真實痛點。

自傳

- 我具備物理與金融背景，善於以邏輯推理與跨領域整合解決問題。在中央研究院參與有機朗肯循環實驗研究時，負責資料記錄、系統設計與數據分析，透過系統化設計與驗證，成功提升實驗數據的一致性與解讀效率，也培養了我對資料處理的敏銳度。
- 在輔仁大學金融碩士期間，我擔任大數據與機器學習課程助教，規劃課堂競賽並主講 XGBoost 與信用風險預測模型，指導學生完成 P2P 違約預測專案。這段經驗不僅加深了我對技術的掌握，也鍛鍊了簡報與跨部門溝通的能力。
- 同時，我協助家族藥局進行數據化營運，導入庫存與帳務分析並經營 LINE 官方帳號，優化了日常營運效率並提升顧客互動。
- 畢業後進入聯邦銀行總行擔任風險管理人員，除了承擔信用風險監控與報表製作，我也與資訊部合作優化 retrain 條件與資料流程，讓模型追蹤更精準並支持即時決策，更體會到數據在商業應用中的價值。
- 為深化工程與系統整合能力，我參與資展國際資料科學計畫，完整學習 JAVA、Flask、SQL、Python、資料視覺化、深度學習、Hadoop 叢集、網頁與 AI 應用等課程，並在結訓專案中主導開發美股預測平台，整合 RNN、SVM、MLP 並運用 Stacking 與 GPU 加速，大幅提升模型效能與訓練效率。這段歷程讓我同時具備紮實的工具知識與完整的專案實戰能力。
- 目前我持續進行 AI 投資分析平台與法律文書生成系統等AI專案，並透過自學與技術社群精進 LLM、RAG、MLOps、AI Agent、AI 工程與系統部署能力。我期許自己能結合理論與實務，在資料科學家或 AI 工程師職位中發揮資料驅動的價值。

- I have a background in Physics and Finance, with strong skills in logical reasoning and cross-disciplinary problem-solving. At Academia Sinica, I participated in an Organic Rankine Cycle (ORC) research project, where I was responsible for data recording, system design, and data analysis. Through systematic design and validation, I improved the consistency and interpretability of experimental results, which also sharpened my sensitivity to data processing.
- During my master’s studies in Finance at Fu Jen Catholic University, I served as a teaching assistant for Big Data and Machine Learning courses, where I designed class competitions, lectured on XGBoost and credit risk prediction models, and guided students in developing P2P default risk prediction projects. This experience deepened my technical expertise while enhancing my presentation and cross-functional communication skills.
- In addition, I assisted my family’s pharmacy with digital operations, introducing inventory and accounting analysis as well as managing its official LINE account. These efforts optimized daily operations and strengthened customer engagement.
- After graduation, I joined the Head Office of Union Bank as a Risk Management Analyst. In addition to credit risk monitoring and reporting, I collaborated with the IT department to optimize retraining conditions and data processes, making risk model tracking more accurate and enabling more timely decision-making. This experience further reinforced my understanding of the critical role of data in business applications.
- To strengthen my engineering and system integration capabilities, I completed the Institute for Information Industry’s Data Science Program, where I received comprehensive training in Java, Flask, SQL, Python, data visualization, deep learning, Hadoop clusters, web development, and AI applications. In the capstone project, I led the development of a U.S. stock prediction platform, integrating RNN, SVM, and MLP models with a Stacking framework and GPU acceleration, significantly improving both model performance and training efficiency. This experience provided me with solid technical knowledge and full project implementation skills.
- Currently, I am continuing to develop AI projects such as an AI Investment Analysis Platform and a Legal Document Generation System, while also advancing my expertise in LLMs, RAG, MLOps, AI Agents, AI engineering, and system deployment through self-learning and participation in tech communities. I aspire to combine theory with practice and contribute to creating data-driven value as a Data Scientist or AI Engineer.

語言能力

英文

聽/中等 | 說/中等 | 讀/中等 | 寫/中等

TOEIC (多益測驗) 660

證照


金融投顧相關證照


信託業業務人員信託業務專業測驗(信託法規乙科)

其他證照


資展國際(前資策會教研所) -Big Data 巨量資料分析班

ShangLin1606/
portfolio




 1


Contributor

 0


Issues

 0

Stars

 0

Forks



github.com/ShangLin1606/portfolio

資料科學作品集