

黃偉嘉



未來市股份有限公司 | Senior Data Engineer / MLOps Engineer

國立交通大學 | 資訊管理研究所 碩士畢業

新北市新店區 | 5~6年工作經歷 | 希望職稱：data engineer, MLOps engineer, data scientist, machine learning engineer

個人資料 男、30歲、役畢(2020/1)
就業狀態 在職中
主要手機 0988-925-558
E-mail tyc30827@gmail.com
通訊地址 新北市新店區德正街***
英文姓名 Willie Huang
聯絡方式 tyc30827@gmail.com, 手機連絡，10:00-22:00 (盡量19:00後)
駕駛執照 普通重型機車駕照、普通小型車駕照
交通工具 普通重型機車

學歷

國立交通大學	2017/9~2019/8
資訊管理研究所 碩士畢業	
國立成功大學	2013/9~2017/6
工業與資訊管理學系 大學畢業	

工作經驗

總年資 5~6年工作經歷

Senior Data Engineer / MLOps Engineer
未來市股份有限公司（電腦軟體服務業 30~100人）
資料工程師 | 台北市中山區

2024/8~仍在職

技術專長 (Tech Stack)：

- 主要程式語言：Python
- 資料平台：Dagster、Airbyte、dbt (SQL)、BigQuery
- 資料儲存：MongoDB、AWS S3
- 資料視覺化：Looker Studio
- 資料串流與佇列：Apache Kafka
- 後端開發：Flask
- LLM / RAG：AWS Bedrock、PydanticAI、Llamaindex、Huggingface、AWS Bedrock Knowledge Bases
- 基礎設施：Terraform、uv、poetry
- 警示系統：Slack

主要工作內容：

- 獨立完成社群平台內容資料標記系統 (AI-labeling system) 設計與開發，整合 LLM (AWS Bedrock、PydanticAI)、Flask、Kafka、Dagster，打造完整資料處理流程：
 - 設計兩階段 prompt 流程，透過 LLM (Claude Sonnet 3.5, Meta LLaMA 4) 自動化標記 MongoDB 資料庫中有關短影音、心情貼文的內容
 - 建立 Flask server，實作 API endpoints 供資料標記與Dagster作為客戶端呼叫後續推論使用
 - 將標記結果透過 Kafka 傳遞至推薦系統使用，並建置 Slack 警示機制監控系統狀況
 - 標記結果已正式應用於公司產品推薦系統以及搜尋引擎中，大幅提升使用者體驗
- 設計與實作Multi-modal RAG系統
 - 研究並應用最新多模態RAG論文，使用Llamaindex搭配Huggingface或者大廠API實作系統
 - 應用Gemini 2.5系列大模型為基底開發baseline multi-modal RAG系統
 - 應用AWS Bedrock Knowledge Bases開發end-to-end RAG系統
 - 開發給團隊用的RAG system實驗環境與Codebase
 - 透過MLflow紀錄GenAI應用的開發過程與追蹤生命週期
 - 提升公司AI Agent產品的檢索能力，幫助純文字RAG提升至可以檢索圖片、PDF等內容的multi-modal RAG
- 設計與實作Structured data RAG系統
 - 使用DuckDB搭配Gemini-2.5-flash-lite，設計出檢索時間與花費成本極有競爭力且使用者上傳檔案即可立刻使用的系統架構
 - 自動推論結構化資料的結構，自動生成對應SQL，自動生成Insight
- 負責資料平台架構優化，系統設計並維運 Airbyte (結合Terraform管理)、dbt、Dagster、BigQuery等基礎設施，推動資料處理流程自動化與標準化。
- 擁有多平台資料排程經驗，熟悉 Apache Airflow 與 Dagster，根據專案需求設計穩定且可擴展的 Data Pipeline。
- 主導公司雲端資料倉儲Google BigQuery系統資料分析設計與優化，對資料分級、提高查詢效率、降低90%的查詢開銷。
- 擔任 BI codebase 的 Code Owner，負責制定技術規範、Code Review 與品質控管，協助團隊維持穩定開發流程。
- 帶領 4 位 Junior 工程師熟悉資料工程基礎觀念、團隊技術文化與系統設計，提升團隊整體技術水準。
- 規劃團隊開發環境升級，推動虛擬環境從 venv 遷移至 poetry，再優化為 uv，提升開發效率與環境一致性。

- 舉辦過2次技術分享活動，向全公司工程團隊介紹 AWS Bedrock 實務應用與介紹資料團隊技術架構，推動跨團隊技術交流與透明度。
- 主講多次工程團隊的各類RAG的分享會，分享商業化技術與前沿研究。
- 熟悉敏捷開發流程，日常透過 Scrum、Jira、Confluence 管理專案與任務，並參與 RFC 制定與技術討論，推動團隊技術決策透明化與制度化。

Senior data engineer

2023/10~2024/8

訊連科技股份有限公司（電腦軟體服務業 500人以上）

11個月

資料工程師 | 新北市新店區

Text mining for customer feedback data

- 建立NLP model (LDA)來對主題進行建模，試圖從原始的使用者feedback data找出潛在的主題或者關鍵字
- 把預測結果以dashboard的方式呈現在Tableau上

Data pipeline used for collecting total mobile install data from various datasource

- 為monetization team搜集多個資料源的mobile app安裝次數資訊
- 資料源有: google play, google ads, app store connect, apple search ads
- 以Airflow每日定期執行撈取任務，分別以下列兩種方法達成
 - 透過selenium網路爬蟲自動化通過驗證去撈取google play, app store connect的資料
 - 透過Airbyte去sync google ads與apple ads API所提供的安裝資訊
- 將資料自動化擷取至Tableau datasource，供內部同仁創建相關dashboard

Data migration of existed projects handled by our data team

- 汰換從MSSQL撈取資料到AWS S3及Athena table的方法
 - 從sqoop這種在hadoop與relational DB之間傳輸的方法改成Airbyte
- 以Airflow自動化執行sync task

#ETL #Machine Learning #AWS #Python #Tableau #Airflow

Data Scientist and Engineer

訊連科技股份有限公司（電腦軟體服務業 500人以上）

軟體工程師 | 新北市新店區

1. 工作內容

- Data Engineering :

(1) Data pipeline，透過Airflow處理上千萬筆/GB等級以上的資料之ETL/ELT

(2) Data warehouse，對big data進行接近real-time分析

(3) 透過Superset, Tableau等視覺化工具對資料視覺化

- Machine Learning, MLOps :

(1) 範本內容檢索系統 (Information retrieval from templates)

對範本內容進行自動標註，幫助使用者根據關鍵字快速探索與查找感興趣的圖像/影片內容

極大幅改善點擊率(CTR)：從不到1%變成30-60%

(2) 個人化推薦系統 (Data product)

從0-1開發公司產品 - "Promo" 的「為你推薦」功能

透過產品互動data訓練deep learning based recommendation system，對訂閱者進行個人化推薦

以MLOps的形式進行開發，透過AutoML自動化訓練、調整參數與模型

表現優於rule-based baseline至少11.2%

經過A/B Test證明模型對轉化率影響通過統計檢定

(3) 基於圖神經網路(GNN-based)的user preference system

透過訂閱者的互動、個人資料建立relation network

以圖神經網路預測訂閱者會感興趣的內容與潛在的訂閱活動

2. Toolkit and Service

Apache Airflow, Apache Kylin, Databricks

Apache Superset, Tableau,

Machine Learning (Recommendation / Computer Vision / Time-series) :

MLflow (Performance tracking, Machine learning life cycle handling)

AWS Service :

S3, Athena, EC2, Neptune, Kinesis,

SageMaker, Rekognition, Personalize

API Gateway, Lambda, SES, and so on)

Python, SQL, PySpark, Gremlin (graph), Java, C++/C

Git / Linux / Docker

#Python #AWS #Apache Airflow #Deep Learning #SQL #Data engineering

求職條件

希望性質 全職工作

上班時段 日班、可配合輪班

可上班日 錄取後一個月可上班

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 data engineer, MLOps engineer, data scientist, machine learning engineer

希望職類 軟體工程師、軟體專案主管、其他資訊專業人員、演算法工程師

希望產業 金融機構及其相關業、電子資訊／軟體／半導體相關業

工作內容 data engineering
machine learning
data science
MLOps
Domain: recommendation system, time-series data, computer vision

語言能力

英文

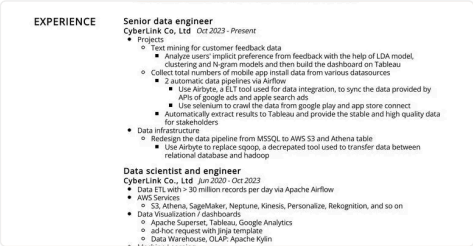
聽/精通 說/精通 讀/精通 寫/精通
TOEIC (多益測驗) 845

台語
精通

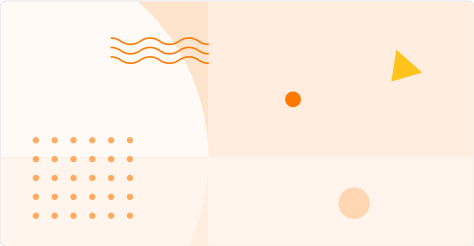
附件



TOEIC_845.jpeg
多益845



Resume_WillieHuang.pdf
English Resume



AI領域頂級期刊第一作者

專長

擅長工具

- 程式語言：Python, SQL
- Database: PostgreSQL, MySQL, SQL Server
- NoSQL: S3/MinIO, MongoDB, BigQuery, DuckDB, Athena
- Data engineering: Airflow, Dagster, Airbyte, dbt (SQL), Apache kafka
- Data analytics: GA4, Looker Studio, Superset, Tableau
- Backend: Flask, Pydantic
- IaC: Terraform, Docker, Git, GitOps/ArgoCD, CICD
- AWS Services: Personalize, Rekognition, Neptune, EC2, Lambda, Kinesis
- GenAI / RAG / MLOps: AWS Bedrock, PydanticAI, LlamaIndex, MLflow, Databricks
- Machine Learning: GNN, Recommendation system, Financial Time-Series, PyTorch, Keras, SageMaker, Huggingface
- Package dependencies management: uv / poetry

#Python #AWS #Machine Learning #Data Architect #Git

工作技能

- **數據管線編排與ELT架構設計**：設計高可靠性、自動化 Airflow/Dagster**資料管線**，使用dbt進行數據轉換與品質管理，具備數據模組化分層設計能力。
- **數據串流處理與分散式系統**：基於 Apache Kafka設計即時數據處理方案；透過BigQuery / DuckDB (本地開發) 等數據倉儲優化大數據查詢。
- **團隊技術領導與指導**：具備帶領初級工程師的能力，能進行工作分解、技術指導與知識傳承。作為商業智慧程式碼庫的**Owner與Reviewer**，專責維護程式碼品質、架構健壯性。
- **ML 模型生產化與生命週期管理 (MLOps)**：MLflow 3.0進行實驗追蹤、版本控制與模型註冊；負責模型的API呼叫生命週期追蹤與監控。
- **LLM應用落地與RAG系統設計**：成功設計並部署基於LlamaIndex的**多模態RAG**以及基於PydanticAI**結構化資料RAG**解決方案；精通 AWS Bedrock / DuckDB服務整合，並透過PydanticAI確保LLM輸出的**結構化與準確性**。
- **後端API服務開發與微服務架構**：使用Flask搭配Pydantic設計健壯且有結構的 RESTful API服務；利用Copilot Agent等工具提升**單元測試**和**整合測試**效率。
- **基礎設施即程式碼IaC**：使用Terraform部署Airbyte，透過GitHub Action控制CI/CD workflow。
- **深度學習與進階模型實作**：具備 PyTorch/Keras實作GNN、推薦系統和財務時間序列模型的經驗，並負責模型的上線驗證。

#軟體程式設計 #Machine Learning #Python #AWS #系統架構規劃 #Cloud #MLOPs #AI #LLM #結構化程式設計

More About Me

對資料科學極度充滿熱情且充滿好奇心，嘗試透過各種data science tools對資料本身挖掘insight，擁抱新技術，極度渴望能在職涯中不斷自我學習，並打下扎實的基本功

CakeResume:

<https://www.cakeresume.com/s--4ib21SBPJWKArgqPRRypSw--/tyc30827>

Linkedin:

<https://www.linkedin.com/in/willie-huang-ba79a9199/>

證照

陸上運輸相關證照

普通重型機車駕照、普通小型車駕照

其他證照

TOEIC 845

專案成就

First author of paper in top-tier computer science journal - Information Fusion

Huang, W. C., Chen, C. T., Lee, C., Kuo, F. H., & Huang, S. H. (2022). Attentive gated graph sequence neural network-based time-series information fusion for financial trading. Information Fusion.

[前往查看 >](#)