

DATA SCIENCE PORTFOLIO

A I 驅動的跨領域 情報自動化系統

🕒 單次執行過濾 86% 無效雜訊，精準鎖定高價值商業情報

整合 **Google Gemini** 語意理解與堅韌爬蟲架構的
Python 全自動化解決方案

簡報人：王譽鈞

痛點與挑戰：資訊雜訊與資料清理的困境



資訊超載

每日面對數萬條非結構化數據，人工篩選耗時且易錯，難以捕捉轉瞬即逝的 Alpha 情報。



時間陷阱

來源網站大量使用 "2天前" 相對時間，導致歷史回測與長期趨勢分析出現嚴重偏差。



語意雜訊

關鍵字匹配缺乏語境（如 "Apple" 指水果還是科技？），傳統規則式爬蟲無法有效區分。

系統邏輯架構：端到端自動化流水線

Process Flow



工程技術亮點：對抗髒資料的三大防線

Python Engineering

Yahoo 時間校正

針對「2天前」模糊時間，實作 JSON-LD 結構化數據解析。

✓ 修正 166 則日期錯誤

🗑 剔除 104 則偽裝舊聞

Google RSS 代理

開發 RSS 代理模組，突破 ETtoday 等網站的站內搜尋頻率限制。

🛡 穩定獲取即時索引

🚫 0% IP Ban Rate

多執行緒平行架構

實作 4 Threads 平行架構，在有限運算資源下達成最佳 I/O 平衡。

⚡ 1,300+ 則 / 10 min

🧠 效能提升 500%

AI 賦能 I：趨勢獵人 (Trend Hunter)

Generative AI

從「靜態關鍵字」自動轉化為「動態商業場景」，擴展情報視野。

INPUT ENTITY

好市多 (Costco)



AI IDENTIFIED CONTEXT

2025 黑色星期五 (促銷季佈局)

INPUT ENTITY

00940 (ETF)



AI IDENTIFIED CONTEXT

成分股 輪動效應 (投資痛點)

INPUT ENTITY

輝達 (NVIDIA)



AI IDENTIFIED CONTEXT

GB200 供貨狀況 (供應鏈追蹤)

AI 賦能 II：嚴格守門員 (Strict Gatekeeper)

Semantic Filter

AI 成功攔截的真實雜訊案例，精準區分「字面符合」與「語意相關」。

搜尋關鍵字	爬取到的標題	AI 決策	商業洞察 / 攔截原因
OpenAI	「馬斯克談 Grok5 的未來展望...」	REJECT	🚫 競品動態 (非本體)
黃金價格	「陽明海運獲頒環保金級獎肯定」	REJECT	⚠️ 形容詞誤判
鴻海	「全球海運聯盟即將進行重組」	REJECT	📦 領域錯置 (航運業)
台積電	「台積電 CoWoS 產能擴充計畫提早」	ACCEPT	🎯 高價值情報

商業影響力：情報密度漏斗

Data Funnel



▼ 總雜訊過濾率達 **86%**，審閱時間 ~~4hr~~ → **15min**

🧪 測試場景：

科技趨勢(輝達/OpenAI/鴻海) x 品牌(星巴克/好市多/特斯拉) x 金融(BTC/黃金/00940)

📅 執行數據：

2025/11/26 (回推3日)

作品展示與 未來展望

1. 自動化報表 Deliverable

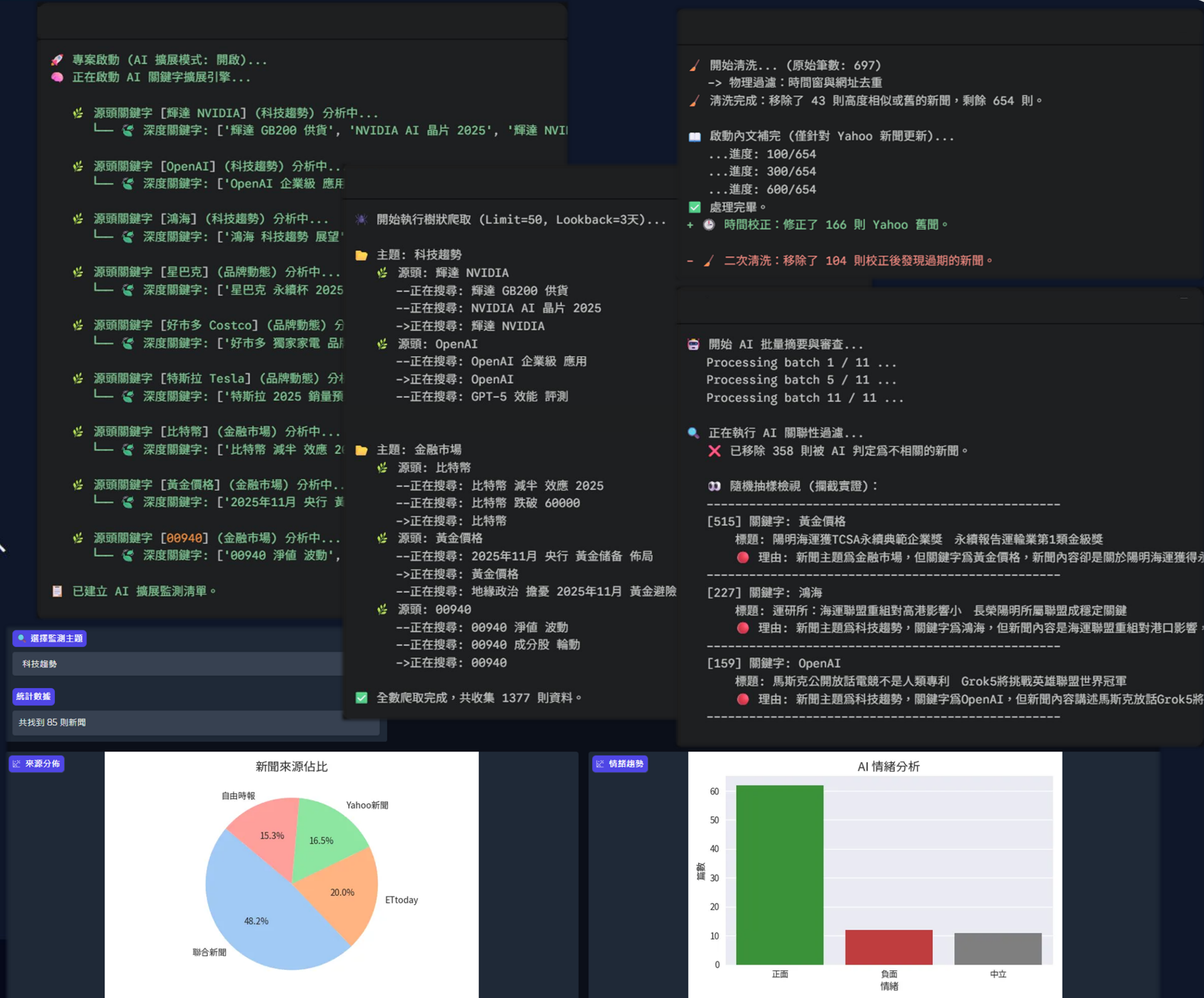
目前系統已能自動生成結構化 Excel 報表，包含來源、校正後日期、AI 摘要與信心評分。使用者可直接透過篩選器查看高價值情報。

2. Next Step: RAG 整合

未來將導入檢索增強生成 (RAG) 技術，建立企業專屬知識庫，允許使用者進行自然語言問答：

User: "上週特斯拉的主要競爭對手動態為何?"

Bot: "根據情報庫, BYD 發布了新的刀片電池技術 ..."



最新高價值情報 (Top 10)			
標題	時間	摘要	AI 評語
Edge AI商機現 展碁NVIDIA DGX Spark銷售看旺	2025-11-25 06:16	展碁國際引進輝達NVIDIA DGX Spark桌上型AI超級電腦，看好AI與Edge邊緣運算需求，預期將為公司挹注營收。	新聞主題為科技趨勢，關鍵字為輝達NVIDIA，新聞內容講述展碁國際引進輝達NVIDIA DGX Spark桌上型AI超級電腦，並預期銷售看旺，符合主題與關鍵字。
展碁搶輝達超級電腦商機	2025-11-26 00:17	展碁國際宣布引進輝達NVIDIA DGX Spark桌上型AI超級電腦，看好AI與邊緣運算需求，首批產品已開始出貨。	新聞主題為科技趨勢，關鍵字為輝達NVIDIA，新聞內容講述展碁國際引進輝達NVIDIA DGX Spark桌上型AI超級電腦，並預期銷售看旺，符合主題與關鍵字。
傳Google賣晶片「一度跌6%」輝達反發文恭喜：但我領先一個世代	2025-11-25 21:48	傳Google推出自家晶片導致輝達股價一度下跌6%，輝達發文回應表示其技術仍領先業界一個世代。	新聞主題為科技趨勢，關鍵字為輝達NVIDIA，新聞內容提及Google賣晶片導致輝達股價下跌，並引用輝達發文回應技術領先，符合主題與關鍵字。
AI霸主遭Google挑戰 輝達反擊「技術仍領先業界一代」	2025-11-26 02:35	Google被指挑戰輝達AI霸主地位，輝達回應表示其技術仍領先業界一代。	新聞主題為科技趨勢，關鍵字為輝達NVIDIA，新聞內容講述Google挑戰輝達AI霸主地位，輝達則反擊技術領先，符合主題與關鍵字。
輝達AI大客戶變心！Meta遭爆買Alphabet TPU 分散晶片來源	2025-11-25 04:14	輝達AI大客戶Meta被爆購入Alphabet的TPU晶片，以分散晶片來源，顯示客戶策略轉變。	新聞主題為科技趨勢，關鍵字為輝達NVIDIA，新聞內容講述輝達AI大客戶Meta遭爆購買Google TPU晶片以分散晶片來源，符合主題與關鍵字。