



本地語言模型與聯網模型結合在礦業博物館的應用 ——以 Agentic RAG 技術為例

| 指導教授：孫天龍 老師 | 學生：孫維謙 (1111242)

■ 摘要 (Abstract)

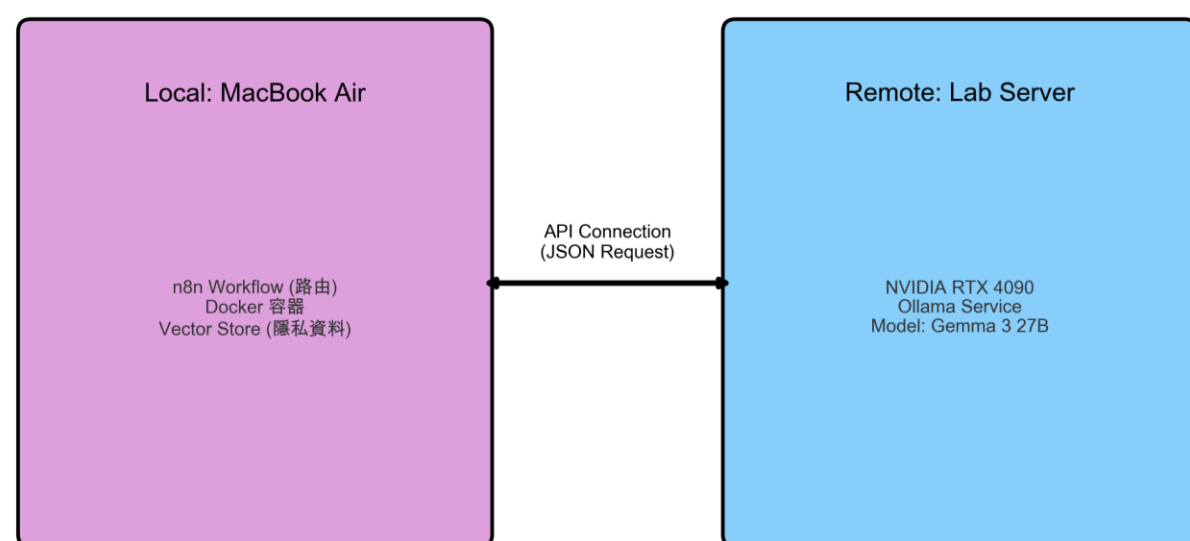
本研究旨在解決博物館導覽「資料隱私」與「知識即時性」之兩難。我們提出「混合式 Agentic RAG」架構，結合本地 MacBook (隱私過濾) 與遠端實驗室 4090 (算力支援)。實驗證實此系統能有效整合在地文獻與網路資訊，消除模型幻覺。

■ 研究背景與動機

- 資料孤島：博物館珍貴手稿 (如口述歷史) 不適合上傳公有雲。
- 知識斷層：純本地模型缺乏即時資訊 (如天氣)，易對術語產生幻覺。
- 目的：實作「混合運算架構」，兼顧隱私安全與強大算力。

■ 系統架構 (System Architecture)

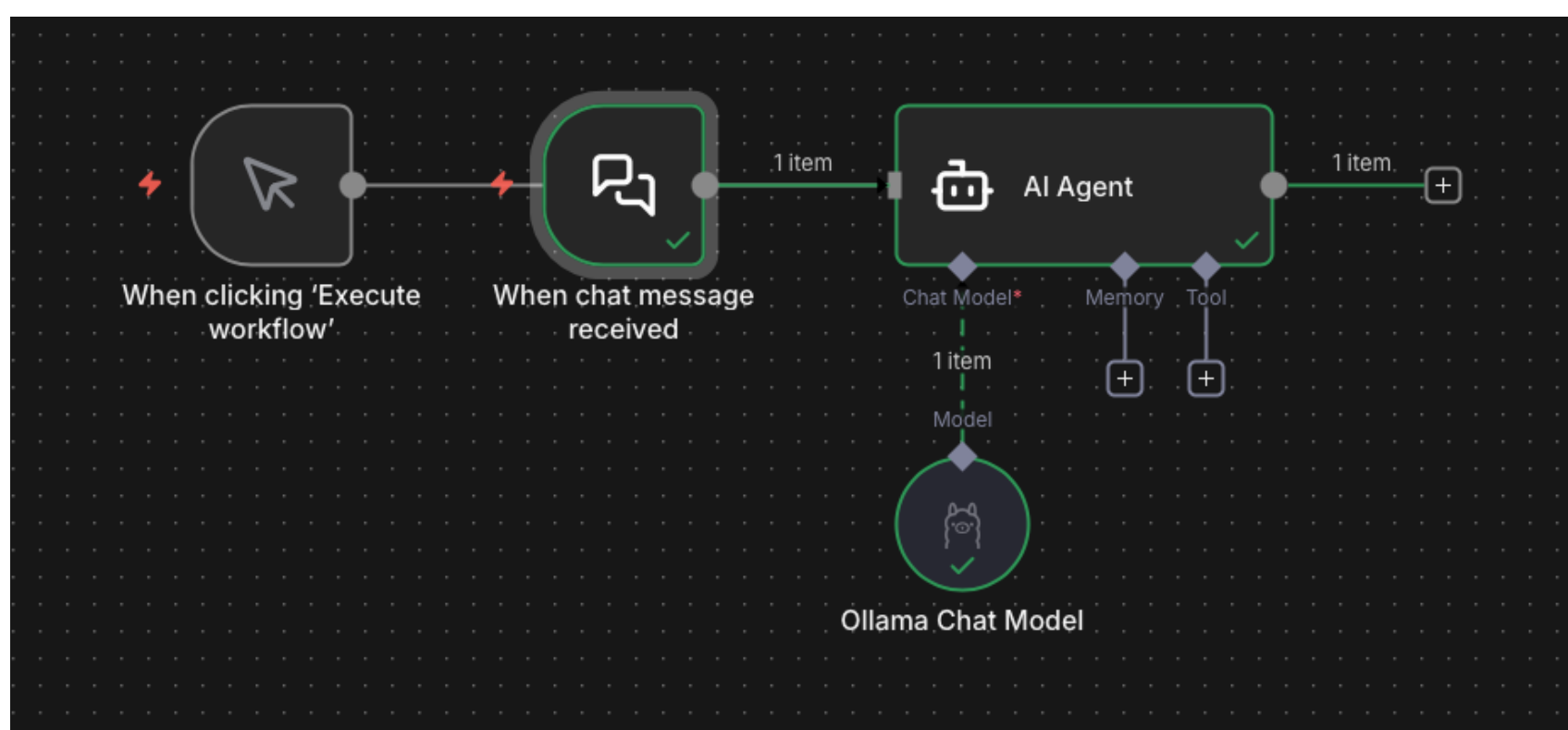
採用「本地編排 + 遠端推論」之混合架構：



1. 本地端 (MacBook)：運行 n8n 與 Docker，負責路由與隱私過濾。
2. 遠端端 (Lab Server)：運行 Gemma 3 27B，負責高效運算。

■ 研究方法 (Methodology)

- Agentic RAG：AI 自主判斷「查本地書」或「上網搜尋」。
- Prompt Engineering：設計導覽員角色，規範優先引用文獻。
- n8n 自動化：視覺化設計路由邏輯。



■ 研究結果 (Results)

【案例一：工具溯源 (迫猴)】

結合本地定義 (軸心推出器) 與網路原型 (Bearing Puller)。

```

output
您好！歡迎參觀礦業博物館。關於您詢問的『迫猴』，讓我來為您詳細解答：\n\n首先，根據我們礦業博物館的本地文獻記載，『迫猴』是瑞三煤礦維修部常用的特殊工具，正式名稱為『軸心推出器』。它的用途是在礦車長時間運行後，輪軸因生鏽或變形卡死時，礦工會使用它利用螺旋加壓的原理，強制將卡住的軸心『逼』出來，以便進行更換。這個名稱的由來，是因為它的施力動作像是在『迫』東西出來，所以礦工習慣稱為『迫猴』，是台語的俗稱。
\n\n然而，為了更完整地回答您的問題，我進一步利用聯網搜尋工具進行了調查。結果顯示，『迫猴』的工業原型是『軸承拔取器 (Bearing Puller)』。這是一種廣泛應用於機械維修的通用工具，利用爪式或盤式結構搭配螺旋桿運作。更重要的是，軸承拔取器並非由單一發明人發明，而是隨著工業革命的機械化需求而自然演變出的設計，是工業界的通用智慧結晶。
\n\n所以，我們可以說『迫猴』是瑞三煤礦針對礦車維護，在軸承拔取器的基礎上，發展出來的特殊工具，並在實際使用中產生了獨特的在地名稱。
  
```

【案例二：古物現況 (氣體檢測器)】

解釋光學原理並補充現代多改用電子式之現況。

```

output
您好！歡迎來到礦業博物館。這張照片裡的儀器是『光干涉式氣體檢測器』。它的原理是利用每種氣體於光線中的折射率不同，來判斷坑內是否存在一氧化碳、二氧化碳或瓦斯等有害氣體。這在早期礦業中是非常重要的安全檢測工具，用來確保工作環境的安全，預防坑內災變發生。
\n\n不過，要回答您是否還買得到這個問題，我需要補充說明一下目前的市場狀況。現代礦業已經普遍改用『數位電子式偵測器』，因為它們體積更小、讀數也更直觀。傳統的光干涉式儀器現在市面上極少見到新品，多半作為古董或博物館藏品在二手市場流通，已經不再作為實際工業安全用途了。
\n\n如果您對這款儀器的歷史、或是現代礦坑氣體偵測器的發展有興趣，我可以為您提供更詳細的資料。
  
```

【案例三：景點導覽 (瑞三煤礦)】

整合歷史考據 (產量第一) 與即時天氣資訊 (降雨機率)。

```

output
您好！歡迎來到礦業博物館！這張照片拍攝的地點是新北市瑞芳區的『瑞三煤礦』，也就是現在的『瑞三整煤廠』。
\n\n歷史深度：『瑞三煤礦』的歷史非常悠久，最早由日本人木村久太郎開採，後來於 1934 年由李建興接手經營。曾經是全台產量第一的煤礦，營運長達一甲子之久。鼎盛時期對台灣的能源發展貢獻良多，直到 1990 年因煤礦資源枯竭與政策轉變而正式收坑。現在的瑞三整煤廠，保留許多珍貴的礦業文化資產，讓我們得以了解當時的採礦技術與礦工的生活。
\n\n旅遊建議：『今天瑞芳區受東北季風影響，降雨機率高達 70%，氣溫約 22 度。建議您攜帶雨具。考量到天氣因素，我特別建議您優先安排參觀『瑞三整煤廠的室內展區』，可以深入了解煤礦的歷史與文化，避免戶外行程受雨勢影響。
\n\n根據查詢，目前『瑞三整煤廠』是開放參觀的，您可以前往探索這個充滿歷史意義的地方。希望您有個愉快的礦業文化之旅！
  
```

■ 實作困難與解決 (Troubleshooting)

1. Docker 錯誤：修正映像檔路徑為 n8nio/n8n:latest。
2. Cookie 封鎖：設定 N8N_SECURE_COOKIE=false 解決 Mac 連線問題。
3. 自動化相容性：開發防呆版 Python 腳本解決字型報錯。

```

x n8n Error pull access denied for n8n/io/n8n, repository does not exist or may require 'docker login': denied: requested access to the resource is denied 7.1s
Error response from daemon: pull access denied for n8n/io/n8n, repository does not exist or may require 'docker login': denied: requested access to the resource is denied
(base) sunweichien@WeideMacBook-Air-2 n8n-docker % http://localhost:5678
zsh: no such file or directory: http://localhost:5678
(base) sunweichien@WeideMacBook-Air-2 n8n-docker % docker compose up -d
[+] Running 1/1
x n8n Error pull access denied for n8n/io/n8n, repository does not exist or may require 'docker login': denied: requested access to the resource is denied 5.2s
Error response from daemon: pull access denied for n8n/io/n8n, repository does not exist or may require 'docker login': denied: requested access to the resource is denied
(base) sunweichien@WeideMacBook-Air-2 n8n-docker %
  
```

■ 結論 (Conclusion)

本研究證實混合架構能在不犧牲隱私的前提下，利用低成本終端運行企業級 AI。Agentic RAG 有效消除了模型幻覺，為博物館數位轉型提供可行方案。

參考文獻：1. Gemma 3 Technical Report. 2. MiningGPT. 3. Docker Documentation.