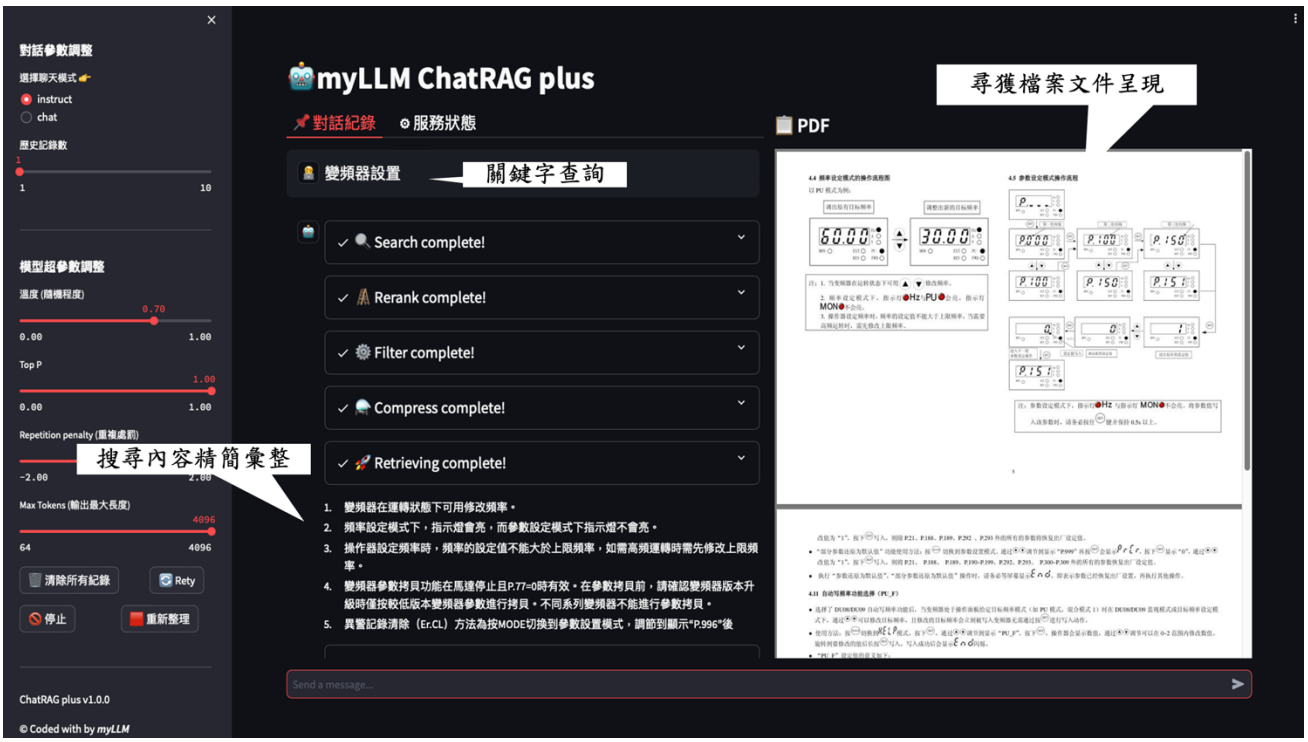
RAG plus：開創企業超強資料檢索和學習力

在當今數位轉型的浪潮中，企業正面臨著前所未有的數據和資訊管理挑戰。從技術文檔到機台說明書，海量的文本資料需要被有效地檢索、管理和利用。針對這一需求，我們公司推出了一套革命性的解決方案：進階強化式檢索生成（RAG plus）系統，專為企業打造的智慧文本管理中心。

一、進階強化檢索生成（RAG plus）

智慧企業文本管理中心

本公司開發之進階強化式檢索生成結合大型語言模型既快、狠又準，在高階個人電腦上只需透過少量關鍵字查詢即可於短時間內歷遍海量公司文件含技術文本、機台說明書等樣樣皆得心應手，並提供高度相符檢索成果、尋獲檔案文件呈現與內容精簡彙整，極為適合企業採納用作內部智慧文本管理中心。



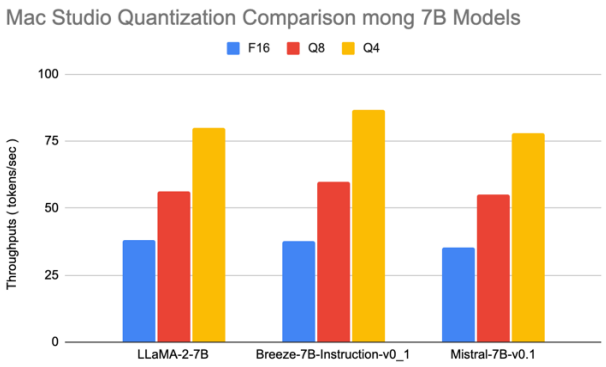
RAG plus 的無盡可能

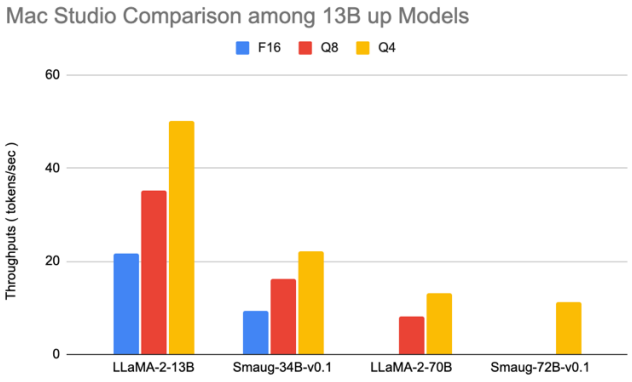
除企業文管中心外，智慧工具代理自動編寫程式碼、探索功能解決問題，到常見問題集助手和會議演講摘要整理等各種應用，本公司開發的各式智慧軟體功能將在企業和個人用戶中無所不能。

無論是面對工程師端的技術挑戰，客服端的客戶應對，亦或是內部開會演說的場合。只要擁有一台個人電腦，企業可以輕鬆地搭配本公司提供的高階軟體技術支援，從而無限運用於不同的職場環境中。無論是提高生產力、優化工作流程，還是創造更加智慧的工作環境，myLLM 都將成為企業的得力助手。

框架部署與實際測量

經多項框架部署與實際測量，MAC Studio 大容量隨機存取記憶體與強大的 M2 Ultra 晶片使其可運行各項大型語言模型(LLM)，包含聯發科技的 Breeze-7B、Meta 的 LLaMA-2 7B、13B、70B(Int8)、Smaug 34B、72B(Int4) 等都擁有不錯推論速度表現。





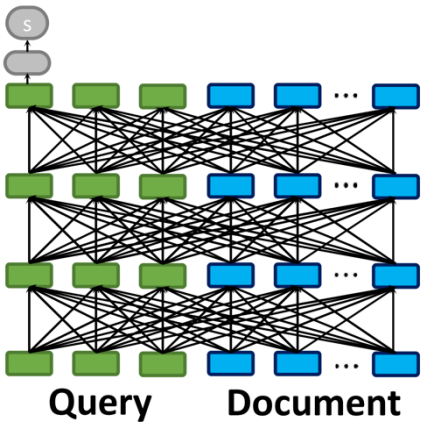
二、從 RAG 到 RAG plus

從 RAG 到 RAG plus 的進化不僅標誌著技術的進步，更象徵著智慧檢索系統在解決實際應用中遇到的挑戰上取得的突破。面對 RAG 框架在檢索精準度和文件雜訊處理上的不足，RAG plus 通過一系列創新的優化模組，大幅提升了檢索與生成的性能，為企業和個人用戶在處理大規

模私有資料時帶來了顯著的便利。

Reranker：提高檢索精準度

在傳統的 RAG 框架中，初步檢索出的 top-k 文件常常包含不少雜訊，影響了資料的使用效率。為解決這一問題，RAG plus 引入了基於 Bert 的 Reranker 模組。該模組對初步檢索結果進行再評分和排序，有效地篩選出更加相關的文件，從而顯著降低錯誤文件的干擾，提升了檢索的準確性。



Document Filter & Compresser：文件處理的新紀元

面對大量且複雜的文件資料，RAG plus 透過 Document Filter & Compresser 的創新設計，進一步提升了檢索的精確度。利用大型語言模型（LLM）作為技術核心，結合高效的推理平台，不僅實現了對大規模文

RAG plus 優化技術

1. 文件相關度精準排序：通過 Reranker 技術，確保檢索結果的相關性和準確性。
2. 複雜文件內容的過濾與壓縮：透過 Document Filter & Compresser，提高處理大規模數據的效率。

3. LLM 建立文件標籤：使用大型語言模型自動標註文件，便於快速檢索和管理。
4. 文件知識圖譜的建立：利用先進的算法，構建文件間的知識關聯，進一步豐富檢索系統的背景知識庫。

在這個數位轉型加速的時代，企業迫切需要一個能夠高效管理和利用龐大數據的系統。我們的 RAG ^{plus} 系統應運而生，它不僅提升了資料檢索和學習的效率，更透過進階技術大幅減少檢索時的雜訊，確保了訊息的精確性和可靠性。RAG ^{plus} 的創新之處在於其融合了最先端的大型語言模型和自主研發的優化模組，從而使企業能在短時間內精準地檢索到所需的文件和資訊。此外，其應用範圍廣泛，不僅能夠提高企業的生產力和工作效率，還能夠促進知識的共享和創新。RAG ^{plus} 無疑為企業帶來了無限的可能，開創了智慧文本管理的新紀元。