

生成式 AI 輔助檔案銷毀智慧審查

Generative AI-Assisted Records Destruction Smart Review

邵成武 Shao, Chen-Wu

凌群電腦 AI 先進實驗室總處長

SYSCOM COMPUTER ENGINEERING CO. Executive Director

摘要

國家發展委員會檔案管理局過去使用分辨式 AI 技術進行檔案銷毀審查，但因訓練資料不足，模型準確率受限。生成式 AI 的興起為此帶來突破，透過生成式 AI 可透過上下文學習及少量資料提示的特性，提高銷毀審查的準確性並推動流程自動化。此外，生成式 AI 也讓經驗傳承系統化可以有效落實。

生成式 AI 導入的挑戰之一是避免「幻覺」問題，需透過微調（Fine-Tuned）和中間層架構（檢索擴增生成 RAG）來提升準確率和實用性，包含向量化技術、語意搜尋及提示工程等，有效整合生成式 AI 和內部系統。並且，將檔案銷毀規範和案例向量化存入資料庫，進行檢索擴增生成後，組成提示語（Prompt），由生成式 AI 產生審查建議，整體流程將實現從檔案銷毀目錄彙送、AI 智慧審查到最終核准銷毀檔案目錄下架自動化，並持續通過回饋提升 AI 準確性。

Abstract

The National Archives Administration of National Development Council had used “Discriminative” AI technology for records destruction review, but due to insufficient training data, the accuracy of trained model inferences was limited. The rise of “Generative” AI brings breakthroughs to this. By leveraging In-Context-Learning and Few-Shot-Prompting, Generative AI can improve the accuracy of destruction reviews and enable the automation of the process. Additionally, it enables systematic transfer and effective implementation of expertise.

One of the challenges in implementing Generative AI is to prevent the issue of “hallucinations.” To improve accuracy and practicality, Retrieval Augmented Generation (RAG) architectures must be applied to the process, including vectorization techniques, semantic search, and prompt engineering. It can effectively integrate Generative AI with internal systems. After the regulations and cases of records destruction are vectorized into database, RAG is used to construct prompts, allowing Generative AI to provide reviews.

The overall process will be fully automated—from the compilation of records destruction catalogs and AI-powered intelligent review, to the final approval and removal of record destruction catalogs. The system will continuously improve AI accuracy through feedback mechanisms.

關 鍵 字 | 生成式 AI、檔案銷毀、檢索擴增生成（RAG）、智慧審查

Keywords | Generative AI, record destruction, Retrieval Augmentation Generation (RAG), intelligent review

壹、前言

國家發展委員會檔案管理局（以下簡稱檔案管理局）過去利用分辨式 AI（Discriminative AI）試作檔案銷毀審查。分辨式 AI 擅長學習區分不同類別的資料，可用於分類和預測，就基於給定的輸入數據來判斷其所屬的類別的運用，已有初步成果。惟分辨式 AI 的準確率有其侷限性，主要原因在於由於訓練資料較少，不易訓練出良好的模型。就目前檔案管理局的現況，即使將訓練資料極大化，對分辨式 AI 來說仍不足，因為以上限制，在檔案管理局利用分辨式 AI 進行檔案銷毀審查，則精準度難以進一步提升。

近期，生成式 AI（Generative AI）興起，有別於分辨式 AI，利用大語言模型（Large Language Model, LLM）具備的上下文學習（In-Context Learning）技術（Dong et al., 2022），可以利用少量資料提示（Prompt）的特性，進行銷毀審查推論，可以彌補資料較少的缺憾。利用生成式 AI 的新架構、新技術，將可提升檔案管理局銷毀預審的精準度，以此為基礎，進一步將流程自動化，取代現行人工及紙本作業，進行數位轉型。

導入生成式 AI 的另一個附加價值，則是解決經驗傳承的問題。過去銷毀目錄審查經驗，仰賴資深承辦人員專業判斷，無法系統化傳承給新人，人員異動易形成經驗斷層。透過生成式 AI 檔案銷毀智慧審查，讓經驗傳承系統化可以有效落實。

貳、從現行檔案銷毀流程探討生成式 AI 可利用之處

根據《國家發展委員會檔案管理局機關檔案銷毀審核作業須知》，檔案銷毀作業依據主要為檔案法、檔案法施行細則、機關檔案保存年限及銷毀辦法、檔案保存價值鑑定規範、機關檔案

管理作業手冊、機關共通性檔案保存年限基準（GRS）等法源或規範。

現行檔案銷毀作業程序，實務上的流程主要為：

1. 檔案銷毀目錄製作
2. 檔案銷毀目錄內部審核
3. 檔案銷毀計畫、目錄或審查紀錄送核
4. 執行檔案銷毀
5. 相關目錄註記及彙送

分析實務流程，目錄形式審核、銷毀計畫審核自動化等項目，以傳統資訊管理系統技術即能做到，不是本文探討的重點，不多做贅述。本文焦點集中於較耗費人力的「3. 檔案銷毀計畫、目錄或審查紀錄送核」的保存年限判斷。保存年限大致可分為「永久保存」及「定期保存」，「永久保存」的檔案又可分為「列為國家檔案」及「機關永久保存」，定期保存又可分為「屆期後鑑定」及「依規定程序銷毀」，主要判斷依據為機關共通性檔案保存年限基準（GRS）及機關保存年限區分表，可以利用生成式 AI 進行推論及判斷。

首先有規則可循者，可直接引用規則來推論，判定不得銷毀，例如：

- 不動產撥出，重要關鍵字如土地屬、不動產撥用等。
- 本機關行政規則，重要關鍵字如要點、辦法、實施辦法、實施要點等。
- 同機關曾經收錄成國家檔案的檔案，可能有漏徵集情形，在機關彙送銷毀目錄時檢查發現。

沒有規則可直接判定者，則透過相似機關共通性檔案保存年限基準（GRS）、機關保存年限區分表及過去案例，產生提示詞（Prompt），並利用生成式 AI 上下文學習（In-Context Learning）及少量資料提示（Few-Shot Prompting）的特性，進行銷毀審查推論，這也是本文要探討的核心理論。

但並非所有規則，生成式 AI 都能推論，規則必須具體，不能太抽象或需仰賴外部資料，抽象或需仰賴外部資料的檔案目錄保存年限，人工都難以判斷，遑論 AI，此部份則需另外開發輔助系統判斷，現階段則以人工處理為主，例：

- 重大輿情：以見報與否認定是否為重大輿情

重大輿情本身是很抽象的定義，何謂重大輿情？此部份難以量化或具體化，必須透過輔助系統固定時間至網路上搜尋新聞，做為重大輿情的依據。這類檔案銷毀目錄可能只是銷毀審查的極小部份，投入大量成本開發這個輿情搜尋系統，是否值得有待商榷，因此現階段以人工處理為主。

- 具國家或機關重要行政稽憑價值：判斷檔案是否具有價值

國家或機關重要行政稽憑價值也是很抽象的定義，同樣難以量化或具體化，仍需人工處理。

- 組織沿革、主要業務運作及職能演變：參考各機關官方網站

此部份需參考外部大量資料，難以實作，如同重大輿情的判斷，投入開發也不符經濟效益，仍需人工處理。

針對可以導入生成式 AI 智慧審查的部份，進行流程改造，從檔案銷毀目錄彙送、到檔案管理局審核通過、到目錄下架，全部流程自動化並輔以 AI 判定，整體概念簡述如下：

1. 機關透過國家檔案及機關檔案整合平臺臺（編輯註），彙送檔案銷毀目錄
2. 檔案銷毀目錄自動檢測，目錄形式審核正確後，自動送審上級機關
3. 上級機關預審，審查通過後自動送檔案管理局審查
4. 生成式 AI 智慧審查，判定是否可銷毀，及不得銷毀原因

5. 檔案管理局承辦人員依 AI 判定結果進行覆核，若不同意 AI 判定結果，案例及意見將回饋給 AI，使 AI 下次依此結論審核，增加準確率

6. 核准銷毀檔案目錄，機關審視結果進行檔案銷毀並註記銷毀日期，於國家檔案及機關檔案整合平臺自動下架檔案目錄

上述自動化流程的檢核錯誤訊息，都將回饋機關，進一步提升機關檔案銷毀目錄內部審核的能力。

參、生成式 AI 檔案銷毀智慧審查論述

訓練大語言模型必須有強大的運算能力，動輒數十億美金投資，並非檔案銷毀智慧審查可以採取的方案。檔案銷毀智慧審查發展的方向是利用已中文繁體微調（Fine-Tuned）好的模型，發展新應用。

生成式 AI 這麼美好，為什麼一般機關尚在觀望？主要原因是大語言模型共同存在的問題，也就是「幻覺（hallucinations）」，ChatGPT 的使用者應有體驗到生成式 AI 的強大能力，但也會發現 ChatGPT 會很認真回答「錯」的答案，因此開發出一個能避開幻覺，機關可以輕易上手、結合營運流程並且能夠確保資安與隱私的應用方式，成為本文要闡述解決方案的重要方向。

現有模型微調示意圖如圖 1。

透過對大語言模型不知道答案的知識進行微調或持續訓練，可以解決部份幻覺問題。但仍會有一定比例的錯誤，試想一桶水倒入大水池，必定混合在一起，很難有效挑出該桶水回答問題。

進一步的規劃，則是在人跟生成式 AI 中間，設計一個中間層。前置處理並內嵌（向量化）檔案目錄銷毀的規範、核准案例及不得銷毀案例，將內嵌向量化資料存入向量資料庫，利用語意搜尋找到相似的範例，組成提示語（Prompt）詢

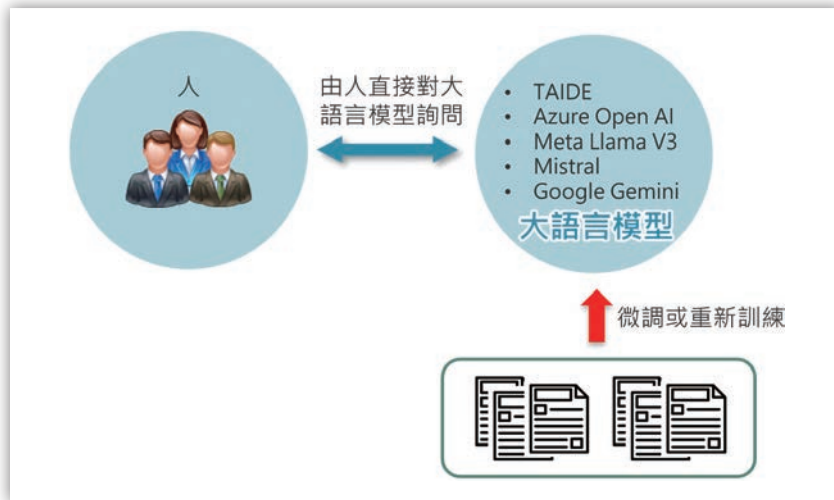


圖 1 大語言模型微調示意圖

資料來源：作者繪製

問生成式 AI，得到最終答案，判定是否核准的建議。

AI 中間層作法獨立，可串接不同大語言模型包含 TAIDE、OpenAI、Meta Llama V3、Mistral、Google Gemini...等，整合的技術包含檢索擴增生成（RAG, Retrieval-Augmented Generation）（Lewis et al., 2020）、提示工程（Prompt Engineering）、內嵌向量化（Embedding）、少量資料提示（Few-

Shot Prompting）、思維鏈（Chain of Thought）（Wei et al., 2022）、文章摘要（Text Summarization），以下章節將逐一說明。

此作法優點：

1. 可整合公有雲、私有雲或在地端的大語言模型

市面上多數解決方案採用公有雲方式，此方式有資訊安全及整合疑慮，如：



圖 2 「大語言模型 + AI 中間層」架構示意

資料來源：作者繪製

■ 公部門資料具有業務機敏性，無論上傳過程或是向量資料庫放置於公有雲，本身都有資訊安全疑慮。

■ 大語言模型廠商提供公有雲服務，不易於機關內部系統整合，公部門有些內部系統不能對外連線。

AI 中間層整合公有雲、私有雲或地端的大語言模型，特別是在地端，可確保資訊安全及整合延伸。

■ 向量資料庫放置在地端，保有資訊安全。

■ 易於與機關內部系統整合。

■ 客製化程度較高。

2. 不用微調或重新訓練大語言模型

資料有變動時，如檔案目錄銷毀的規範修正、新的核准案例及不得銷毀案例異動，不必微調大語言模型，只要將修正或異動資料內嵌向量化資料存入向量資料庫即可。

基於上述作法，規劃一個可延伸的架構，不

單是為了檔案銷毀，也為後續可繼續延伸不同功能的智慧應用，架構示意圖如圖 3。

此生成式 AI 整合平臺，主要由上述中間層概念為主所組成之架構，中間層利用向量資料庫儲存內嵌向量化資料，整合大語言模型各項技術，後續擴充將文本解析為知識地圖（Knowledge Graph）（Chen et al., 2020），進一步精煉文章，以知識三元組萃取重要資訊，以精煉過的資料品質來涵蓋答案，示意圖中之智能客服、語意關聯、知識地圖都是未來可擴充的方向。

內嵌資料如果未作任何抽象化或預處理，即表達文本的原意。依原意來尋找答案，回答問題，就形成了智能客服。當文本進一步精煉，擷取知識三元組可進一步做語意關聯。當文本要進行內嵌向量化時，先進行預處理，將一些特定機關、人名、地名進一步抽象化，用另外抽象文字代替如 [機關]、[人名]、[地名]，則文本能更進一步抽象化變成特定案例範本，將會增加語意搜尋的精準度，進而生成更符合答案的結果。

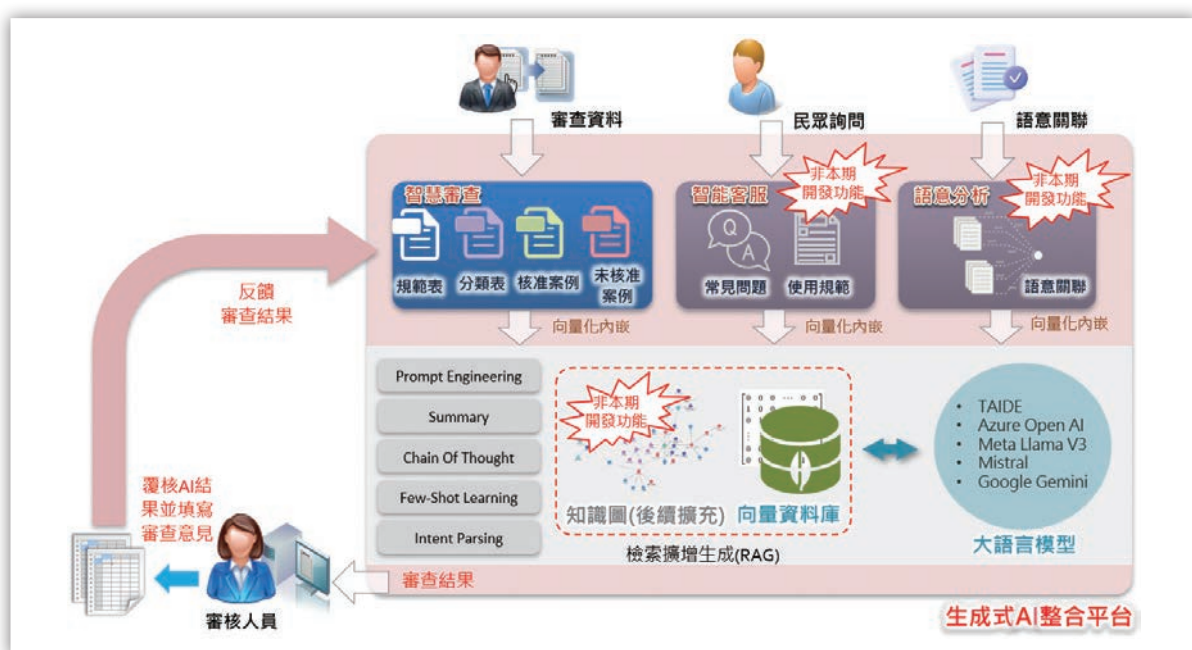


圖 3 生成式 AI 架構示意圖

資料來源：作者繪製

肆、生成式 AI 整合平臺關鍵技術說明

在說明整個檔案銷毀運作之前，先針對本案用到的關鍵技術進行說明。

一、內嵌向量化 (Embedding)

上述論述中一直反覆提到內嵌向量化？為什麼文本要向量化，直接利用原始的文本會有什麼問題嗎？AI 的基礎是向量，所有資料不管文本、語音、圖片、影像都要向量化，向量化後，數學或統計的理論才能應用，進而有現在百花齊放的榮景。

例如，要尋找一篇特定文章，大眾最熟悉的做法就是全文檢索。建立全文檢索引，利用不同關鍵字的 AND 或 OR 來檢索，比對到關鍵字的文章就標記出來。這類作法能有一定程度的效率，也解決大部份資料查找的問題。但僅能進行文字上的比對，不曾了解文章所要表達的意思，當需要檢索相似文章時，就捉襟見肘。

把文字轉成向量 (Latent Matrix)，投影到向量空間 (Latent Space)，相近的意思會被放置在一起，可於相似性文章的檢索，產生很好的效果。如查詢「交通工具」，可以找到汽車、機車，全文檢索只能找到文章中有出現過「交通工具」關鍵字的文章，向量空間 (Latent Space) 是一個多維度的空間，一般人不易想像，我們利用二維空間來解釋，如圖 4。

文本切割進行內嵌向量化，當轉成向量後，文本相似度比對問題，就變成是數學問題。如上圖二維向量空間我們發現 man 跟 king 很接近，woman 跟 queen 很接近，但 man 跟 woman 就離下方 London 很遠，我們可以很簡單算出兩點距離，當距離小表示兩個向量 (語意) 很接近。

二、中間層向量資料庫及搜尋

當文本內嵌向量化後，對於越來越多的文本，必須有個向量資料庫能儲存這些大量資料，並能快速搜索。向量資料庫現在商用及免費都有

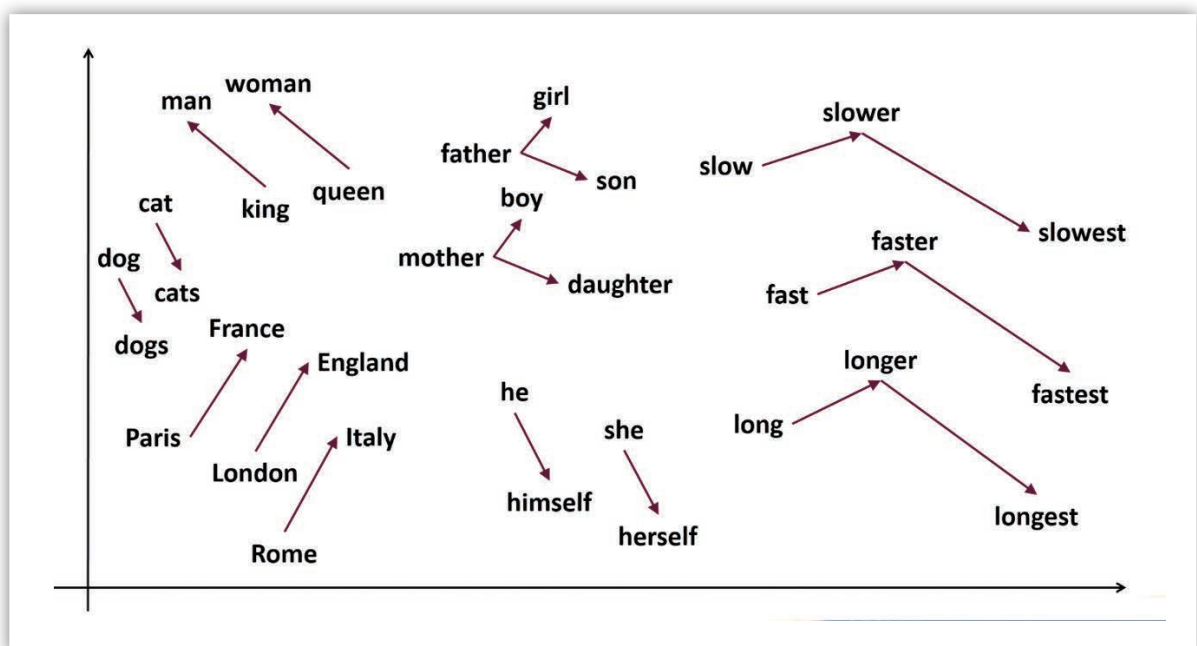


圖 4 二維向量空間示意圖

資料來源：A Deep Dive into NLP: From Word2vec to BERT

很多，也內建搜索的函數進行相似句比對，以下為向量資料庫的示意圖如圖 5。

向量資料庫比對到的資料將取前面幾筆相似度較高的文本，利用 prompt 請生成式 AI 判定是否核准。

三、檢索擴增生成 (Retrieval-Augmented Generation, RAG)

檢索擴增生成是一個重要關鍵技術，要說明檢索擴增生成前，先談一下大型語言模型的局限性。

大語言模型具有強大的自然語言處理能力，這些模型透過大量的數據進行訓練，讓模型能夠理解和生成出人類使用的語言，甚至能夠在一定程度上理解上下文和意圖。然而，大語言模型也存在明顯的局限性。首先，它們缺乏真正的理解能力，只能根據過去訓練數據中的經驗與方法進行回應，在面對不曾見過的問題時，無法正確回答，回覆的內容可能會有偏見或不準確的問題，因為模型的回答只是反映了訓練數據中的想法，這是大語言模型共同有的問題——幻覺。

如何避開幻覺？最好的方法就是協助大語言模型獲得新知識，此外也可以限制其回答的答案在我們給定的知識庫中。而獲得新知識的方式若採用之前段落提的微調，則需要較多的運算力及資源，且微調出來後尚有部份會產生幻覺，檢索擴增生成變成較容易克服幻覺的解決方案。

RAG 架構主要由兩個部分構成：檢索和生成。從向量資料庫中「檢索」相關知識訊息後，送到大語言模型進行生成處理，利用檢索到的知識生成回應，示意圖如圖 6。

RAG 要能完美運作有幾個關鍵因素：

(一) 內嵌至向量空間的準確性

大語言模型是由英語系的國家所發展出來，英文文本內嵌向量化模型的訓練已臻成熟，繁體

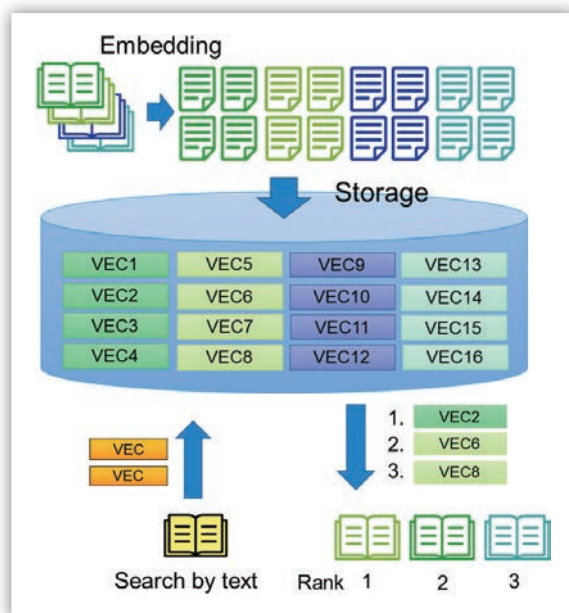


圖 5 向量資料庫示意圖

資料來源：臺達電閻志克博士演講 ChatGPT-based Hybrid Enterprise Knowledge Management 簡報

中文的內嵌模型則較為缺乏，市面上充斥使用簡體中文的內嵌模型，選擇內嵌模型至關重要。

(二) 長文本切割內嵌的大小

除了慎選內嵌模型之外，文本如何切割大小做內嵌，也至關重要。若把不相關的文本放在一起內嵌，文本的語意會跑掉，造成大語言模型誤解其本意。切割需要注意以下幾點重點：

- 問答集，一個問答一個內嵌，不同問答不能放在一起，且同一個問答的「問」跟「答」需要一起內嵌。
- 長文本設定切割的大小區塊（chunk），把長文本切成一個一個的小區塊，而當切成區塊後，區塊與區塊之間的語意則會不連續。而如何讓文章意義能連貫，則可以設定下一個區塊可以跟前一個區塊重疊（overlap）多少個字，例如 500 字切一個區塊，重疊 200 字，下一個區塊不是從 501 開始，而是從 301 開始，區塊及重疊



圖 6 檢索擴增生成 (RAG) 示意圖

資料來源：作者繪製

的概念示意圖如圖 7。

(三) 大語言模型的推理 (Reasoning) 能力

大語言模型的推理 (Reasoning) 能力，也是所謂的上下文學習 (In-Context Learning)，預先提供知識時，模型是否能讀懂知識中的內容，並正確回答問題。

四、大語言模型 (Large Language Model, LLM)

採用 RAG 架構比較獨立，後面生成的大語言模型可以接雲端的 Azure OpenAI、Google Gemini、Claude 等模型，在地端可接 Llama V3、Mistral、TAIDE 等模型。由於資料上雲端有機敏性的問題，檔案銷毀智慧審查將以在地端大語言模型為主。

五、提示工程 (Prompt Engineering) 及上下文學習 (In Context Learning)

利用提示工程來規範生成式 AI 的行為，結合提示 (Prompt) 及大語言模型中的上下文學習 (In-Context Learning)，本案將會用到大語言模

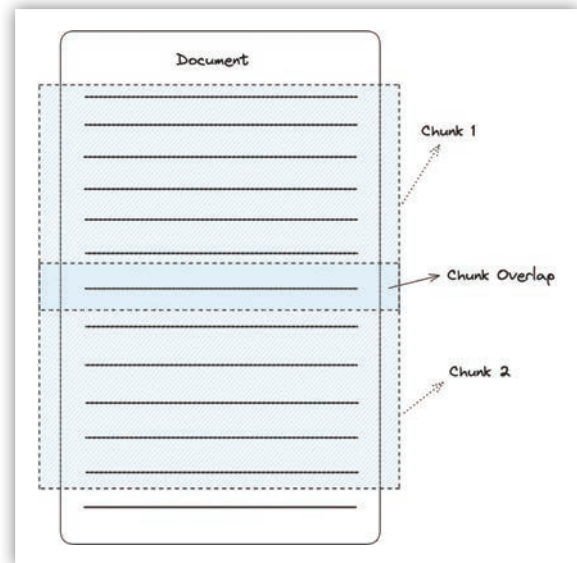


圖 7 區塊及重疊的概念示意圖

資料來源：LangChain 怎麼玩？用 Document Loaders / Text Splitter 處理多種類型的資料

型中少量資料提示 (Few-Shot Prompting) 能力，讓大語言模型讀懂機關共通性檔案保存年限基準 (GRS)、機關保存年限區分表、核准案例及不得銷毀案例，判斷檔案是否可銷毀，如下列提示的示例，灰色底色的部份是可變動的，依不同銷毀檔案目錄，擷取不同相似資料：

使用者提供一個檔案目錄，描述檔案目錄的分類編號、分類名稱、內容描述，將提供下文相似的分類、相似的標準保存年限規範、過去核准銷毀的案例、過去永久保存案例，提供判斷檔案目錄保存年限。

-- 相似分類

{RAG 中插入的相似機關共通性檔案保存年限基準 (GRS)、機關保存年限區分表 }

-- 過去核准銷毀的案例

{RAG 中插入的相似核准案例 }

-- 過去永久保存案例

{RAG 中插入的相似永久保存案例 }

請用使用者提供的檔案目錄中的 [內容描述] 去理解內容，並對照上文意思相近的，[內容描述] 不用完全相同，必要時考慮 [分類號]，得到保存年限，請說明保存年限判斷的原因，摘要成 500 個中文字以內，請用以下格式及中文繁體回答，輸出格式如下：

——輸出格式

保存年限判斷結果：保存年限判斷結果

保存年限判斷原因：保存年限判斷結果，請說明原因

使用者提供檔案目錄：

{ 機關要判斷是否可銷毀的檔案目錄 }

六、大語言模型彙整回答

大語言模型加上彙整提示彙整回答，回饋審查結果，以下是彙整摘要提示：

使用以下上下文來回答問題，

請針對問題回答，如果您不知道答案，只需說 "" 資訊不足，問題再描述詳細一點 ""，不要試圖編造答案。

{RAG 中插入的檢索結果 }

請用中文繁體回答，摘要成 300 個中文字以內

問題：{ 使用者詢問的問題 }

回答：

如圖 8 範例，使用者詢問「請問檔案的保存年限，從何時起算？」。

實際上在向量資料庫中語意搜尋到的答案有多個，由大語言模型總結回答如下：

問題：請問檔案的保存年限，從何時起算？

答案：檔案的保存年限從文件產生日期起算。案件層級以文件產生日期為起點，案卷層級則從案卷內



圖 8 大語言模型總結多項答案生成結果

資料來源：作者繪製

最晚文件產生日期起算。文件產生日期的判定依據為：機關創簽文件以核判日期為準，發文以發文日期為準，來文存查則以來文機關的發文日期為準。

七、覆核 AI 結果回饋

審查人員覆核 AI 結果，並線上填寫審查意見，系統將其審查結果回饋至向量資料庫，使向量資料庫越來越智慧。

此步驟非常重要，AI 判定保存年限若與審查人員認知不同，審查人員覆核的保存年限結果將回饋至向量資料庫。未來若有同類型案例，將以覆核後的結果組成提示詞，增加保存年限判斷的正確性，若沒有回饋，AI 判定保存年限會跟以前一樣的結果，無法使 AI 變得更聰明。

伍、檔案銷毀智慧審查運作機制

一、二階段審查

檔案銷毀智慧審查實質運作機制分兩階段自動審查。第一階段基於不得銷毀規則審查，第二階段進階審查。第一階段將過去不得銷毀的特殊情形審查、銷毀轉移轉經驗、審查經驗經過生成式的提示進行基於不得銷毀規則審查，第一階段

已判定不得銷毀，則不必進入進階審查。

第一階段不得銷毀規則如下：

- 叛亂犯判決及處置
- 戒嚴時期情治機關執掌與分工、預算
- 情治機關對特定團體之監控
- 戒嚴時期執政黨承租借用及轉讓房地
- 戒嚴時期執政黨與機關之聯繫
- 戒嚴時期執政黨附隨組織之經費補助
- 黨職併公職
- 出版品查禁
- 對外軍援、軍購、國際協定簽署等
- 對匪作戰、資匪案等
- 臺灣省物資局機關增設及精簡裁併、主管會報、代辦木材外銷墊付費用、倉儲租賃管理、
- 嘉義縣附屬單位預算綜計表
- 原臺中縣政府總會計報告
- 35 至 38 年印鑑證明申請書及廢止之印鑑案
- 921 震災之後續處置
- 臺灣省防救天然災害及善後處理辦法
- 大甲溪水資源開發案
- …（以下省略）

進階審查將資料抽象化及向量化，從中間層知識庫，擷取相似機關共通性檔案保存年限基準（GRS）、保存年限區分表、銷毀核准案例、永久保存案例等文件組成提示語由大語言模型判定是否核准銷毀，提示語的組成如示意圖 9。

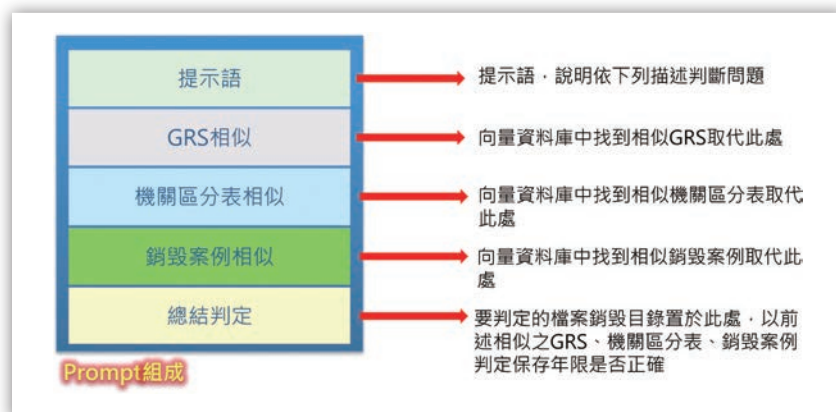


圖 9 檔案銷毀智慧審查進階審查提示語組成示意圖

資料來源：作者繪製

二、資料前置處理

此二階段檔案銷毀智慧審查要能真正發揮效能，必須先針對資料進行前置處理，包含第一階段不得銷毀規則設定，以及會用來判斷的特殊情形審查、銷毀轉移經驗、審查經驗等。第二階段資料抽象化及向量化，包含機關共通性檔案保存年限基準（GRS）、保存年限區分表、銷毀核准案例、永久保存案例等文件，這部份也是最耗時的部份。

（一）第一階段基於不得銷毀規則前置處理

把不得銷毀之特殊情形審查、銷毀轉移經驗、審查經驗，利用生成式提示工程（Prompt Engineering）把特殊規則變成提示，請大語言模型判定是否符合。一個規則一個提示，如下例：

你在回答問題時，將會被要求判斷問題是否為與「**叛亂犯判決及處置**」相關，答案只要回答「是」或「不是」，在說明中說明原因，格式如下：

問題：要判斷是否為「**叛亂犯判決及處置**」的描述

答案：回答「是」或「不是」

說明：說明判斷為是或不是的原因

-- 例子開始

問題：臺北縣三重鎮住民顏○○○憑日據時代戶口調查籍謄本申請更正年齡為民國三十年一月十二日出生及更正出生別為次男，應予照准。（併案）

答案：**不是**

說明：問題中描述內容是要更正年齡及更正出生別，與叛亂犯判決及處置無關

-- 例子結束

-- 例子開始

問題：案奉臺灣省政府○○○○○○○號密代電開：「一，案據農林廳林產管理局號○○○○○○○代電以『准臺灣省保安司令部代電囑查明因參加叛亂組織已處刑確定職教員之介紹人保證人及單位主管姓名分別議處見復一案。』

答案：**是**

說明：問題中描述內容與叛亂犯判決及處置相關

-- 例子結束

問題：{ 要詢問的檔案目錄內容描述 }

其中「**叛亂犯判決及處置**」為特定規則，可取代為不同的特殊情形審查、銷毀轉移經驗、審查經驗，此部份透過設計畫面，使用者可輸入特定規則及例子，由程式自動產生提示，{ 要詢問的檔案目錄內容描述 } 為使用者提出的檔案目錄內容描述，大語言模型，只需回答「是」或「不是」，如果「是」就判定為不得銷毀，不會進入第二階段。

（二）第二階段進階審查前置處理

進階審查將資料抽象化及向量化，從中間層知識庫，擷取相似機關共通性檔案保存年限基準（GRS）、機關保存年限區分表、銷毀核准案例、不得銷毀案例等文件組成提示語，由大語言模型判定是否核准銷毀，示意圖如圖 10。

1. 資料前置處理—共通性檔案保存年限基準（GRS）

GRS 共計 19 大類（曾經有 20 類，教育類於 97 年停用，現為 19 類），其中「01 政風類」、「03 主計類」、「04 人事類」及「06 行政類」四類為全臺各機關之通用類別。以每一個分類號作為一筆資料進行內嵌向量化，因 GRS 為機關共通性檔案保存年限基準，只需保留一份，存在向量資料庫同一個集合中。

將舊資料進行一次性處理，將 GRS 不同格式資料轉換成統一格式的 JSON 檔，再解析 JSON 檔向量化每一筆分類，示意圖如圖 11。

GRS 雖不常變動，但仍有變動的可能性，系統運行時提供畫面供使用者上傳更新的 GRS，由系統重新內嵌內量化，取代舊的內嵌向量化資料。

2. 資料前置處理—機關保存年限區分表

機關保存年限區分表為各機關編制之保存年限表，以每 1 個分類號作為 1 筆資



圖 10 檔案銷毀智慧審查進階審查示意圖

資料來源：作者繪製

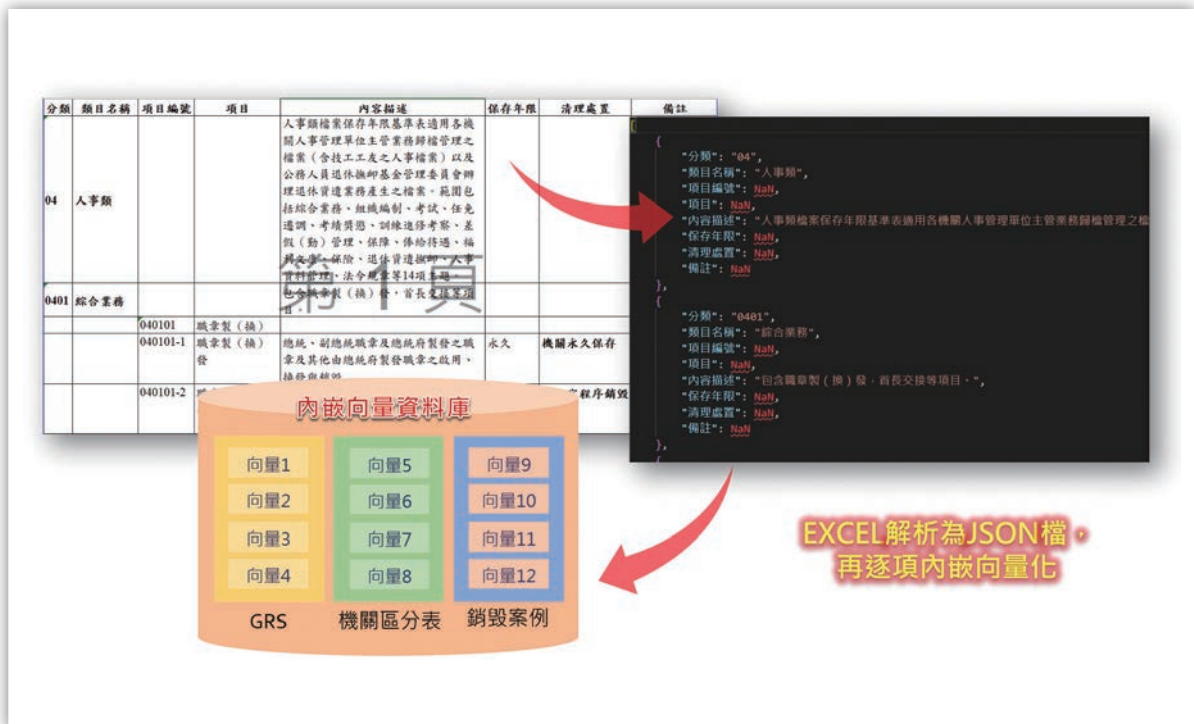


圖 11 共通性檔案保存年限基準（GRS）向量化示意圖

資料來源：作者繪製

料進行內嵌向量化。由於各機關之內容及分類號不盡相同，單一機關之保存年限區分表僅能供該機關使用，因此機關保存年限區分表的資料量並不大。每個機關的保存年限區分表獨立內嵌在單一資料集，彼此互不影響，機關保存年限區分表的新增、刪除、修改，也會透過國家檔案及機關檔案整合平臺連動並內嵌向量化異動。

將機關保存年限區分表不同格式資料轉換成統一格式的 JSON 檔，再解析 JSON 檔，向量化每一筆分類，如圖 12 示意。機關保存年限區分表目前並無資訊化，有多種格式，每一種格式代表需要建立一種解析規則。在建置初期，機關保存年限區分表的解析將會是一項無可避免的辛苦且耗時的工程，未來將以國家檔案與機關整併的新格式進行解析。

3. 資料前置處理—銷毀核准案例、不得銷毀案例

除了機關共通性檔案保存年限基準（GRS）、機關保存年限區分表，特別加入了核准銷毀目錄與不得銷毀目錄，主要原因如下：

- 機關區分表及 GRS 表達語意太窄，無法涵蓋較大範圍。
- 過去的經驗可以做為日後參考。
- 每次的審核結果可以回饋給平臺，使平臺判斷能越來越精準。

內嵌核准銷毀目錄與不得銷毀目錄必須做資料抽象化處理，如果未作任何抽象化處理，即表達文本的原意，則當文本中有一些機關、人名、地名，將使文本的語意跟這些字綁在一起。透過將這些特定機關、人名、地名進一步抽象化，用抽象文字代替如 [機關]、[人名]、[地名]，則文本能更進一步抽象化變成特定案例樣板，增加



圖 12 機關保存年限區分表向量化示意圖

資料來源：作者繪製

語意搜尋的精準度，進而能生成更符合答案的結果，如下例子：

案由：

所送被繼承人張三先生所遺座落新莊市新樹段三三三
地號土地之逾期未辦繼承登記土地列冊管理單乙案，
上開標（的業已註銷列冊管理，請查照。）

抽象化後：

所送被繼承人[人名]先生所遺座落[地名]土地之逾
期未辦繼承登記土地列冊管理單乙案，上開標（的業
已註銷列冊管理，請查照。）

上述案例，將人名及地名抽象化，使其更適合後續進行語意搜尋，所以必須能把文本中的[機關]、[人名]、[地名]取代。這在過去不易利用程式撰寫達成，但使用生成式 AI 則可藉由提示語達成。

銷毀核准案例、不得銷毀案例先轉成統一格式的 JSON 檔，經由上述之資料抽象化後，再內嵌向量化。核准案例與不得銷毀案例以不同資料集存放以做區隔。

陸、運行流程及預期效益

一、運行流程

前面章節已闡述檢索擴增生成是本案一個重要關鍵技術，RAG 用到的整合技術包含提示工程、內嵌向量化、上下文學習、少量資料提示、文章摘要等。本文著重於流程 AI 化的部份，如前所述，經過兩階段自動審查，第一階段基於不得銷毀規則審查，第二階段則是進階審查。第一階段是基於不得銷毀規則審查，將過去不得銷毀的特殊情形審查、銷毀轉移轉經驗、審查經驗經過生成式提示工程進行審查，若第一階段已判定不得銷毀，則不必進入進階審查。進階審查將資料抽象化及向量化，從中間層知識庫，擷取相似機

關共通性檔案保存年限基準（GRS）、機關保存年限區分表、銷毀核准案例、永久保存案例等文件，組成提示語由大語言模型判定是否核准銷毀。

運行流程涉及機關、上級機關、國檔機關整合平臺、智慧審查（AI）、檔案管理局承辦人等相關角色之間的交互關係，僅就 AI 部份主要流程與角色之交互關係說明如示意圖 13。

上述示意圖，以左邊不同角色為主軸，說明如下：

（一）機關、上級機關角色

機關自國檔機關整合平臺彙送檔案銷毀目錄，透過國檔機關整合平臺檔案檢測銷毀目錄剛性規則，不符合即退回，機關修正後重新彙送，檢測通過，上級機關可透過平臺進行預審。檢測失敗或審查結果，機關皆可於系統上檢視，或下載回饋可解析資料，提供機關後續應用。

（二）國檔機關整合平臺角色

提供機關彙送檔案銷毀目錄並檢核，也提供上級機關預審，預審通過則進入智慧審查（AI）。

（三）智慧審查（AI）角色

第一階段基於不得銷毀規則，第二階段進階審查，第一階段已判定不得銷毀，則不必進入進階審查。進階審查將案例抽象化及內嵌向量化，從中間層知識庫，擷取相似機關共通性檔案保存年限基準（GRS）、機關保存年限區分表、銷毀核准案例、永久保存案例等文件組成提示語由大語言模型判定是否核准銷毀。AI 智慧審查主要註記 3 個 AI 欄位，包含 AI 審查意見、AI 審查保存年限、AI 審查是否銷毀。

（四）檔案管理局承辦人角色

檔案管理局承辦人根據 AI 審查意見、AI 審查保存年限、AI 審查是否銷毀，搭配機關彙送

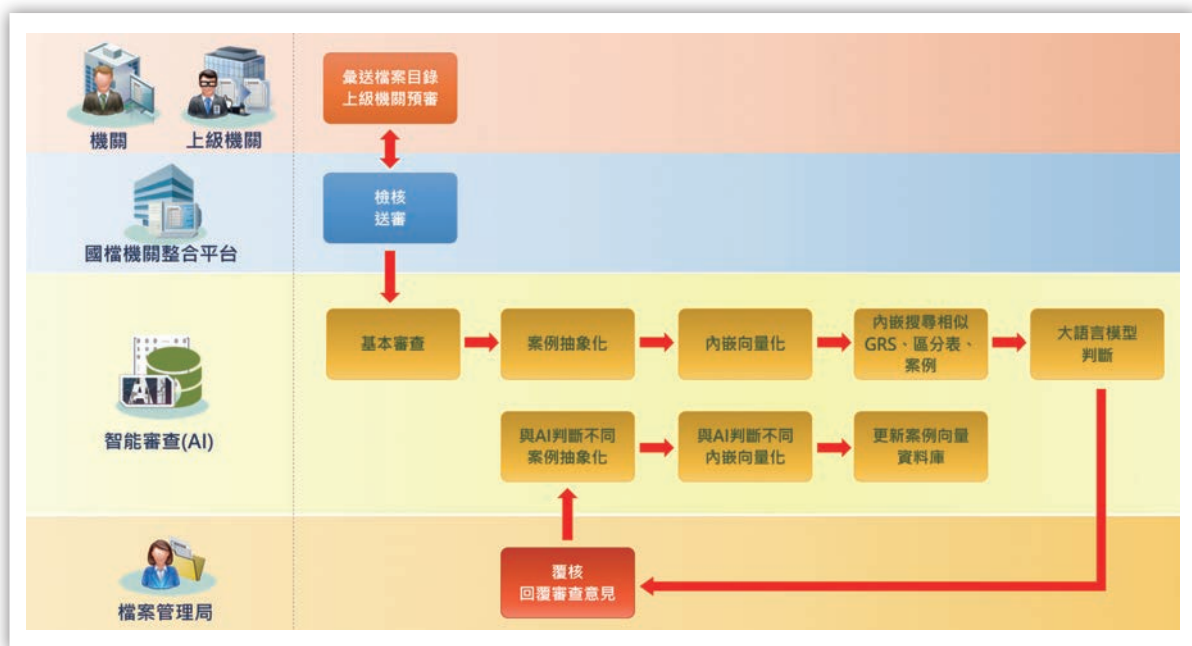


圖 13 檔案銷毀智慧審查流程（AI 部份）示意圖

資料來源：作者繪製

的檔案銷毀目錄及保存年限，進行覆核。如果不同意 AI 建議結果，可以說明建議保存年限、審查意見或不得銷毀意見。

覆核完成後，針對不同意 AI 建議結果之案例，進行抽象化及內嵌向量化，將此修正判定更新向量資料庫，做為下一次判斷基準，AI 將越來越智慧。

二、預期效益

如本文一開始所闡述以生成式 AI 新架構、新技術來提升檔案管理局銷毀預審的精準度，預期效益為：

1. 科技洪流生成式 AI 應用亮點

- 利用大語言模型的自然語言理解能力，提示語少量資料提示，解決訓練資料不足的問題，提升檔案銷毀智慧審查精準度。
- 利用大語言模型的摘要能力，總結結果。

2. 覆核功能強化案例回饋 AI

- 提供查詢篩選、整批覆核。
- 審查結果回饋系統，案例越來越多，AI 將越來越精準。

3. 流程自動化、經驗系統化

- 提供二階段 AI 智慧審查，結合大語言模型判定，並進一步改造現行流程，使流程自動化。
- 解決經驗傳承的問題，人員異動易形成斷層，透過生成式 AI 檔案銷毀智慧審查，讓經驗傳承系統化可以有效落實。

4. 數位轉型

流程自動化、AI 智慧審查，取代傳統的紙本審查，達到數位轉型的目標。

三、未來發展方向

1. 功能性擴充

- 初期 AI 檔案銷毀智慧審查將由試辦機關進行功能驗測，並回饋改善意見，待系統

更完善後，推廣至全國機關。

- AI 檔案銷毀智慧審查目前實作在檔案管理局的管理端。未來將進一步推動服務至機關端，由機關端自行針對要銷毀的檔案目錄，進行 AI 預先判斷，判定可以銷毀再送至檔案管理局，減輕檔案管理局承辦人負擔。

2. 領域擴充

相關技術或平臺，未來不僅使用在檔案銷毀，也可應用於不同領域。例如透過解析出重要人、事、時、地、物，可進行跨文本之間的語意關聯，進一步可以文找文，進行所有相關資料的快速查找。

參考文獻

- Luo, M., Xu, X., Liu, Y., Pasupat, P., & Kazemi, M. (2024). In-context learning with retrieved demonstrations for language models: A survey. *arXiv preprint arXiv:2401.11624*.
- Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., ... & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*, 2.
- Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., ... & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35, 24824-24837.
- Chen, Y., Li, H., Li, H., Liu, W., Wu, Y., Huang, Q., & Wan, S. (2022). An overview of knowledge graph reasoning: key technologies and applications. *Journal of Sensor and Actuator Networks*, 11(4), 78.

註釋

編輯註 本文所提之「國家檔案及機關檔案整合平臺」，係為作者所規劃之預想平臺。目前檔案管理局之目錄銷毀審查程序中未利用此類平臺，正在整體規劃建置中。