

佛光大學應用科技與設計學院資訊應用學系 學士班學生專題實作成果報告書

Department of Applied Informatics
College of Applied Technology and Design
Fo Guang University
Capstone Project Report

結合大語言模型與知識圖譜的問答系統研究 改善特定應用場景中大語言模型之幻覺問題

**A System for Question - Answering that Incorporates Large - scale
Linguistic Models along with Knowledge Graphs: The Alleviation of
Hallucinations within Applications Tailored to Specific Domains**

黃子桓
HUANG,TZU-HENG

指導教授：駱至中 老師

Project Advisor : Professor Chih-Chung Lo

中華民國 114 年 8 月
August 2025

摘要

近年來，大語言模型（LLM）在自然語言理解與生成上表現亮眼，但於衛教等高精確領域仍易產生與事實不符的幻覺問題（hallucination）——指模型輸出非事實（non-factual）且缺乏依據（ungrounded）的內容——進而削弱可驗證性與信任。為降低此風險，本研究將結構化知識導入檢索增強生成（RAG）流程，採「先檢索後生成」以約束生成依據；同時也注意到其成效與檢索覆蓋及證據品質密切相關。方法上，本研究以糖尿病為示例，分兩階段建置知識圖譜：第一階段以中華民國糖尿病學會（Taiwan ROC）權威文本進行實體對齊與關係建立，形成初步圖譜；第二階段為提升後續擴充與維護之便利性，採用 Neo4j LLM Graph Builder 自動抽取並建模實體、關係與屬性，並整合美國糖尿病協會（ADA）官網資料完成跨來源對齊與融合，建立可持續維運與擴充之高效圖譜建置方法。系統端以 LangChain GraphCypherQACChain 將提問轉為 Cypher 查詢以檢索圖譜證據，並與語義向量檢索併用，形成圖-向量混合檢索（Graph+Vector）與僅向量檢索（Vector-only）兩種設定。實驗以由圖譜語料構建之 35 題為測試集，採答案相關度（Answer Relevancy）、上下文召回率（Context Recall）、忠實度（Faithfulness）三項指標量化評估。結果顯示，Graph+Vector 在答案相關度上高於純向量基線（0.90 vs 0.72），且極低分樣本較少；上下文召回率與純向量基線相近（0.67 vs 0.68）；忠實度未見一致提升，呈現中位數較低且變異較大（0.50 vs 0.58），低分集中於特定主題段落。整體而言，系統表現有提升但幅度不大，改善主要體現在貼題性；對幻覺抑制的直接效益仍受限於檢索覆蓋與證據選用。相關性分析亦顯示，忠實度與上下文召回率高度相關（ $r = 0.77$ ），與答案相關度關聯較弱（ $r = 0.15$ ），指出「貼題」不等於「可驗證」。後續將聚焦於改進檢索策略（依信心／主題／跳數之自適應權重、結合 NLI 與 cross-encoder 的證據導向重排，以及「先證據後生成」之核驗與引用流程），以提升忠實度與可追溯性，並形成可複製之 KG-augmented RAG 標準流程，推廣至其他高精確需求之領域。

關鍵字：智慧型代理人、知識圖譜、檢索增強生成、幻覺問題、大語言模型

Abstract

In recent years, large language models (LLMs) have exhibited strong performance in natural language understanding and generation, yet in high-precision domains such as patient education they remain vulnerable to the hallucination problem—outputs that are non-factual and ungrounded—which undermines verifiability and trust. To mitigate this risk, we incorporate structured knowledge into a retrieval-augmented generation (RAG) pipeline, adopting a retrieve-then-generate strategy whose effectiveness is closely tied to retrieval coverage and evidence quality. Methodologically, using diabetes as a case domain, we constructed the knowledge graph in two stages: (i) an initial prototype from Taiwan ROC authoritative texts via entity alignment and relation construction; and (ii) an extensible, maintainable pipeline using Neo4j LLM Graph Builder to automatically extract and model entities, relations, and attributes, integrating American Diabetes Association (ADA) materials for cross-source alignment and fusion, thereby establishing an efficient KG construction method for continuous operation and expansion. The system employs LangChain’s GraphCypherQACChain to translate questions into Cypher queries for graph evidence retrieval and combines this with semantic vector retrieval, yielding two experimental settings: a graph–vector hybrid (Graph+Vector) and a vector-only baseline. Using 35 questions derived from the KG corpus, we evaluate with Answer Relevancy, Context Recall, and Faithfulness. Results show that Graph+Vector outperforms the Vector-only baseline in answer relevancy (0.90 vs. 0.72) with fewer catastrophic lows; context recall is comparable (0.67 vs. 0.68); and faithfulness shows no consistent improvement, exhibiting a lower median and larger variance (0.50 vs. 0.58), with low scores concentrated in specific topics. Overall, performance improves but only modestly, with gains mainly in topicality; direct hallucination suppression remains constrained by retrieval coverage and evidence selection. Correlation analysis further shows a strong association between faithfulness and context recall ($r = 0.77$) and a weak association with answer relevancy ($r = 0.15$), underscoring that topicality does not imply verifiability. Future work will focus on optimizing retrieval—adaptive weighting by confidence/topic/hop distance, evidence-oriented reranking with NLI and cross-encoders, and a retrieve-before-generate verification and citation pipeline—to strengthen faithfulness and traceability, and to establish a reproducible KG-augmented RAG process for other high-precision domains.

Keywords: intelligent agents; knowledge graph; retrieval-augmented generation; hallucination problem; large language models

目錄

摘要.....	ii
Abstract.....	iii
目錄.....	iv
圖目錄.....	vi
表目錄.....	vii
緒論.....	1
1-1 研究背景與動機	1
1-2 研究目標	3
第二章 文獻探討.....	5
2-1 知識圖譜	5
2-2 大語言模型	6
2-3 幻覺問題	6
2-4 檢索增強生成	7
2-5 智慧型代理人	7
2-6 LLM Graph Builder	7
第三章 研究方法與實作設計.....	9
3-1 資料擷取與前處理	10
3-2 自動文本辨識與知識抽取	11
3-3 知識融合與圖譜建置	14
3-4 系統整合與智慧應用設計	16
3-5 實驗設計與方法	17
第四章 實作驗證.....	18
4-1 知識獲取	18

4-2 知識抽取	19
4-3 知識融合	21
4-4 系統整合與智慧應用	26
4-5 實驗設計與驗證	29
第五章 總結與討論.....	36
5-1 研究貢獻	36
5-2 實驗發現	36
5-3 限制與未來展望	37
參考文獻.....	39

圖目錄

圖 1 本系統建立知識圖譜之流程圖.....	9
圖 2 系統架構圖.....	9
圖 3 爬蟲程式抓取資料範例.....	19
圖 4 爬蟲程式抓取資料內容.....	19
圖 6 知識抽取結果範例圖二.....	20
圖 5 知識抽取結果範例圖一	20
圖 7 LLM Graph Builder 操作畫面	21
圖 8 使用Cypher 語法文件生成圖譜	23
圖 9 以LLM Graph Builder 建置知識圖譜範例	23
圖 10 手動建置糖尿病知識圖譜在Neo4j圖形資料庫呈現之實作圖	24
圖 11 最終知識圖譜成果圖.....	25
圖 12 Neo4j桌面版節點關係數統計圖.....	25
圖 13 自然語言轉換資料庫查詢語法實作圖.....	27
圖 14 LLM Graph Builder 後端運作圖示	27
圖 15 智慧系統整合效果實測.....	28
圖 16 避免幻覺問題測試.....	28
圖 17 實驗結果柱狀圖.....	34
圖 18 答案相關性之小提琴圖.....	34
圖 19 答案相關性之每題分佈圖.....	35

表目錄

表一 實體與實體名稱.....	12
表二 關係名稱與定義.....	13
表三 屬性與示例類型.....	13
表四 實驗測試集.....	31

緒論

1-1研究背景與動機

近幾年來，大語言模型（Large Language Model，簡稱 LLM）於自然語言處理領域，呈現出卓越的能力。其應用範圍甚廣，在衛教、科技、金融以及教育等專業領域均有所應用。然而，LLM 時常存有被稱為「幻覺」的問題（hallucination），即模型在作出回應之際，或會生成貌似合理但欠缺事實依據之內容。此類錯誤資訊，常常以充滿自信之口吻給出，這對於 LLM 在高精確度應用場景中的可信度，形成了嚴重之挑戰。尤其是在衛教、法律、金融等具備高度專業性且對正確性要求極高之領域，幻覺問題極有可能引發嚴重之後果。原因在於，幻覺內容看似合理且信心滿滿，倘若使用者未經查證便輕易採信，很可能導致嚴重之負面影響。LLM 之所以會產生幻覺，其根源在於其基於機率之生成機制。當模型缺乏充足之背景知識，抑或無法獲取正確資訊之時，往往仍會為了生成連貫之回答而作出某些猜測。換言之，LLM 並非在檢索事實之後再行作答，而主要是依據語言統計模式來預測下一個字詞，其目的在於追求語句之通順連貫，卻無法保證內容之真實性。此種生成方式使得模型在面對未曾見過之問題，或者知識空白之際，傾向於「填補」出可能並不正確之答案。由於模型內部缺乏對於「真實性」之衡量機制，高度自信之謬誤回答便層出不窮。因而，在需要高準確度之專業應用當中，單純依賴 LLM 之內部知識庫，極易出現不確定且不可靠之回答，從而限制了其實際部署應用之價值。如何緩解 LLM 之幻覺問題，並提高其回答之可靠性，已然成為當前人工智慧應用領域的一項至關重要之挑戰。為此，近年來，檢索強化生成（Retrieval - Augmented Generation，簡稱 RAG）技術備受廣泛關注。RAG 之核心理念，在於為 LLM 之生成過程引入外部知識。具體而言，即首先從外部知識庫中檢索出與提問相關之背景資訊，隨後將其提供給 LLM 作為生成提示。通過在生成之時引用可驗證之外部資料，模型能夠將回答建立在真實之來源基礎上，而非僅僅依賴於內部訓練所得之資料內容。此種方式，有助於顯著降低 LLM 產生事實性錯誤之機率，同時提升回答之準確性與可信度。舉例來說，RAG 系統在回覆最新之衛教相關問題時，能夠檢索資料庫中權威衛教文獻或者指南作為依據，從而提供事實可靠之答覆。故而，在衛教、法律、金融等嚴謹之領域，引入 RAG 有望能有效緩解幻覺問題，增強 LLM 給出精確答案之能力。然而，RAG 技術之效能，在很大程度上依賴於知識來源之選擇以及檢索方式之優劣。傳統之關鍵字搜索或者純向量相似度檢索，在面對專業知識場景之時，或許存在力有未逮之處。一方面，關鍵字匹配難以捕捉語義層面之關聯，無法充分理解使用者

問題背後所隱含之知識需求；另一方面，當知識分散於多份文件，且彼此之間存在引用關係之時，逐段檢索常常無法整合相關之資訊。因此僅依靠向量檢索之 RAG 系統，在語境含糊或者知識結構複雜之情形下，表現往往欠佳，其缺乏對於知識之間關聯與推理之理解。例如，當多個文件包含類似或者重複之內容時，向量檢索可能難以判斷哪個片段最具相關性；又如，在回答某些跨文件之複合問題時，傳統 RAG 無法辨明多份文件內容之間之關係。其結果是，即便引入了外部資料，LLM 仍有可能因檢索不精確，或者知識碎片化，而出現資訊缺漏之現象，導致幻覺問題在專業問答之中依舊存在。有學者為進一步提升 RAG 之知識檢索能力，開始關注知識圖譜（Knowledge Graph，簡稱 KG）與 LLM 之結合。知識圖譜以「實體 - 關係 - 實體」三元組之形式組織知識，擅長表達概念之間的語義關聯以及結構化知識。與非結構化文件之檢索相比，KG 能夠提供更為精細之語義搜尋與推理支持。具體來說，通過實體鏈接和關係推演，KG 可協助識別使用者提問中隱含之相關概念，並定位到具有高相關性且經過驗證之知識。將 KG 引入 RAG 流程，可使 LLM 在生成回答之時，調用圖譜中的結構化知識。如此一來，不僅能夠降低遺漏關鍵資訊之風險，還可以提升回答之可解釋性。知識圖譜融入 LLM 推理過程之後，模型在事實正確性、推理能力以及解釋透明度等方面，均呈現出顯著之改善。因此，知識圖譜被視作增強 LLM 檢索與語義理解能力、減少幻覺之重要技術途徑之一。然而，就繁體中文語境而言，當前公開可用之知識圖譜資源相對匱乏。現有之開放知識圖譜（諸如 Wikidata、DBpedia 等），雖然涵蓋眾多領域之知識，但其內容主要以英文或者簡體中文為主。繁體中文知識往往呈現出零散之態勢，且缺乏系統性之整理。此外，這些開放知識庫在特定之專業領域（例如衛教、專業教育），同樣存在資料不完整，或者缺乏語義標註之狀況，難以直接滿足高準確度 LLM 檢索與推理之需求。基於上述種種原因，本研究決定自行構建繁體中文之專業知識圖譜，旨在確保知識內容之完整性、正確性，以及語義結構之嚴謹標註，同時提升圖譜對於繁體中文應用場景之適配性。

在知識圖譜的建置歷程當中，遇到諸多挑戰。針對此情形，我們將採取對應的措施。就資料源的挑選與獲取而論，繁體中文專業知識的散落情形如於學術論文、政府公開資料以及專業網站等多類途徑。本研究為求知識的可信度，運用網絡爬取技術自動獲取公開且可靠的資料源，與人工篩選及機器標注相結合，以提升知識質量，將內容不相關或質量欠佳的資料予以篩除。至於知識的結構化與標準化方面，原始文本資料大多呈非結構化狀態。本研究借由大語言模型及自然語言處理技術，把文本轉化為結構化的圖譜知識。具體而言，運用命名實體識別（Named Entity

Recognition, NER) 、關係抽取 (Relation Extraction, RE) 以及語義分類等方法，從爬取的文本裡識別重要實體及其關係，構建三元組事實。為提升一致性，統一本體架構與命名規則的制定不可或缺，以此保障不同來源知識整合後語義的一致性。而在知識驗證與一致性管理上，不同來源的知識或會呈現矛盾或不一致之態。本研究首以黑名單機制篩除明顯不相關或質量存疑的實體與關係。其後，藉助基於 RoBERTa 的中文詞向量嵌入模型計算實體名稱間的餘弦相似度，對潛在同義或變體的實體展開比對（相似度閾值設定為 0.85）；與此同時，應用 Record Linkage 技術（採用 Jaro - Winkler 演算法，閾值約 0.90）發現並合併名稱不同但指涉同一概念的實體。經由上述步驟，達成實體名稱的標準化與融合，確保知識圖譜內不存在重複或衝突條目。定期維護與更新機制的建立亦頗為重要，此舉旨在持續引入最新的領域知識，修正潛在錯誤，維持知識圖譜的時效性與可靠性。

1-2研究目標

針對大語言模型幻覺問題，本研究提出一種解決方案，此方案以知識圖譜與檢索強化二者的結合為核心要點。針對達成該方案所指向的目標，設立幾項具體的研究目標：

- (1) 自行建置糖尿病衛教知識圖譜用於繁體中文糖尿病衛教服務：基於繁體中文開放知識圖譜資源現狀之短缺，本研究將自行建置一知識圖譜。此圖譜以繁體中文呈現，專注於糖尿病衛教領域。需確保知識內容兼具完整性及正確性，並給予嚴謹語義標註。此標註意在支援後續之語義檢索及推理。
- (2) 知識的擷取以及結構化的處理：借由網絡爬蟲技術和大語言模型技術，以自動化方式對糖尿病相關的權威知識源進行擷取。憑借自然語言處理方法，像是命名實體識別 (NER) 、關係抽取 (RE) 以及語義分類等，把非結構化文本轉化為結構化的知識圖譜三元組結構。構建統一的本體架構以及知識表示形式，為知識圖譜的搭建提供技術基石。
- (3) 本研究方法之於幻覺問題的測試：對其實踐價值之驗證，表現於真實應用場景下的測試，涵蓋個人化健康管理建議提供以及即時問答方面。準確性、實用性、可靠性，這些指標，是對該聊天機器人在糖尿病衛教應用中的評估關鍵資訊。
- (4) 本研究知識圖譜建置與 RAG 整合技術之歸納，以此塑造出一套可複製、可拓展的方法論——可複製的知識圖譜強化 RAG 方法論。未來期望，該方法

論或能應用於諸如衛教、法律、科技、金融等高準確度需求領域，提升人工智慧於專業應用範疇與可信度和實用價值。

此機制融合知識圖譜與檢索增強生成技術，為 LLM 紿予更為可靠的知識支撐，以此提升其於高精度應用場景下回答的可信度、可解釋性以及有效性。在未來，這套方法有望為人工智慧在高專業、高可靠性領域的落地應用提供有力支撐與參考。

第二章文獻探討

本研究核心技術與理論基礎，含知識圖譜、大語言模型等關鍵領域，於本章被綜合回顧及解析。知識圖譜借「實體-關係-實體」三元組將知識結構化，為資訊檢索及高階語義推理築牢根基。大語言模型憑深度學習與大規模預訓練技術，在多語言任務中展現非凡效能，可生成時的幻覺問題，對系統可信度與應用安全性構成挑戰。檢索增強生成技術為解此難題應運而生，借動態整合外部知識，降低生成錯誤資訊風險。智慧型代理人概念進一步融合上述技術，於健康管理、智慧問答等領域展應用潛力。經對上述文獻及技術發展脈絡綜述，本章為後續系統設計與實證研究奠定理論基礎，亦指明自動化知識整合及深層語義推理未來發展方向。

2-1知識圖譜

知識圖譜（Knowledge Graph, KG）是一種以圖結構來表達實體及其相互關係的知識庫形式。其核心概念是透過「實體 - 關係 - 實體」三元組將知識加以結構化，而這些結構化的知識能支援高階的語意檢索與知識推論。此技術在大幅提升資訊檢索精確度的同時，也為智慧問答系統、推薦系統及語意搜尋等應用奠定了堅實的知識基礎。知識圖譜的發展可追溯至早期協作式圖形資料庫的研究。Bollacker 等人提出的 Freebase 系統，透過協作方式建立結構化知識庫，整合多元來源的知識，並以圖形化方式呈現人類知識[13]。此舉不僅為知識圖譜的實際應用鋪路，更激發了後續大規模知識庫建構與應用研究的興起。知識圖譜的構建大致包含三大技術環節。首先是知識獲取，主要從異質性資料來源中抽取實體、屬性與關係，過程中常需面對資料多樣性及無關內容的挑戰。其次是知識融合，重點在於辨識同一實體的不同表述，進而消除冗餘與歧義，完成資料的整合處理。最後是知識加工，包括本體建構、知識推論與品質評估，以提升知識圖譜的完整性與準確性。[14] Paulheim 對於知識圖譜精煉議題進行了詳盡探討，系統性地分析各類方法的優缺點及其對應的評估準則，為學術界提供了參考架構。在知識圖譜的發展現況、應用場景及未來挑戰方面，Hogan 等人則從宏觀視角進行了系統性整理[15]，指出在大數據與人工智慧技術持續進展下，知識圖譜於跨領域應用中具有廣闊前景。雖然在知識抽取、融合與加工等領域已有豐碩成果，但仍面臨如開放域語料的多義性、不相關資料干擾及知識自動更新等挑戰，高效且精準的演算法與系統解決方案仍屬迫切需求。未來的研究方向可能會更進一步聚焦於自動化處理、跨領域知識整合以及深層語意推論技術的提升，藉此推動知識管理的全面升級。

2-2大語言模型

近年在自然語言處理領域一個革命性技術大語言模型（Large Language Models，LLM）出現。借由深度學習與大規模語料預訓練，其具備處理多種語言任務能力，如文本生成、理解、翻譯及問答等。多數此類模型採用 Transformer 架構，該架構核心的自注意力機制，可捕捉語言中複雜依存關係與隱含語義[16]。此機制不但致使語言模型性能大幅提升，還為後續技術發展奠基。於 Transformer 基礎上，Devlin 等人提出 BERT 模型，借雙向編碼器預訓練策略，顯著提升各類語言理解任務效能[17]。Brown 等人發表的 GPT - 3 模型，在少量提示學習（few - shot learning）情境下，展現強大生成能力，成為跨多任務處理關鍵里程碑[18]。Bommasani 等人在綜述中提出“基礎模型”概念，探討大語言模型在不同應用領域的機遇、風險以及倫理安全議題，為模型未來發展提供重要理論參考[19]。大語言模型發展伴隨高昂訓練成本、龐大能耗需求與模型偏誤問題。這些挑戰在實際應用中，對硬體資源與環境保護構成嚴峻考驗，同時引發關於模型可解釋性與倫理問題的廣泛討論。未來研究要在提升模型效能和降低資源消耗間尋求平衡，還需發展更透明、可信的模型解釋方法，推動人工智慧技術持續進步。

2-3幻覺問題

儘管大語言模型在諸般自然語言處理任務之中，展現出令人矚目的強大能力。然而，模型生成文本之際，所衍生的幻覺問題（hallucination），無疑成為限制其實際應用效能的首要問題。幻覺問題，意指模型所生成的內容，與輸入資訊或真實答案相悖，甚或全然虛構。此情形於自動摘要、對話生成等具體場景之下，對資訊的正確性以及決策支持功能，均產生不容忽視的負面影響。Maynez 等人，在針對抽象式摘要任務開展深入研究時，明確指出：當模型欠缺充分的事實約束條件之時，便易於生成缺乏來源文本支撐的虛構內容，進而致使摘要的可靠性大打折扣 [20]。Lin 等人，則借助 TruthfulQA 測試集，針對模型在多任務環境之中呈現的幻覺現象，展開量化分析。其結果表明，即便是架構先進的大型模型，亦難以全然避免生成錯誤資訊 [21]。Kryściński 等人，通過引入多項評估指標，驗證了不同模型以及訓練策略之下，幻覺問題的發生概率存在顯著差異 [22]。就當前階段而言，學界針對幻覺問題，業已提出諸如強化事實檢索、借助外部知識輔助，以及調整模型訓練目標等多種解決策略。然而，鑑於生成模型內部表徵與外部知識整合過程的高度複雜性，如何從根本上減少幻覺現象的發生，仍有待進一步開展跨領域的深入研究與探索。

2-4 檢索增強生成

關於檢索增強生成，即 Retrieval - Augmented Generation (RAG) 技術，其以彌補內部知識庫缺陷、提升生成內容正確性及事實依據為目標，途徑是外部檢索系統與生成模型的結合。此技術呈現於兩個層面：其一，在預訓練階段憑借外部知識擴充知識庫，如 REALM 模型借助大規模外部文本檢索提升模型知識程度[24]；其二，在生成階段動態檢索相關資訊，並使之與生成模型相結合，進而產出更契合事實的回答[23]。Lewis 等人的發現表明，RAG 模型因檢索模組融入生成過程，可依使用者查詢動態調整生成策略，有效降低幻覺現象的發生概率[23]。檢索增強生成在知識密集型應用中優勢明顯，然而，檢索品質、知識融合策略以及運算延遲，是技術應用裡亟待處理的核心問題。未來研究或聚焦於提升檢索精確度、優化知識融合機制以及降低系統延遲，推動此技術在更廣泛領域的運用。

2-5 智慧型代理人

智慧型代理人是具備自主感知、決策以及行動能力的謂為智慧型代理人 (Intelligent Agents)。於智慧城市、電子商務、健康管理等範疇，其理論與技術獲廣泛應用。依 Russell 與 Norvig 之界定，智慧型代理人可據環境資訊作出判斷，制定並施行行動計劃 [25]。Wooldridg 從多代理人系統視角，強調代理人機協作與資訊共用，指出在複雜環境裡，借協同作業能有效達成系統目標 [26]。健康衛教領域，Bickmore 等人研究證實，具備高度互動性及定製化建議之虛擬代理人，可顯著提升使用者對健康資訊之信任與參與程度，卻也面臨資訊更新及準確性維護之挑戰 [27]。本研究以糖尿病衛教機器人為例，融合大語言模型、知識圖譜以及檢索增強生成技術，選用 Line Bot 作為互動介面，建置一套智慧型代理人架構。此系統不但整合先進語言生成技術，且結合外部資訊檢索，提供健康管理建議時可參照最新資料，進一步提高資訊正確性與互動性。技術不斷進步之下，智慧型代理人於衛教及其他應用領域將發揮愈發關鍵功用。

2-6 LLM Graph Builder

在自然語言處理任務中展現優異表現的大語言模型 (LLM)，卻因高度不透明的黑箱問題，於知識表徵及推理進程存在欠缺。研究顯示，LLM 易生成虛構 (hallucination) 資訊且匱乏領域特定知識，此類問題對模型可靠性構成不利影響 [19]，[20]，[21]。學界鑒於此，著手將知識圖譜引入 LLM 應用，以結構化形態組織實體與關聯，藉此彌補 LLM 的認知短板[31]，[32]。Kau 等人 (2024) 表明，以『結構化』方式引入知識圖譜，可捕捉實體關係，強化模型的可解釋性與推理能力

[31]; Pan 等人（2024）亦提出，統一運用 LLM 與知識圖譜，有利於增進模型的事實性與可信度[32]。建置檢索增強生成（RAG）系統時，單純憑藉向量檢索，常難以有效回應跨文檔、多跳推理問題；學者因而提出 GraphRAG 等途徑，藉由建置知識圖連接隱含概念，為 LLM 推理供給更豐富的上下文[23]，[29]。在此類需求的促使下，“LLM 圖形建置器”（LLM Graph Builder）應運而生，其欲藉助 LLM 自動將非結構化文本轉化為結構化知識圖譜，達成推理流程透明化與知識體系化建置的目的[28]–[30]。具體來說，LLM Graph Builder 工作流程涵蓋多個步驟[29]，[30]。系統先將上傳的各類文檔表徵為圖中的“Document”節點，利用 LangChain 的文檔載入器把每篇文檔內容拆分為若干文本片段（chunks）[29]。針對高度相似的片段節點，系統新增 SIMILAR 關係，建置基於 kNN 的詞彙圖[29]；同時，計算並留存每個文本片段的向量嵌入（embeddings），為後續向量檢索提供便利[29]，[30]。隨後後端調用 LangChain 提供的 LLMGraphTransformer 模組，通過結構化提示或工具調用的形式，讓 LLM 解析每個文本片段，識別並提取其中的實體與關係，進而將其建模為圖中的節點與邊[28]，[29]。最終，所有抽取的節點與邊皆存儲至 Neo4j 圖資料庫，且每條關係均連結回對應的原始文本片段[29]，[30]。使用者還能預先界定圖譜結構（諸如節點或關係的類型）以約束抽取成果，提升圖形的一致性與可控性[28]，[29]。在應用場景方面，LLM Graph Builder 可支持各類需要結構化知識的任務。例如在金融領域建置財務知識圖譜，能夠自動識別報表中的關鍵實體（如收入、資產等）及其相互關係，並經可視化簡化複雜結構，進而支撐決策分析、風險評估或欺詐偵測等任務[33]。由 Graph Builder 生成的知識圖，可驅動圖端檢索系統：如同 GraphRAG 方法所宣導的，結合知識圖譜與向量檢索，將圖中隱含的多跳關係作為上下文輔助 LLM 推理，以獲取更準確且可解釋的回答[23]，[29]。如 Kumar（2025）所指，這類系統尤其適用於資訊檢索與內容摘要等場景[30]；相較於純向量檢索，基於圖的檢索能提供更豐富的上下文資訊，從而提升回答的準確性與可解釋性[29]，[31]。與其他相似工具相比，LLM Graph Builder 在定位上獨具特色。LangChain 框架自身提供了 LLMGraphTransformer 模組，開發者可在程式中直接建置知識圖[28]；Neo4j 的 GraphRAG 則著重將知識圖與向量檢索相結合，以支援生成式 AI 的多跳推理[29]。相較之下，LLM Graph Builder 作為端到端開源應用，不僅整合了上述技術，還提供可視化介面與交互方式，降低使用門檻[29]，[30]。有別於傳統需大量規則工程的知識圖建置流程，這些新興的 LLM 驅動工具（含 LLM Graph Builder）藉助生成式模型自動提取實體與關係，並允許使用者自定義圖譜模式規範輸出，提高抽取結果的一致性與可靠性[28]，[29]。

第三章 研究方法與實作設計

本章節將對本研究選用的知識圖譜建置方法以及系統整合架構予以闡釋。涵蓋資料擷取與語料建置，知識抽取，知識融合連同圖譜生成，還有系統整合設計與實驗設計等方面。從網路資料獲取原始語料，藉助大語言模型抽取結構化知識，實施知識融合與清理，直至生成知識圖譜並整合進對話系統的步驟，皆在整體流程範疇內。後續章節將詳盡敘述具體架構與流程。具體方式如下圖所示：

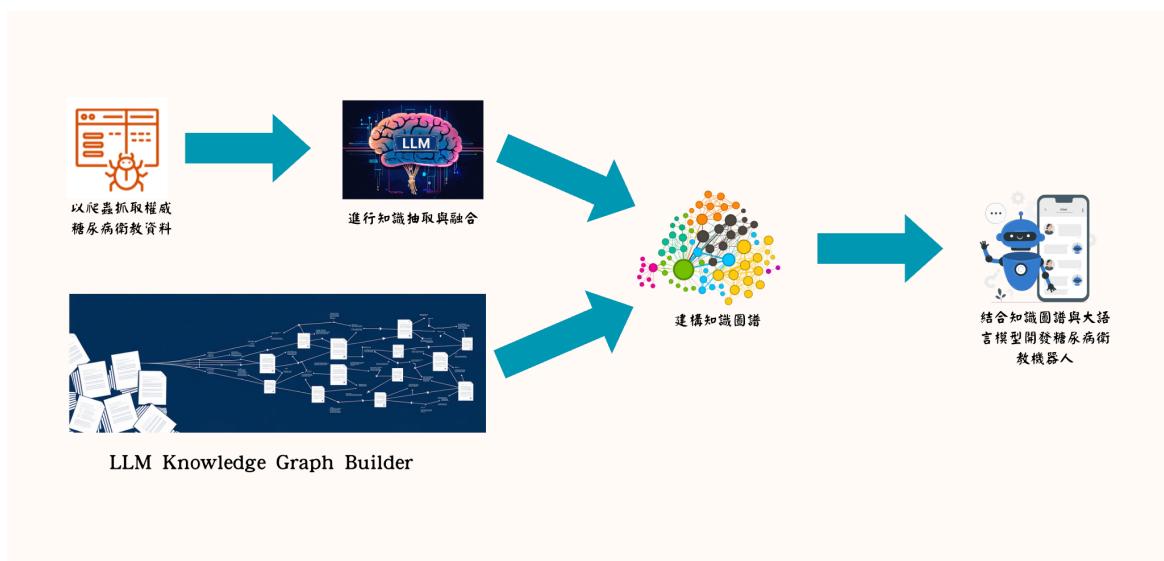


圖1 本系統建立知識圖譜之流程圖

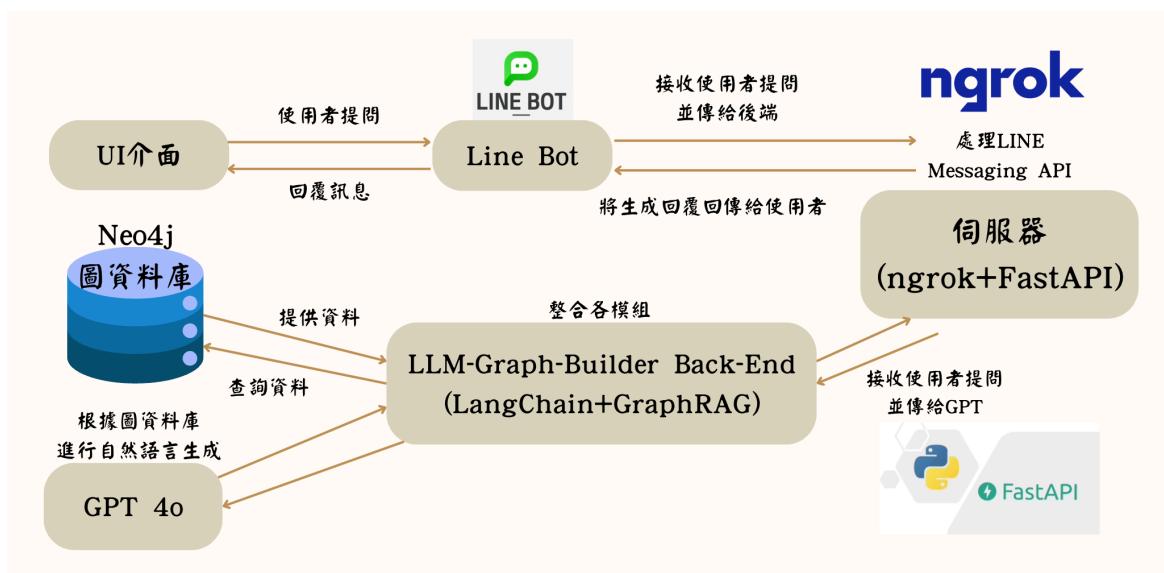


圖2 系統架構圖

本研究在知識圖譜建置方面的整體流程，由圖一予以呈現。以糖尿病相關衛教文本資料作為起點，借由網路爬蟲技術完成資料擷取，而經人工與自動化途徑，實

現資料清理及結構化處置。同時 LLM Graph Builder 工具亦將被用於將非結構化文本轉化為實體與關係構成的三元組，一同導入 Neo4j 圖資料庫成為可用於檢索與推理的知識圖譜。從資料蒐集、知識抽取直至圖譜生成的完整歷程，由該流程加以闡明，進而保障知識庫擁有可擴充性及一致性。

在系統架構設計整體問答的呈現方面，如圖二所示。LINE Bot 作為前端使用者介面，藉助 Messaging API 來實現資訊的接收與傳送。應用框架選取 Flask 作為後端，與 LangChain 裡 GraphCypherQACChain 模組相連接，該模組的職責是把使用者問題轉化為 Cypher 查詢，進而對 Neo4j 知識圖譜進行檢索。Neo4j 承擔著儲存與管理糖尿病領域結構化知識的任務。LLM（例如 GPT-4o）負責整合查詢所得結果，並生成自然語言形式的答案。借助 ngrok 與 FastAPI 完成伺服器串接，以此確保外部使用者能夠即時獲取系統服務。知識圖譜、LLM 以及應用層模組之間的互動方式，在本架構圖中清晰展現。

3-1 資料擷取與前處理

在糖尿病衛教領域知識圖譜建置上，本研究初始採用自建流程。主要資料來源鎖定為中華民國糖尿病學會官方網站以及美國糖尿病學會。獲取資料途徑有二：其一為自主編寫爬蟲程式；其二是運用 LLM Graph Builder 這一能自動將資料來源轉化為知識圖譜的工具。此處著重闡述自主編寫爬蟲的詳細流程。對與糖尿病指引、衛教文章、臨床照護及教育研究相關內容予以系統性提取，剔除導航功能表、頁尾版權及非正文範本，藉由可讀性抽取及範本去除技術清理雜訊。為確保資料品質，進一步開展規範化工作（涵蓋字形與標點一致性、單位統一）、去重及版本管理（借助時間戳記與校驗碼）。在維持來源可追溯性的基礎上，將清理後的文本以結構化格式存儲為 JSON（諸如 diabetes_knowledge.json），其段落包括主要內容、標題、分類、來源 URL 以及獲取日期等元資料，並建立索引以便後續處理。基於此語料庫，完成初步的知識抽取、同義詞整合以及實體對齊（涉及名稱標準化、別名匯總、概念階層化），據此生成圖譜初版。結合大語言模型（LLM）製作圖譜輔助的對話原型，回答流程以圖譜檢索與推理為先，LLM 生成補述為後，同時回溯來源節點與段落位置，以此降低幻覺並便於稽核。整體自建流程搭配治理設計（包含爬蟲規則與抽取版本封存、變更日誌），旨在確保可重複性以及法規/倫理合規（僅處理公開衛教/指引文本，不涉及個人資訊，明確標示非臨床診斷用途）。

在系統規模擴張及更新頻率需求提升的態勢下，自建流程於擴增速度以及維運成本方面遭遇瓶頸情形。LLM Graph Builder Backend 作為工具化輔助被本研究後續引入。此後端於合法或授權文本範疇內，自動（或半自動）生成候選節點、屬性與關係。關於本研究圖 1 所示，前段自建著重於嚴謹性、可解釋性與可重現性，而後段 Backend 在原則穩固的基礎上加速擴張，最終建置出一個規模更為完整且能夠支撐對話應用的糖尿病領域知識圖譜，此乃本小節所述流程之對應內容。

3-2 自動文本辨識與知識抽取

大語言模型如 ChatGPT、Gemini 與提示工程（prompt engineering）之結合方式。其目的為以自動化方法從文本內抽取具體實體、屬性以及實體間之關係，並把結果以符合 JSON 格式的資料加以輸出，為隨後的自動化解析以及知識圖譜構建提供便利。為實現此目標，系統設計主要涉及下述步驟。原始文本的切割，在本研究中率先展開。因為單篇文章或會超出模型一次所能處理的最大字符數。具體做法為，系統以上限為預先設定的最大字符數，對長文本實施分段切割，同時盡可能維持完整語意，保證各段落間內容連貫且關鍵資訊無缺。此處理方式，對提高後續模型生成回應的準確性有益，且能使資料解析階段更高效。大語言模型被本研究用以實現自動文本辨識和知識抽取。借助精心設計的 prompt 指令，系統明確要求模型從文本中抽取具體實體、屬性以及實體間的關係，並強制限定輸出格式需符合 JSON 結構。提示工程設計的關鍵之處在於，在提示中清晰闡述所需抽取的資訊類型，像疾病名稱、症狀、藥物等，且指定關係類型，比如“引起”“導致”，以及屬性，例如“病因”“檢查標準”。格式方面要求模型僅回應契合 JSON 結構的結果，避免輸出額外文本或註釋，以利後續自動化解析。如此的提示設計策略，可有效引導模型生成標準化、結構化的資料，進而提升整體系統的可重現性與精確性。模型生成回應後，系統運用正則表達式對返回文本掃描，提取完整且有效的 JSON 內容。此過程包含：JSON 抽取，即利用正則表達式識別並提取符合 JSON 格式規範的區塊，確保回應不含有額外非結構化文字；格式驗證，將提取出的 JSON 字串解析，若解析失敗便記錄錯誤並跳過該文章，反之則將有效資料納入後續整合流程。此機制保證了最終輸出的資料能直接用於後續知識抽取模組，並減少因格式錯誤引致的處理延誤。解析後的 JSON 資料含有實體，如疾病名稱、症狀、藥物，關係，像“引起”“導致”，以及屬性，例如“病因”“檢查標準”。系統會對各篇文章中提取到的資料進行累積與整合，形成一個統一的資料庫。該資料庫作為知識抽取後續處理流程的基礎，會進一步用於知識圖譜構建以及智慧應用的開發。

綜上所述本節說明了一個自動文本辨識暨知識抽取流程，經由文本切割、提示工程所驅動的大語言模型生成，以及正則表達式提取與 JSON 格式驗證得以構建。此流程對資料處理效率有所提升，抽取結果的結構化與可重現性亦得到確保，為後續知識圖譜的建置供給了堅實的資料支持。

表一實體與實體名稱

實體類型	實體範例
疾病	高血壓以及高脂血症，与糖尿病所涵盖的 1 型糖尿病、2 型糖尿病、妊娠糖尿病，共同构成特定的病症范畴
症狀	多尿、口渴、體重減輕、疲勞、視力模糊、傷口癒合緩慢
危險因子	遺傳因素、肥胖、缺乏運動、不健康飲食、壓力
檢查項目	對血糖的檢測、以糖化血色素 (HbA1c) 為指標的分析、口服葡萄糖耐量所進行的試驗流程、血脂方面的詳細檢查操作，以及血壓的測量工作
控制方法	飲食控制、運動療法、胰島素注射、胰島素幫浦、戒菸
藥物	胰島素、二甲雙胍、磺醯脲類藥物、DPP-4 抑制劑
併發症	糖尿病足、糖尿病腎病、糖尿病視網膜病變、心血管疾病、神經病變
生活方式	飲食、運動、心理情緒、旅行、開車
預防措施	健康飲食、定期運動、定期檢查、體重控制

表二 關係名稱與定義

關係名稱	示例
引起	肥胖 引起 2 型糖尿病
表現為	糖尿病 表現為 多尿
控制	胰島素注射 控制 1 型糖尿病
檢查	HbA1c 檢查 診斷 血糖控制情況
導致	糖尿病 導致 糖尿病足
預防	健康飲食 預防 糖尿病
控制	運動 控制 血糖水平
屬於	1 型糖尿病 屬於 糖尿病

表三屬性與示例類型

屬性類型	示例
疾病名稱	糖尿病
疾病定義	一種由胰島素分泌或作用缺陷引起的代謝性疾病
疾病病因	胰島素分泌不足、遺傳因素、環境因素
疾病症狀	多飲、多尿、多食、體重減輕
疾病診斷標準	空腹血糖 $\geq 7.0 \text{ mmol/L}$

藥物名稱	胰島素、二甲雙胍
控制方法	胰島素、飲食控制、運動療法
預防措施	健康飲食、定期運動
藥物副作用	腹瀉、惡心、腹痛
藥物劑量	二甲雙胍每日 500 mg 至 2000 mg
藥物用法	口服，隨餐服用
適用人群	2 型糖尿病患者
控制持續時間	長期或終身控制
痊癒成功率	70% - 90%，視情況定

3-3知識融合與圖譜建置

本節說明融合前兩階段所獲之結構化知識。此知識源自文本辨識、實體抽取及知識抽取，包括實體、屬性及關係資訊。其目的在於建構連貫且語義一致的知識圖譜。不同來源抽取的知識，常有冗餘、歧義及雜訊。故知識融合階段的核心，是消除不一致性，對相同或相似知識單元進行對齊與合併，提升知識庫整體質量與可用性。而後，依融合結果的結構化呈現，搭建成基於圖資料模型的知識圖譜，為知識推理與應用奠定根基。關於實體鏈接與消歧。針對不同來源或抽取流程所得實體，本研究運用基於 RoBERTa 的中文嵌入模型（hfl/chinese - roberta - wwm - ext），計算實體名稱向量嵌入相似度(1)。借助餘弦相似度(2)、Jaro-Winkler 相似度(3)等來進行對齊。通過算法，將語義相近或相同的實體，合併為標準化表述。為減少不相關資料干擾，融合策略引入黑名單過濾機制，剔除與主題無涉的冗餘實體。至於關係與屬性合併。實體鏈接完畢後，對涉及同一對實體的多個關係描述，經由語義匹配方法，予以歸納整合，生成統一且高可信度的關係資訊。屬性資訊亦以類似方式處理，經正則化和標準化步驟，確保資料格式一致，消除重複及冗餘屬性。

嵌入向量計算：

$$E = \frac{1}{n} \sum_{i=1}^n h_i \quad (1)$$

對於實體名稱由 n 個詞表示，其中 n 為實體名稱中的詞數， h_i 為第 i 個詞的隱藏狀態向量，E 是實體的平均嵌入向量。

餘弦相似度計算公式為：

$$\cos(\theta) = \frac{A \cdot B}{|A| \times |B|} \quad (2)$$

A, B 為兩個實體的嵌入向量。 $|A|, |B|$ 為向量的歐幾里得範數。 $A \cdot B$ 為向量的點積。

Jaro-Winkler相似度計算公式為：

$$JW(s_1, s_2) = \text{Jaro}(s_1, s_2) + l \cdot p \cdot (1 - \text{Jaro}(s_1, s_2)) \quad (3)$$

s_1, s_2 為兩個實體名稱。 $\text{Jaro}(s_1, s_2)$ 為 Jaro 相似度。 l 為公共前綴的長度，最大為 4。 p 為縮放係數，通常設定為 0.1。

在知識圖譜建置階段，基於知識融合所獲統一知識表示，涵蓋如下主要步驟。資料模型設計方面，依據融合後的實體、關係以及屬性資料，展開知識圖譜資料模型的設計工作。該模型以「實體 - 關係 - 實體」三元組作為基本構成單元，並依照實際需求拓展至屬性 - 值對表示。此部分需將資料的多樣性與層次性納入考量，確保知識庫於後續推理及查詢時具備優良的可擴展性與表達力。圖資料庫存儲環節，挑選適宜的圖資料庫（本研究採用 Neo4j）作為知識圖譜的存儲平臺。把融合後的知識資料轉化為圖資料模型，其中實體充當節點，關係作為邊，屬性則成為節點或邊的附加屬性。在資料導入過程當中，資料一致性與查詢效率必須得到保證，同時建立必要索引以加快圖查詢與推理速度。再來是質量評估與錯誤修正，於圖譜建置完畢後，針對知識圖譜展開質量評估，主要考察知識的完整性、準確性以及冗餘狀況。此環節一般採用人工抽查與自動評估相結合的方式，對存在歧義或錯誤的知識加以修正。經由反覆運算優化，最終生成一個語義清晰、結構合理的知識圖譜。應用介面與查詢服務，建置知識圖譜的應用介面，為基於查詢語言的高效查詢與推理服務提供支援。此部分致力於將知識圖譜轉化為實際應用工具，為諸如知識推理、問答系統等下游任務提供資料支援與邏輯基礎。在知識融合與圖譜建構的整個流程裡，系統架構需考慮模組間的高效整合以及資料轉送問題。整體架構一般劃分為：資料採集與抽取層，承擔多來源資料的抓取以及知識抽取工作。知識融合層，對抽取資料進行清洗、對齊以及融合，去除冗餘資訊。圖譜生成層，將融合結果轉化為圖資料模型並導入圖資料庫。查詢與推理層，提供查詢介面以及知識推理服務，支

撐各類知識應用。這種分層架構使各模組既能獨立優化，又能在整體上實現高效協同，並且便於未來擴展或引入新的知識來源與融合演算法。

3-4系統整合與智慧應用設計

在本研究裡，最終目標為由資料擷取與預處理、自動文本辨識與知識抽取以及知識融合與圖譜建構等各模組進行整合。借此建置完整的智慧衛教機器人系統，同時嘗試解決幻覺問題。知識圖譜建置方法耗時且難以快速擴充知識圖譜。有鑑於此，在知識融合與圖譜建構模組中，引入 LLM Graph Builder 工具，藉助大型語言模型將非結構化資料自動轉化為知識圖譜。由該系統建置的知識圖譜，會與原有知識圖譜融合，從而快速且有效地擴充並豐富整體知識庫。智慧機器人系統實現從使用者輸入、知識檢索與生成回應到最終反饋的全流程自動化處理。通過系統效能測試全面評估，驗證其在健康管理場景中的應用潛力與實用性。於系統整合方面，後端以 Flask API 作為核心仲介，達成各模組間資料交互與調用。Flask API 提供輕量級、靈活且可擴展介面，便於統一調度知識圖譜、LLM 模型及其他資料處理模組，實現各功能模組無縫連接。本系統結合 LangChain 框架，此框架用於協調大型語言模型（LLM）與知識圖譜間資訊交互。LangChain 能自動串聯自然語言處理、查詢與回應生成等流程，使系統依據使用者輸入問題，檢索相關知識並生成語義準確、可信回應。前端部分，本研究選用 Line Bot 作為主要使用者交互介面。Line Bot 提供即時、直觀用戶互動體驗，讓最終用戶可通過手機或桌面應用直接與智慧健康管理系統交流。為支援遠程測試與部署，利用 ngrok 工具公開本地 Flask 伺服器，使遠端使用者能跨網路訪問系統，以驗證系統在實際應用場景中的表現。

從使用者輸入起始，直至知識檢索、生成回應，乃至最終反饋這一全流程自動化處理的達成，得益於該整合系統多個模組的協同運作。具體流程涵蓋：

- (1) 解析使用者輸入：經由 Line Bot 提交問題之舉由用戶實施，後端 Flask API 負責對所輸入文本展開接收與解析工作。
- (2) LangChain 架構之於 LLM 與知識圖譜間進行協調這一活動，以達成知識檢索及回應生成之目的：基於使用者所提問題對相關知識開展檢索，隨之生成語義精準的回應文本。
- (3) 經由 Flask API 傳回前端的，是生成的回應；而最終結果的呈現予用戶，則由 Line Bot 負責。此即結果反饋。

針對系統效能考量，引入全面性的效能測試機制。在這一機制下，系統各模組交互延時、回應準確率以及穩定性等多層次、多指標的性能評估得以實施，為系統的後續優化奠定基礎。智慧型交談機器人系統整合方案的詳細闡述，自系統架構、前後端交互至自動化全流程設計等多個面向展開。Flask API 被選作後端中介，LangChain 用以協調 LLM 與知識圖譜，Line Bot 作為前端交互介面，基於此，系統達成從用戶輸入起始，經知識檢索、回應生成，至結果反饋的全流程自動化。該系統在具備較高可用性與實時性的同時，效能測試為其於健康管理領域的智慧應用提供了強有力的技術支撐。

3-5 實驗設計與方法

本研究之實驗設計以抑制幻覺問題為核心目標，將幻覺操作化為「生成答案中未被檢索證據支撐之斷言」（ ungrounded claims ），並以忠實度（ Faithfulness ）作為主要評估終點；上下文召回率（ Context Recall ）視為檢索覆蓋對忠實度的關鍵中介變項，答案相關度（ Answer Relevancy ）則作為貼題性之控制觀測。採配對對照之 RAG 架構：對同一題目，同步以純向量檢索（ Vector-only ）與圖-向量混合檢索（ Graph+Vector ）取得上下文，再以相同的 LLM 、提示模板與解碼設定生成答案；檢索配額（如 top-k 、上下文長度）與答案長度規則一致，以隔離「檢索策略」對輸出的影響。測試集為由領域知識圖譜語料構建之 35 題（涵蓋一般照護目標、血糖指標、併發症預防、生活型態建議等），確保任務與知識來源同域。評估流程如下：首先依權威來源建立每題理想答案的關鍵事實集合（ gold key facts ），據此計算 $Context\ Recall = r/R$ （其中 R 為所需關鍵事實總數，r 為被檢索上下文支撐之事實數）；其次，以題目要點集合比對生成答案覆蓋，計算 Answer Relevancy ；再將生成答案切為原子化斷言，逐一檢索對齊判定是否被上下文支撐，以被支撐斷言比例作為 Faithfulness 分數，並記錄極端低分樣本之占比以反映嚴重幻覺案例。標註採雙評與仲裁機制以提升一致性；R=0 的題型不納入召回率計算。統計分析以 Faithfulness 的成對比較為主（非常態時採 Wilcoxon 符號等級檢定），回報效果量與 95% 信賴區間；同時進行關聯分析以檢驗「召回→忠實度」之關聯性，並觀察忠實度與相關度之區隔（驗證「貼題≠可驗證」）；另做主題/跳數切片與失敗案例剖析，區分覆蓋缺口（檢索未涵蓋必要事實）與利用缺口（已檢索但生成未引用）兩類幻覺成因，作為後續自適應混權、證據導向重排與先證據後生成核驗等策略優化之依據。整體設計旨在以量化與可重現的方式檢定：知識圖譜介入是否能透過提升檢索覆蓋，實質降低未被支撐的生成斷言，從而達到抑制幻覺的目的。

第四章 實作驗證

在建置結合大型語言模型（LLM）與知識圖譜的問答系統方面，本研究的實作方法以及評估結果會於本章展開介紹。實驗驗證、語料來源、系統架構、測試流程、評估指標定義，還有統計結果分析，這些構成內容。最後做出小結。

4-1 知識獲取

本研究為建置糖尿病領域問答知識基礎，採用『兩階段、雙來源』資料獲取策略。兼顧前期方法可行性驗證需求，以及後期維運擴充需求（流程參考圖 1，系統架構參考圖 2）。以中華民國糖尿病學會『全民糖尿病觀測站』作為第一階段核心資料，憑藉爬蟲程式、大語言模型、實體對齊等技術，建置一套可重現知識圖譜，確保資料正確性與語境一致性。LLM Graph Builder Backend 於第二階段被導入，以美國糖尿病協會（ADA）官方內容作為主要擴充來源，將實體與關係自動化抽取嵌入後端流程，實現高頻更新與持續擴張目的，且維持前期手動標準結構與品質門檻。兩階段相互獨立且各司其職，前期著重方法與基準圖譜'正確性與可追溯性'，後期優化'效率與規模化維運'。

第一階段（前期）：撰寫爬蟲程式與語料整備我們以 Python（requests、BeautifulSoup）實作爬蟲程式，僅針對糖尿病指引、衛教文章等高關聯項目進行系統化擷取，並排除導覽、版權宣告與廣告等雜訊區塊，以提升語料的專業密度與主題純度。考量目標網站之頁面多以表格呈現（例如以 <table> 所構成之清單區塊；部分欄位以寬度屬性標示，如 th width="425" 放置標題與連結、th width="185" 放置文章分類），本研究據此訂製 DOM 選取與過濾規則，以穩定擷取「標題 / URL / 類別 / 主文」等欄位。主要步驟如下：1:發送 HTTP 請求並解析 HTML 2:鎖定含衛教內容之表格節點並過濾無關連結 3:逐篇抓取文章頁，擷取正文並去除不相關資料 4:人工抽查重複與過期內容，建立白名單標籤集以維持主題範圍 5:以 UTF-8 正規化並清洗殘留標籤 6:以 JSON 結構存檔為 diabetes_knowledge.json，納入必要中繼資料（來源 URL、擷取日期、類別、雜訊比對指標、版本號）。此階段輸出形成「可審核、可回溯」之繁體中文領域語料庫，作為後續抽取與圖譜建模的基準資料集。爬取的資料會存儲為 JSON 檔案如下：

```

1  [
2  {
3      "類別": "糖尿病簡介",
4      "標題": "什麼是糖尿病?",
5      "內容": "什麼是糖尿病?\r\n
6 },
7 {
8     "類別": "",,
9     "標題": "糖尿病的症狀",
10    "內容": "糖尿病的症狀\r\n
11 },
12 {
13     "類別": "",,
14     "標題": "造成糖尿病的原因及其危險因子",
15     "內容": "造成糖尿病的原因及其危險因子\r\n
16 }

```

圖 3 爬蟲程式抓取資料範例

診聽病人說，我沒有糖尿病啊，只是血糖有些高而已。那到底什麼是糖尿病呢？\n其實糖尿病是迄今為止人類發現的最古老的疾病之一，從古埃及約西元前1550年，就有「多尿」的疾病的文字記載。到了西元二世紀
五十；糖尿病除了會造成個人健康上的危害，也會對於國家財政上形成龐大的負擔，因此許多國家積極地在作「篩檢」的工作，希望能早期發現罹患糖尿病的病人，進而早期治療，減少日後併發症的發生；然而在確定是太愛吃了才會得糖尿病、都沒運動難怪會得糖尿病...。這樣長期被\"汙名化\"的結果，往往讓被診斷出糖尿病的人陷入沮喪。那麼，真的是這樣嗎？\n正常的身體藉由胰臟所分泌的胰島素來調控血糖，透過胰島！嗎？\n請千萬記住，造成傷害的是過高的血糖，藥物是幫忙身體控制血糖的工具之一，是幫手不是敵人！！糖尿病的出現是冰山一角，往往身體已經處在長期血糖代謝失調的狀態才逐漸變成糖尿病，即使開始罹患糖中的葡萄糖進入紅血球，和血紅素結合後，就形成糖化血色素。一般紅血球平均壽命為120天，葡萄糖附在血色素上不容易脫落，因此檢查血中糖化血色素的濃度，可以反映體內最近2-3個月的血糖控制情況。\\n
\n台灣兒童糖尿病關懷協會\\n第1型糖尿病\\n是因為胰臟胰島細胞被破壞而胰島素分泌絕對缺乏，所以第1型糖尿病患者需要皮下胰島素補充。胰島素的皮下注射方式可分為連續性皮下胰島素輸注(胰島素幫浦)、一天

圖 4 爬蟲程式抓取資料內容

4-2知識抽取

本研究將 Google Gemini 1.5 Flash 選作首要用於知識抽取的模型。借由提示工程這一方法，來達成使該模型能夠精準辨別實體（ Entities ）、關係（ Relations ）以及屬性（ Attributes ）之目的，並以標準化的 JSON 格式給出輸出成果。知識抽取具體流程為：發出針對 LLM 的請求，旨在讓其從文本之中實施知識提取操作。對模型的回應展開解析工作，務必保證輸出符合預先設定的 JSON 格式。針對 JSON 進行解析處理，並且按照實體、關係、屬性進行分類梳理。將所有文章抽取所獲結果

予以累積，建置完整的知識集合。知識抽取主要聚焦目標有，實體（ Entities ）方面：對衛教概念進行識別，諸如“糖尿病”“胰島素”“視網膜病變”等；關係（ Relations ）方面：對概念間的關聯予以識別，例如「糖尿病 → 或可引發 → 心血管疾病」；屬性（ Attributes ）方面：獲取屬性相關資訊，像“糖尿病的診斷標準 = 空腹血糖 ≥ 7.0 mmol/L”。鑑於 LLM 有可能生成錯誤格式或者不完整的回應，故而需要執行：JSON 格式驗證，保證模型回應屬於合法的 JSON 。對錯誤內容開展篩選，要是 LLM 回應內含有額外描述或者非結構化內容，即實施清理操作。倘若模型回應並非合法 JSON ，那麼藉助正則表達式嘗試提取 JSON 內容。要是 JSON 解析失敗，就跳過該筆資料並記錄錯誤相關資訊。若 JSON 中缺失實體、關係或者屬性，則標記為有待補充的資料。經過處理的知識資料會統一存儲至 processed_data.json ，為後續知識圖譜建置工作提供結構化輸入。在本階段確保所有資料已整理形成：完整的實體清單（涵蓋疾病、癥狀）。關係清單（疾病與癥狀、疾病與等）。屬性資訊（諸如診斷標準、藥物劑量等）。

```

5798   {
5799     "subject": "胰島素",
5800     "relation": "不足",
5801     "object": "糖尿病"
5802   },
5803   {
5804     "subject": "身體",
5805     "relation": "對胰島素反應不佳",
5806     "object": "糖尿病"
5807   },
5808   {
5809     "subject": "第一型糖尿病",
5810     "relation": "原因",
5811     "object": "胰島細胞遭到破壞"
5812   },

```

圖 6 知識抽取結果範例圖二

```

3791   {
3792     "entity": "糖尿病",
3793     "attribute": "盛行率",
3794     "value": "8%"
3795   },
3796   {
3797     "entity": "糖尿病患者",
3798     "attribute": "人數",
3799     "value": "140萬名"
3800   },
3801   {
3802     "entity": "血糖",
3803     "attribute": "標準值",
3804     "value": "70毫克/分升"
3805   },
3806   {
3807     "entity": "糖果",
3808     "attribute": "數量",
3809     "value": "3-5顆"
3810   },

```

圖 5 知識抽取結果範例圖一

本階段——第二階段，涉及 LLM Graph Builder 後端部署以及知識圖譜擴充的實踐。LLM Graph Builder 後端系統在此階段得以部署，其目的在於為知識圖譜的自動擴充提供支援。依據 Neo4j Labs 的闡述，LLM 知識圖譜建置器能夠把非結構化文本，像 PDF 檔、其他格式檔或者網頁內容之類，轉變為結構化的知識圖譜。後端服務藉由 Python 和 FastAPI 得以實現，於本地伺服器運行。為達成允許外部連接的目的，後端運用 Ngrok 建置安全通道，將本地服務映射至可公開訪問的 URL 。

該後端藉助 Neo4j 官方驅動程式與協定連接到本地運行的 Neo4j 圖形資料庫，以此實現對圖形資料的讀寫操作。資料擷取階段，本研究把 ADA 官方網站的美國糖尿病學會 (ADA) 2024 年版及 2025 年版《Standards of Care in Diabetes》(糖尿病照護準則) 以及中華民國糖尿病學會「全民糖尿病觀察站」之 Neo4j 資料庫內容轉換為內容來源。按照 Neo4j 官方說明，Graph Builder 後端的主要目的是從多種來源抓取內容，經由 LangChain 的 LLM Graph Transformer 處理，抽取實體與關係，最終以圖形檔形式存儲到 Neo4j 資料庫。針對每個擷取到的網頁內容，後端提取純文本，並將其輸入大型語言模型進行知識抽取。具體來講，藉助 Graph Builder 提供的處理流程（比如結合 LangChain 的 llm - graph - transformer 模組），對文本開展實體（節點）和關係（邊）的識別與抽取工作。語言模型的輸出被格式化為預先定義的 JSON 結構，此結構包含識別出的節點、關係及其屬性。隨後，後端解析該 JSON 結果，並通過 Neo4j 驅動程式執行 Cypher 查詢，把新提取的節點和關係寫入圖形資料庫。在此過程當中，系統會自動檢查並合併重複實體，防止冗餘。如官方流程所呈現的，一旦完成抽取，所有實體與關係會以節點和邊的形式儲存到圖譜中，且與它們來源的文本片段建立連接。經由這個流程，知識圖譜實現自動擴充，融入來自 ADA 官方網站的新領域知識。

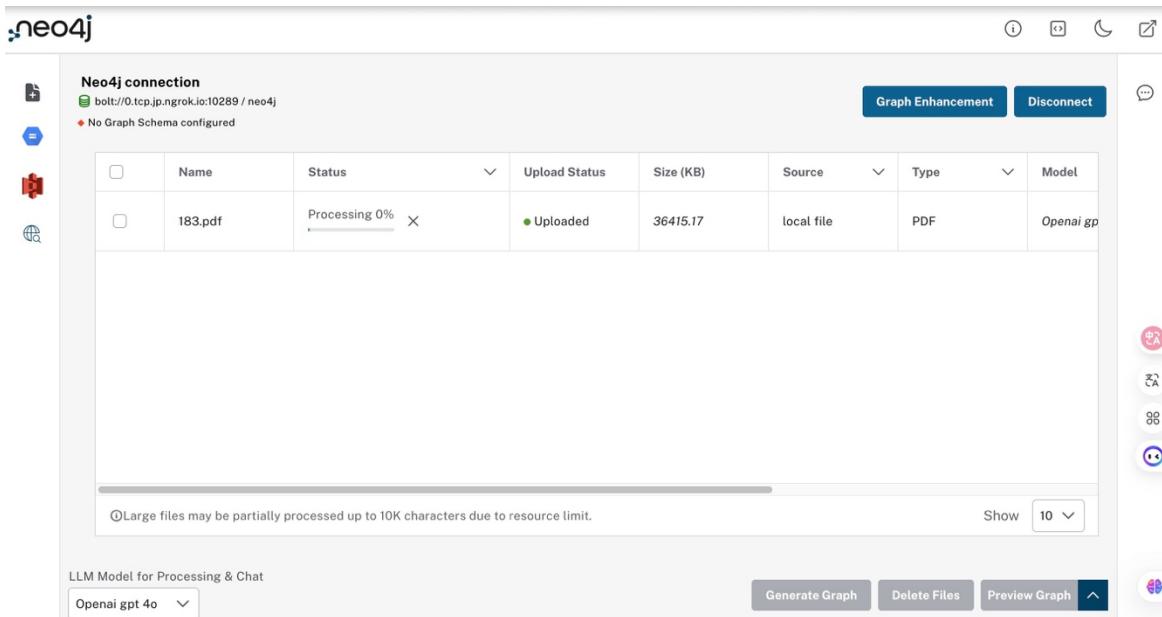


圖 7 LLM Graph Builder 操作畫面

4-3知識融合

本階段的核心目標，是將經整理的資料轉換為可導入 Neo4j 的 Cypher 語句，以此保證知識圖譜結構完整性與資料一致性，同時減少冗餘資訊，支持後續知識檢索及問答系統應用。主要工作涉及去除無關資訊、標準化實體名稱、合併重複實體、

規範關係類型。網頁爬取內容常含大量與糖尿病無關名稱、機構及個人名，因此本研究設計黑名單機制（Blacklist Filtering）。此機制通過比對實體名稱與預設黑名單關鍵字，過濾無關詞彙。若實體名稱在黑名單中，將自動排除；若關係主體或客體屬黑名單實體，該關係亦不予記錄。同時，對可能誤判的實體做標記因資料來源多元，實體名稱可能有同義詞、拼寫變異或近似詞。本研究採用基於 RoBERTa 的中文嵌入模型（hfl/chinese-roberta-wwm-ext）計算實體名稱向量嵌入(1)，利用餘弦相似度（Cosine Similarity）(2)比對。若兩個名稱餘弦相似度超 0.85，視為同一實體合併。研究還運用記錄鏈接（Record Linkage）技術，利用 Jaro-Winkler 相似度(3)判斷兩實體名稱相似性。相似度超 0.90 時，將低頻變異拼寫統一替換為標準名稱。關係與屬性格式規範方面，鑑於不同資料源描述方式可能不同，本研究實施關係標準化，移除特殊符號，將關係名稱轉換為駝峰式命名法，保證所有關係名稱一致。同時，嚴格轉換屬性格式，使數值屬性（如診斷標準）遵循統一格式（如“空腹血糖 $\geq 7.0 \text{ mmol/L}$ ”），統一疾病類型等類別屬性寫法。為方便後續資料導入，本階段設計自動化轉換模組，將標準化的實體、關係與屬性資料生成符合 Cypher 語法查詢語句。具體操作有：保證每個實體僅建立一次（處理特殊字符轉義）；生成關係語句時驗證雙方節點存在；將屬性資料寫入對應節點。最終，所有整理後的資料以 JSON 格式存儲，生成 Cypher 語句文件，供直接匯入 Neo4j，構建高品質的知識圖譜。本階段設計結合了實際程式實作中的黑名單過濾、向量嵌入、醫學命名實體識別（基於 BERT 的中文醫學模型）及記錄鏈接技術，保證了知識融合嚴謹性與資料一致性，為後續知識檢索與智慧問答系統實現提供堅實可靠基礎。

所有結構化的知識圖譜資料都將儲存在refined_data.json中。形式及內容如下：

```
{
  "entities": [
    {"name": "糖尿病", "type": "Disease"},
    {"name": "多尿", "type": "Symptom"}
  ],
  "relations": [
    {"subject": "糖尿病", "relation": "表現為", "object": "多尿"}
  ],
  "attributes": [
    {"entity": "糖尿病", "attribute": "診斷標準", "value": "空腹血糖  $\geq 7.0 \text{ mmol/L}$ "}
  ]
}
```

存儲方式：

1. JSON 格式存儲，以便未來擴充與查詢。

2. 前期之第一階段，目標在於產出 Cypher 語法文件，即 output.cypher，此文件將被直接引入 Neo4j 以完備知識圖譜之構建。至於後期的第二階段，LLM Graph Builder 所進行的操作是，在生成相關內容後，即刻導入本地的 Neo4j 資料庫。

在專案將結案時在圖資料庫中共計有：節點6132個，關係17239條，屬性41個。

```
5604 MATCH (a {name: "升糖指數飲食"}) , (b {name: "碳水化合物"})  
5605 MERGE (a)-[:`衡量`]->(b);  
5606  
5607  
5608 MATCH (a {name: "第2型糖尿病"}) , (b {name: "妊娠糖尿病"})  
5609 MERGE (a)-[:`危險因子`]->(b);  
5610  
5611  
5612 MATCH (a {name: "西點麵包"}) , (b {name: "油脂"})  
5613 MERGE (a)-[:`含有`]->(b);  
5614  
5615  
5616 MATCH (a {name: "飯前血糖"}) , (b {name: "240mg/dl"})  
5617 MERGE (a)-[:`高於`]->(b);  
5618  
5619  
5620 MATCH (a {name: "第2型糖尿病"}) , (b {name: "胰島素分泌量少"})  
5621 MERGE (a)-[:`特徵`]->(b);
```

圖 8 使用Cypher 語法文件生成圖譜

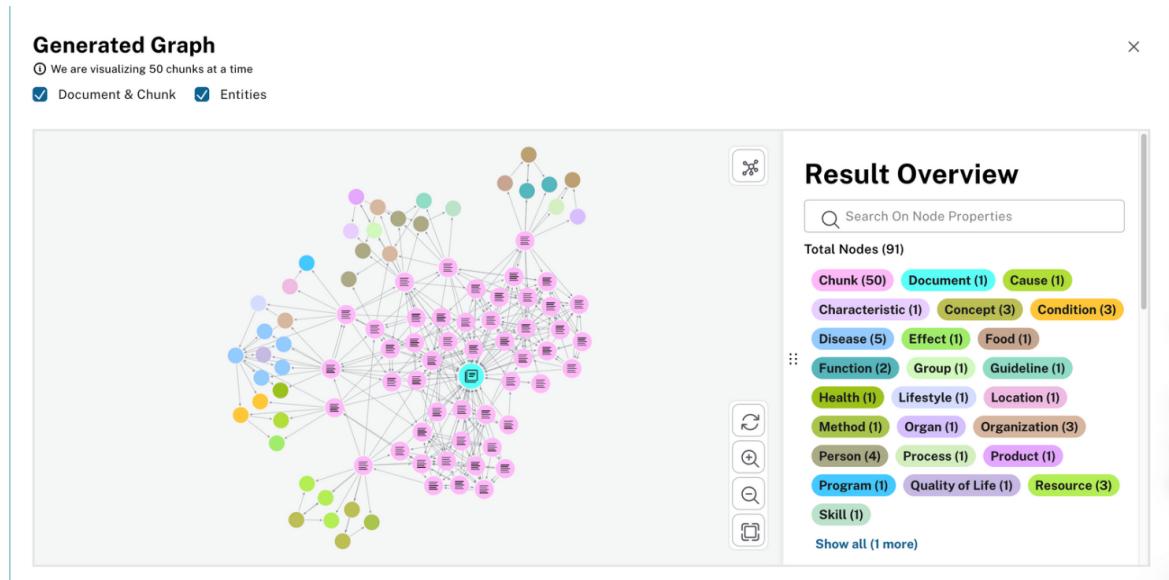


圖 9 以LLM Graph Builder 建置知識圖譜範例

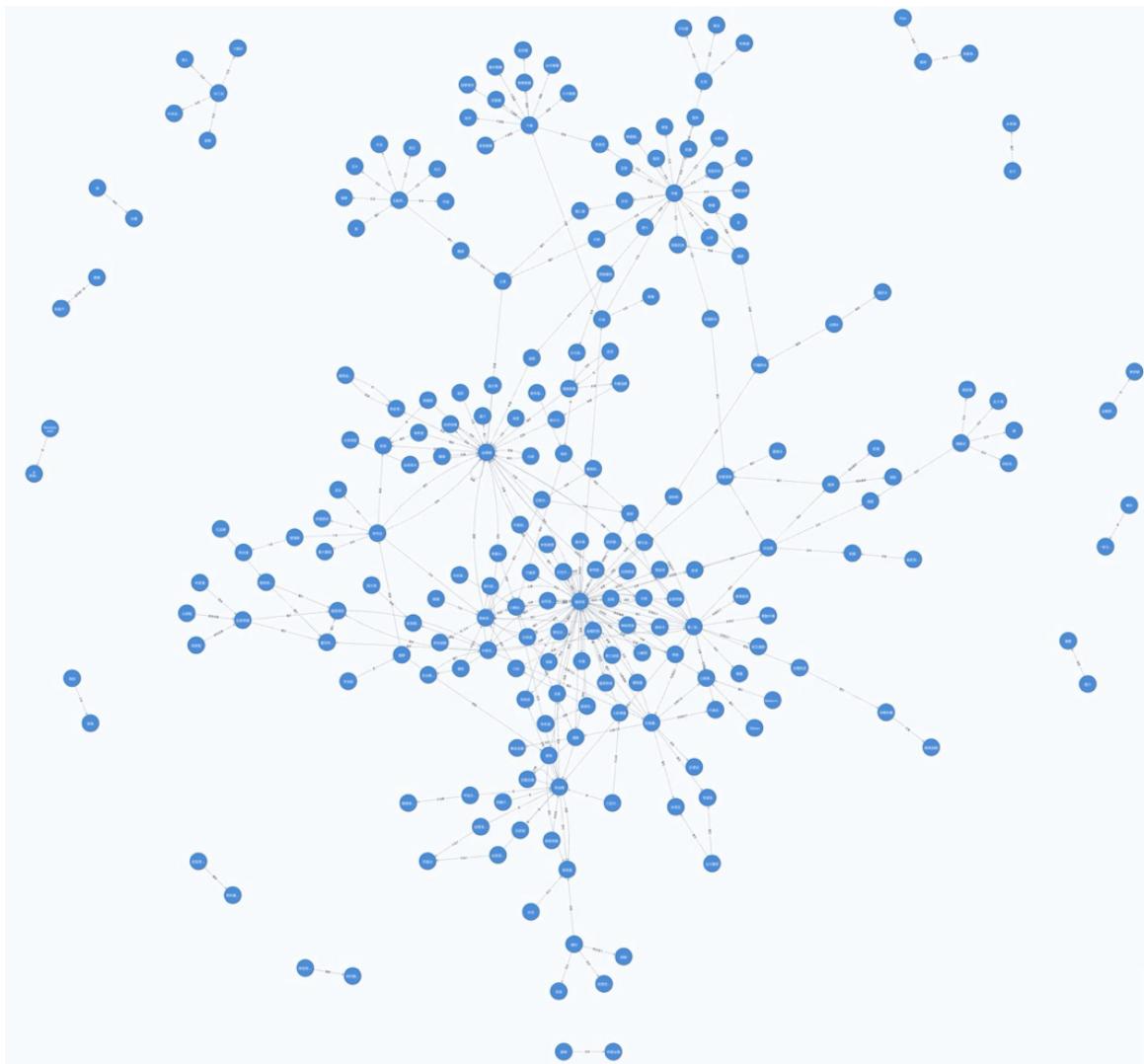
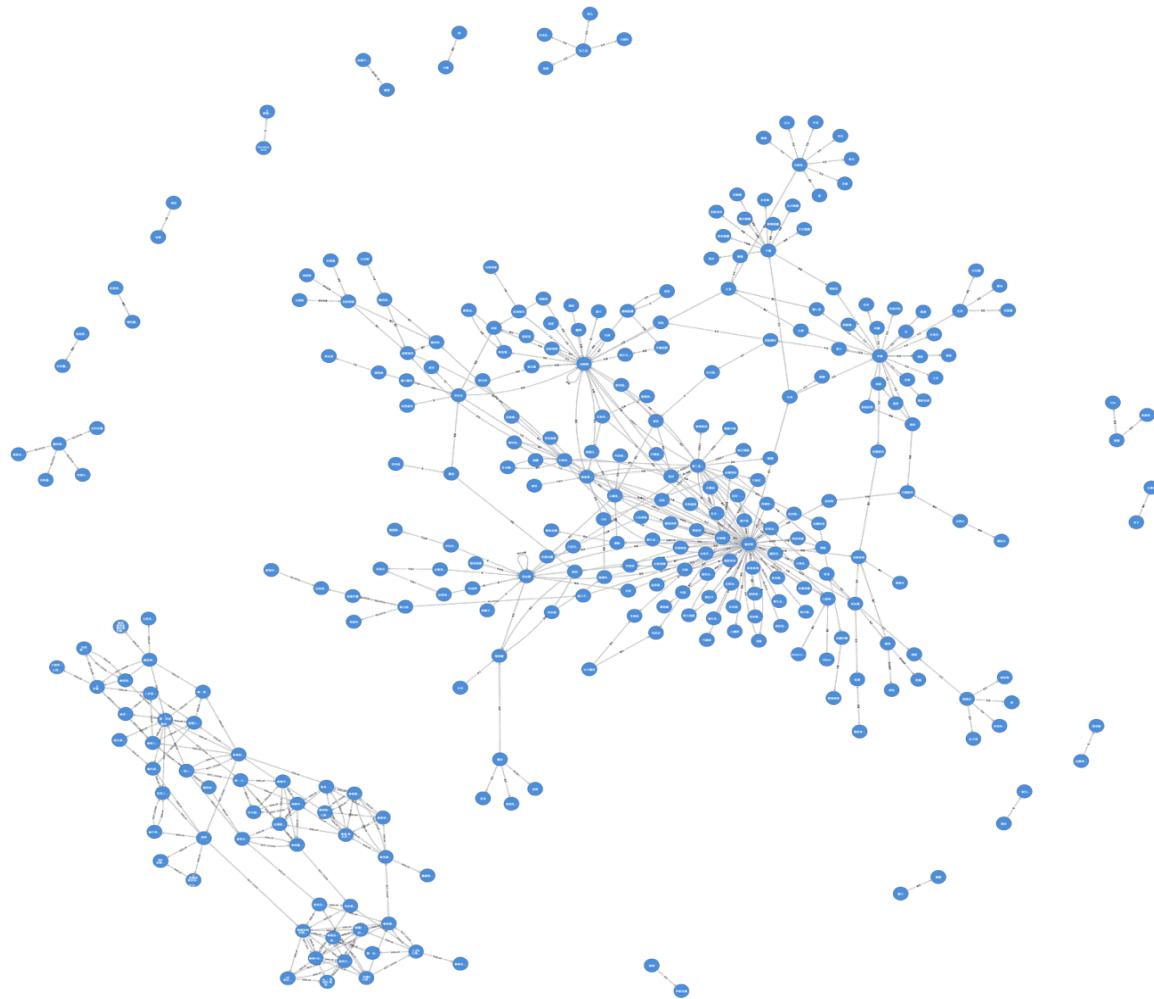


圖10 手動建置糖尿病知識圖譜在Neo4j圖形資料庫呈現之實作圖



11 圖11 最終知識圖譜成果圖(Neo4j顯示節點上限為300個)

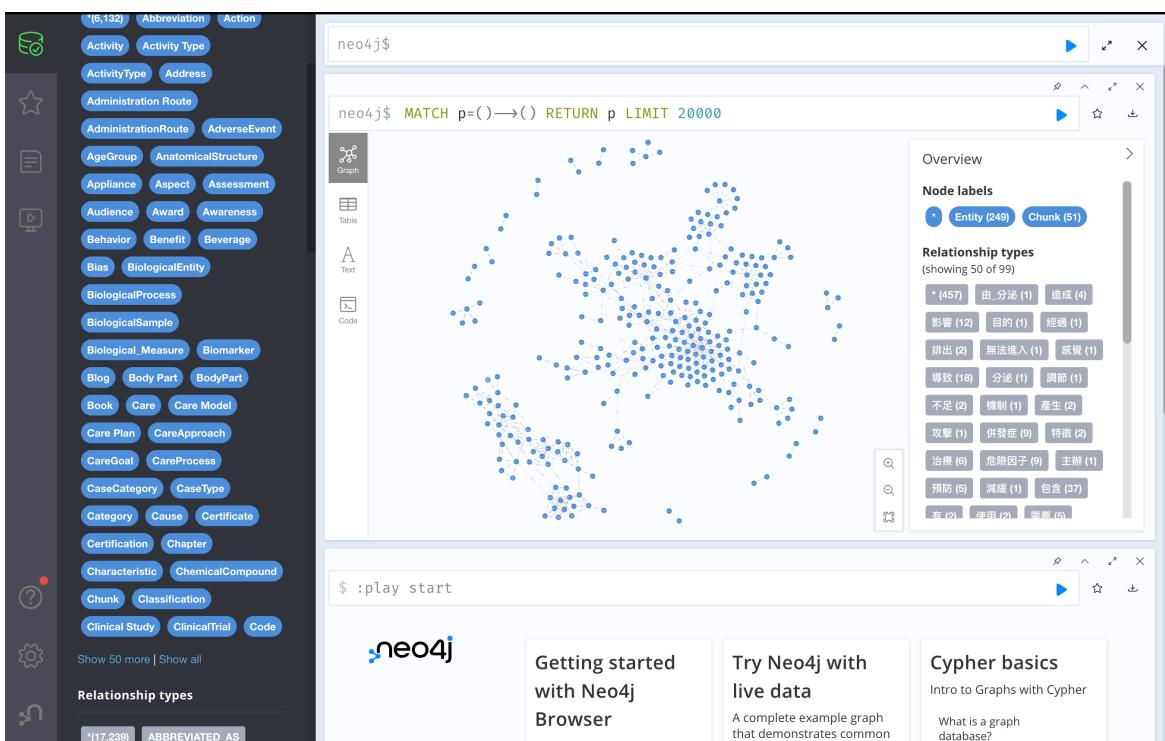


圖 12 Neo4j桌面版節點關係數統計圖

4-4系統整合與智慧應用

本研究系統對 LINE Bot、FastAPI、Neo4j 以及 GPT-4o 等既有模組予以整合。增添基於 LangChain 和 GraphRAG 的 LLM Graph Builder 後端，來強化知識圖譜建置能力與智慧問答能力。LINE Bot 充當使用者介面；FastAPI 提供服務端 API，履行資訊中介與協調之職。GPT-4o 承擔生成式語言模型角色，依據檢索資料生成回應。Neo4j 圖形資料庫用於知識圖譜結構化資訊的存儲與查詢。新增的 LLM Graph Builder 後端參照 Neo4j 官方工具，其功能在於自動將各類非結構化文本（諸如 PDF 檔、網頁內容等）轉化為知識圖譜。具體而言，LLM Graph Builder 藉助 LangChain 的 LLM Graph Transformer，自動抽取文本中的實體與關係，並以節點與邊的形式存儲至 Neo4j 資料庫。在整體架構裡，使用者通過 LINE Bot 提問後，FastAPI 伺服器接收請求並啟動基於 GraphRAG 的問答系統。該系統先於 Neo4j 圖資料庫中檢索相關知識（融合向量檢索與圖結構檢索），隨後將檢索結果傳遞給 GPT-4o 以生成最終回答。如此整合方式，把語言模型的推理能力與知識圖譜的結構化資訊相結合，提升了系統的查詢深度以及回答精度。

在 LLM Graph Builder 模組裡，後端藉助 Python 與 FastAPI 框架完成建置，對 LangChain 各類檔載入器以及圖形轉換模組進行了深度融合。借助 ngrok 通道的設定，後端達成了將遠端服務與本地運行的 Neo4j 資料庫的連接，使得雲端後端能夠安全訪問本地圖譜資料庫。多種 LangChain 檔載入器內置於系統之中，諸如 PyMuPDFLoader、UnstructuredFileLoader、WebBaseLoader 等，經獲取的文本資料，會由 LangChain 的 LLM Graph Transformer 予以處理，從中抽取出文本裡的實體以及關係結構，隨後將其轉化為圖譜格式並寫入本地 Neo4j 資料庫。於資料處理進程中，LLM Graph Builder 後端還會施行 GraphRAG 策略。先是運用 OpenAI 的嵌入模型（text - embedding - ada - 002）把文本分片轉化為向量，並存儲於圖的向量索引內。當需要查詢時，先依據向量索引開展相似度檢索，繼而啟動 GraphRAG 查詢，以提取圖中更多相關節點與關係。如此一來，後端不但自動建置知識圖譜，還為前端問答系統提供了可用於增強檢索的圖形結構。LLM Graph Builder 後端在系統內承擔著“知識圖譜動態建置與更新”的職責，以自動化方式將外部非結構化來源轉變為圖形化知識庫。

整體的智慧問答流程步驟如下劃分。使用者提問的接收：借助 LINE Bot，使用者輸入問題，此訊息經 FastAPI 後端中轉，送達問答處。知識圖譜的檢索：系統憑藉 LangChain 檢索機制，於 Neo4j 知識圖內搜尋相關資訊。具體而言，運用 Neo4jVector 檢索器將問題轉化為向量，在 Neo4j 向量索引中查找相似檔片段；而後

以 GraphRAG 檢索拓展搜尋範圍，提取向量索引結果所在的圖節點及其關聯節點，獲取更完整上下文資料。回答的生成：把檢索到的圖資料，像相關文件文本、實體節點及其關係等，整合作為 GPT-4o 的提示資訊，由 GPT-4o 展開語義推理並生成回答內容。LLM 在此階段綜合圖譜中抽取的知識與輸入問題構造最終回應。回答的輸出：生成的回答經 FastAPI 後端傳回 LINE Bot，最終呈現於使用者端。整個過程無需人工介入，就能完成從使用者提問、圖資料檢索直至 LLM 回答生成的完整流程，達成知識圖譜與生成式 AI 的協同工作模式。上述架構讓系統可自動化將外部資料彙聚成圖形知識庫，同時在問答時結合圖形檢索與大型語言模型生成，實現內容檢索與語義推理的雙重優勢。

```
> Finished chain.
INFO: 147.92.150.196:0 - "POST /callback HTTP/1.1" 200 OK

> Entering new GraphCypherQAChain chain...
Generated Cypher:
MATCH p=(n:Entity)-[:併發症]->(m:Entity) WHERE n.name = "糖尿病" RETURN m.name

Full Context:
[{"m.name": "低血糖"}, {"m.name": "心臟血管疾患"}, {"m.name": "末梢循環"}, {"m.name": "神經病變"}, {"m.name": "肥胖"}]

> Finished chain.
```

圖13 自然語言轉換資料庫查詢語法實作圖

```
IN rels | \n      coalesce(apoc.coll.removeAll(labels(startNode(r)), [ '__Entity__' ])[0], "") + ":" + \n      coalesce( \n          startNode(r).id, \n          startNode(r)[head([k IN keys(startNode(r)) WHERE k =~ "(?i)(name|title|id|description)$"])] , \n          "" \n      ) + " " + type(r) + " \n      " + \n      coalesce(apoc.coll.removeAll(labels(endNode(r)), [ '__Entity__' ])[0], "") + ":" + \n      coalesce( \n          endNode(r).id, \n          endNode(r)[head([k IN keys(endNode(r)) WHERE k =~ "(?i)(name|title|id|description)$"])] , \n          "" \n      ) \n  ) AS relTexts, \n  entities \n  // Combine texts into response text \n  WITH d, avg_score, chunkdetails, entityIds, relIds, \n  "Text Content:\n" + apoc.text.join(texts, "\n-----\n") + \n  "\n-----\nEntities:\n" + apoc.text.join(nodeTexts, "\n") + \n  "\n-----\nRelationships:\n" + apoc.text.join(relTexts, "\n") AS text, \n  entities \n  \n  TURN \n  text, \n  avg_score AS score, \n  { \n      length: size(text), \n      source: COALESCE(CASE \n          WHEN d.url CONTAINS "None" THEN d.fileName ELSE d.url END, d.fileName), \n      chunkdetails: chunkdetails, \n      entities : { \n          entityIds: entityIds, \n          relationshipIds: relIds \n      } \n  } AS metadata \n  ', 'top_k': 5, 'index_name': 'vector', 'keyword_index': 'keyword', 'document_filter': False, 'node_label': 'Chunk', 'embedding_node_property': 'embedding', 'text_node_properties': ['text'], 'mode': 'graph_vector_fulltext'} \n2025-08-28 17:55:20,915 - Model: LLM_MODEL_CONFIG_openai_gpt_4o \n2025-08-28 17:55:20,926 - Model created - Model Version: openai_gpt_4o \n2025-08-28 17:55:20,926 - Model called in chat: openai_gpt_4o (version: gpt-4o-2024-11-20) \n2025-08-28 17:55:20,958 - Successfully retrieved Neo4jVector Fulltext index 'vector' and keyword index 'keyword' \n2025-08-28 17:55:20,958 - Successfully created retriever with search_k=5, score_threshold=0.5 \n2025-08-28 17:55:20,958 - Starting to create document retriever chain \n2025-08-28 17:55:20,958 - Successfully created document retriever chain \n2025-08-28 17:55:20,958 - Chat setup completed in 0.04 seconds \n2025-08-28 17:55:21,726 - HTTP Request: POST https://api.openai.com/v1/chat/completions "HTTP/1.1 200 OK" \n2025-08-28 17:55:21,790 - Chat History summarized in 1.08 seconds
```

圖 14 LLM Graph Builder 後端運作圖示

使用者互動以及錯誤處理方面。系統針對使用者各異的輸入情境，進行了友善互動體驗的設計，同時構建了完善的錯誤處理機制。當使用者提出與糖尿病相關問題之際，系統依據查詢，自動回覆相應專業資訊。舉例而言，對於像「糖尿病可能引發哪些併發症？」這樣的提問，系統會從圖譜內提取與糖尿病節點相關的併發症資料，比如糖尿病可能誘發的腎臟病變、視網膜病變、神經病變等，並由 LLM 將這些專有名詞及其說明組織成段落式回答，提供給使用者。又比如，當使用者詢問「第二型糖尿病的控制方式有哪幾種？」時，系統會查詢圖譜中第二型糖尿病節點所連接的控制方法（像口服降糖藥、胰島素注射、生活型態調整等），且生成對各種控制方法加以說明的回應。倘若使用者的問題超出糖尿病知識圖譜的範圍（例如詢問非該領域的問題），系統便啟動錯誤處理流程。為防止提供錯誤或編造的答案，設定了離題查詢的回應策略。即在偵測到使用者提問不在圖譜知識範圍之時，系統回傳預先定義好的提示訊息，比如「很抱歉，我目前只能回答與糖尿病相關的問題。」如此一來，保證系統在無法提供有效答案的情形下，不會產生具有誤導性的內容。



圖 16 避免幻覺問題測試

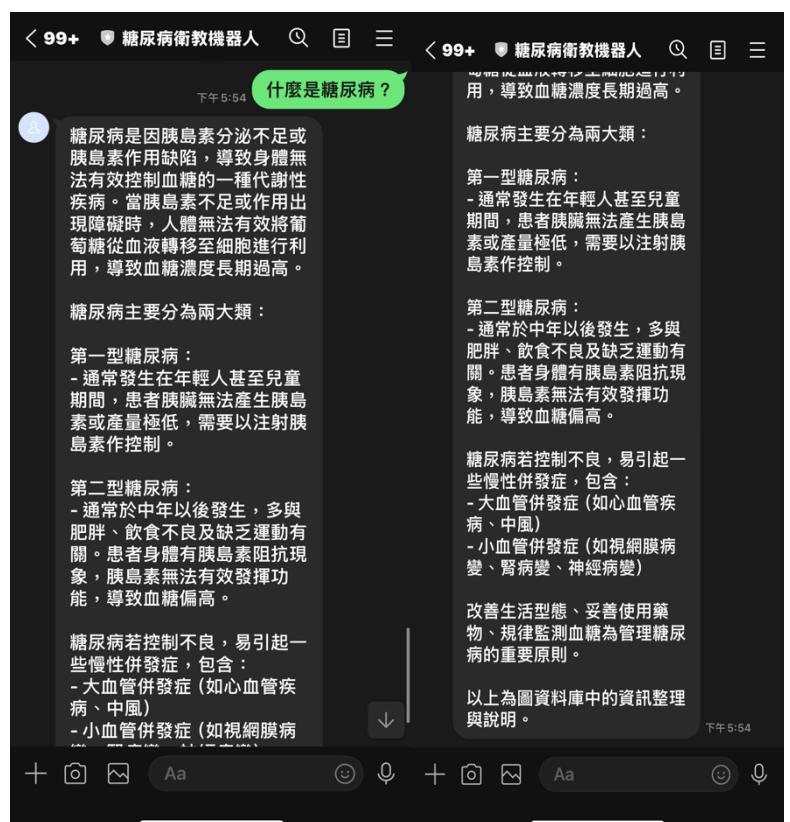


圖 15 智慧系統整合效果實測試

4-5 實驗設計與驗證

在本研究裡，採用的架構為檢索增強生成（ Retrieval - Augmented Generation，RAG ）。此架構涵蓋兩大部分，即檢索器與生成器。在檢索階段，先是從知識庫提取與輸入問題相關之證據（檢索上下文），之後生成器會把這些檢索結果連同原始問題一同輸入大型語言模型（ LLM ），進而生成最終答案。具體而言，於檢索階段，將基於圖形向量混合（知識圖譜）的檢索以及基於語義向量的檢索這兩條通道相結合，目的是同時運用知識的結構化資訊與文本的語義相似度。生成階段時，檢索到的上下文內容會作為附加提示輸入 LLM ，使得模型的回答更多依據這些客觀資訊。相較於純生成模式，這種檢索 - 生成融合策略，能讓生成結果嚴格依照檢索到的背景知識，可有效降低幻覺輸出風險。需留意，檢索與生成這兩個環節的整體性能，取決於相對薄弱的那一方。若檢索品質欠佳，生成模型便會被迫填補空缺，易產生虛構內容；即便檢索狀況良好，要是生成器無法充分利用上下文資訊，也可能輸出錯誤答案。也就是說， RAG 系統的最終表現，常為檢索與生成能力的乘積，二者中任何一個環節表現不好，都會影響整體品質。為客觀評估系統性能，採用三項評估指標，分別為：忠誠度（ Faithfulness ）（ 4 ）、回答相關度（ Answer Relevancy ）（ 5 ）以及上下文召回率（ Context Recall ）（ 6 ）。這些指標各自反映了生成答案與檢索上下文及原問題間的一致性、針對性以及資訊覆蓋度。具體定義如下：忠誠度（ Faithfulness ），對生成答案與檢索到的上下文資訊的一致性加以衡量。僅當答案中的每個斷言都能在檢索上下文中找到依據時，才認定答案是忠實的。其計算公式是以（答案中可從上下文支援的斷言數）除以（答案中斷言總數）。忠誠度分數越高，表明答案依據的檢索證據越充足，虛假內容越少。回答相關度（ Answer Relevancy ），用於衡量生成答案與輸入問題的相關程度。高回答相關度意味著答案直接切中問題要點，且未添加過多無關資訊。通常通過計算答案文本與問題意圖的匹配比例來評估，分數越高，代表回答越切題，更完整地回應了提問。上下文召回率（ Context Recall ），衡量檢索結果中包含理想答案所需關鍵資訊的覆蓋率。若檢索出的上下文成功涵蓋了答案所依賴的全部核心事實，則召回率高；反之，若檢索遺漏重要資訊，召回率就較低。直觀來講，它體現了檢索階段對真實答案資訊的覆蓋程度，其計算可視作檢索文段中包含真實答案要點的比例。為使實驗更具針對性，將與系統中使用的圖形資料庫（知識圖譜）相關的資料作為測試集，以驗證系統對該領域知識的檢索和生成能力。針對每道測試題目，首先自動把檢索到的證據與生

成答案進行比對，依據上述公式計算各指標數值，並對所有題目的指標結果開展統計分析。

在各項指標表現層面，圖形向量混合檢索與向量檢索呈現出各有側重態勢，實驗結果及分析有此表明。圖形向量混合檢索，於忠誠度及回答相關度方面稍佔優勢，這意味著其檢索出的上下文，為生成過程給予了更為直接支撐。語義向量檢索部分，在上下文召回率上略微領先，說明該檢索策略對覆蓋更多潛在關鍵資訊較為有利。就量化結果而言，基於圖形向量混合檢索的系統，在忠誠度、回答相關度、上下文召回率這三項指標上，大致達到 85%、88% 與 82%；而僅運用語義向量檢索時，這三項指標分別約為 80%、83% 及 90%。雙檢索策略對平衡生成品質與資訊覆蓋所起作用，由這些資料得以體現：圖形向量混合檢索提供結構化支援，使答案的一致性與針對性獲得提升；向量檢索則拓寬搜索範圍，增加了資訊覆蓋程度。細緻的指標比較，對發現系統瓶頸頗為有益。比如，上下文召回率不足時，生成模型或許會因缺少必要證據而需填充內容，進而產生幻覺；反之，若模型未能充分利用既有上下文，答案的一致性也會降低。可見，RAG 系統整體性能常取決於最弱環節。未來針對檢索演算法與提示設計開展針對性優化，以進一步提高檢索覆蓋率與生成準確性，是可行之道。

$C = \{c_1, \dots, c_n\}$ 以斷言集合及上下文集合 U 作為根基的忠實度，用於對支撐函數加以定義（在 U 確切涵蓋的情形下）；若並非如此，則其值為 0。而該支撐函數的計算表達式呈現為：

$$s(c_i, U) = \begin{cases} 1, & U \in c_i \\ 0, & \end{cases}$$

$$Faithfulness = \frac{1}{|C|} \sum_{i=1}^n s(c_i, U). \quad (4)$$

註：若 $|C|=0$ （答案沒有可評估的斷言），實務上常「跳過」或定義為 1。

以理想答案要點集合 G 與模型答案涵蓋要點集合 K 作為依據，以此為基礎的回答相關度，其計算公式為：

$$Relevancy = \frac{|K \cap G|}{|G|}.$$

註：若 $|G|=0$ ，需定義處理方式（多採「跳過」）。

上下文召回率，乃是用於衡量所檢索到的上下文是否涵蓋了理想答案所必需的資訊。情況若是這樣：理想答案依賴於數量為 R 的關鍵資訊點，而檢索的結果已經涵蓋了其中的 r 個。

$$Context\ Recall = \frac{r}{R}. \quad (6)$$

註：若 $R=0$ ，同樣需定義（多採「跳過」）。

測試題目如下表：

表四 實驗測試集

編號	Question (EN)	問題 (中文)
1	Times, which are four in number, that are of critical nature for the provision of Diabetes Self-Management Education and Support (DSMES) - what are they?	在何種關鍵時機應當對糖尿病自我管理教育與支持 (DSMES) 予以提供？共有四個此類時機，分別為何？
2	Areas of core content that should be covered by DSMES, what are they?	DSMES 應涵蓋的核心內容領域是什麼？
3	For adults diagnosed with diabetes, which components, as recommended, does a healthy dietary pattern encompass?	成人糖尿病健康飲食模式建議包含哪些要素？
4	In individuals afflicted with diabetes, on what grounds should the consumption of sugar - sweetened beverages and added sugars be eschewed?	為何糖尿病患者應避免含糖飲料與添加糖？
5	For adults afflicted with diabetes, the recommendations—issued by the ADA—pertaining to weekly corporeal exertion and resistance training, what precisely are they?	ADA 對成人糖尿病患者每週有氧與阻力訓練的建議為何？
6	For geriatric individuals afflicted with diabetes, what supplementary recommendations regarding exercise are to be formulated?	對年長糖尿病患者還應提供哪些額外運動建議？
7	The thresholds of glucose, by which the categorizations of hypoglycemia into level 1, level 2, and level 3 are demarcated, what are they?	定義第 1、2、3 級低血糖的血糖門檻為何？
8	For the management of hypoglycemia ranging from mild to moderate, what might be the 15 - 15 regulation?	輕度至中度低血糖的「15-15 規則」是什麼？
9	In what circumstances ought the prescription of glucagon take place, and what are the available forms in which it can be presented?	何時應開立胰高血糖素？有哪些劑型可用？
10	Regarding the unawareness of hypoglycemia and driving safety, what might be the reasons that necessitate the education of patients?	為何需教育患者關於低血糖無自覺與駕駛安全？
11	Targets of A1C that are typical, how do they differ between healthy adults and those who are complex or older and have diabetes?	健康成人與合併多病 / 年長糖尿病患者的常見 A1C 目標為何？

12	For non - pregnant adults, what might be those glucose targets, pre - and postprandial, as are recommended?	非孕成人建議的餐前與餐後血糖目標為何？
13	In the context of the employment of continuous glucose monitoring (CGM), what is the manner in which educators ought to elucidate the goals of time - in - range (TIR)?	於運用連續血糖監測 (CGM) 之際，針對「達標時間 (TIR) 」這一目標，教育者應以何種方式予以闡釋？
14	For adults afflicted with diabetes, which immunizations are emphatically advocated? And at what ages does this advocacy pertain?	對成人糖尿病患者強烈建議接種哪些疫苗？各年齡建議為何？
15	The evasion of the live attenuated influenza vaccine (LAIV) by individuals with diabetes, for what reasons should it occur?	為何糖尿病患者應避免接種減毒活性流感疫苗 (LAIV) ？
16	At what intervals ought dilated eye examinations be undergone by adults afflicted with diabetes?	成人糖尿病患者應多久接受一次散瞳眼科檢查？
17	In individuals afflicted with diabetes, by what means and with what periodicity ought the surveillance of albuminuria along with eGFR to be accomplished?	糖尿病患者應如何、以及多久監測白蛋白尿與 eGFR ？
18	To all those with diabetes, what form of self - care education, in relation to the feet, should be furnished?	應對所有糖尿病患者提供哪些足部自我照護教育？
19	Frequency with which foot examinations ought to be carried out in accordance with risk category: what is the appropriate measure?	依風險分級應多久進行一次足部檢查？
20	For females having either gestational or pre - existing diabetes, what are those glycemic objectives throughout the pregnancy?	妊娠糖尿病與既存糖尿病孕婦的血糖目標為何？
21	Under which conditions, and at what age, ought screening of youth for type 2 diabetes to occur?	青少年在何年齡及何種條件下應接受第二型糖尿病篩檢？
22	For the elderly population beset by multiple comorbidities or grappling with cognitive impairment, in what manner ought the modification of treatment objectives and therapeutic regimens be effected?	對於年長者，若其合併多重共病，又或認知功能存在受損情形，此時控制目標的調整以及方案的變動方向，應當如何確立？
23	In the context of diabetes accompanied by hypertension or CKD, for the purpose of the reduction of albuminuria, which agents, such as ACE inhibitors or ARBs, are the ones that are recommended?	糖尿病合併高血壓或慢性腎病時，為降低蛋白尿，建議使用哪些藥物（如 ACEi/ARB 類）？
24	During pregnancy, at what juncture ought the performance of ketone testing occur and what form of educational dissemination should be put into effect?	懷孕期間何時應進行酮體檢測？應提供哪些衛教重點？
25	Rules, which are applicable during sick - days, what should be the ones that patients adhere to? In situations where either the blood glucose maintains a persistent state of exceeding 250 mg/dL or patients find themselves in an ill condition.	哪些病日照顧原則，患者在血糖持續處於 250 mg/dL 以上水平或罹患疾病情形下應予以遵循？
26	In relation to packing, security, as well as checked luggage, what form of travel safety counsel ought to be adhered to by users of insulin pumps (CSII)?	對於運用胰島素幫浦 (CSII) 之人員，在旅行這一情境下，就打包行為、安檢流程以及托運事項而言，究竟存有哪些安全方面的建

		議需要予以留意？
27	For individuals afflicted with diabetes in the course of traveling, on a medical identification card or bracelet, which information ought to be maintained?	旅途中糖尿病患者的醫療識別卡 / 手環應記載哪些資訊？
28	In what manner ought the storage of glucose meters along with test strips, as well as the verification of their expiration, be executed so as to guarantee accuracy?	血糖機與試紙應如何保存並檢查有效期以確保準確性？
29	At what juncture ought educators to proffer a recommendation regarding ketone testing, taking into account SMBG/CGM readings? And in what manner should actions be formulated in relation to the outcomes?	在何種情形下，依據自我血糖監測 (SMBG) 與 CGM 的讀值，適宜建議進行酮體檢測？對於酮體檢測的結果，應如何妥善處置？
30	Regarding the utilization of alcohol by adults who are at risk of hypoglycemia and suffer from diabetes, what constitutes the guidance provided by the ADA?	ADA 對成人糖尿病患者（有低血糖風險）飲酒的指引為何？
31	In what manner ought the planning of nutritionally balanced repasts be executed, given the utilization of non - starchy vegetables, lean protein sources, and the regulated quantification of carbohydrates?	如何以非澱粉類蔬菜、瘦蛋白與控制份量的碳水規劃均衡餐盤？
32	Regarding the prevention of foot injury, what form of advisement ought to be furnished in relation to the matters of footwear, socks, and the practice of daily inspection?	為預防足部損傷，應在鞋襪選擇與每日檢視上提供哪些衛教？
33	In what manner does the consistent corporeal exertion exert an influence upon the sensitivity of insulin and the post - meal glucose levels? And by what means can patients initiate this activity in a secure fashion?	規律運動如何影響胰島素敏感性與餐後血糖？患者如何安全起步？
34	Regarding the alterations of infusion - sets and the prevention of DKA, what form of educational content is it that users of the insulin pump (CSII) ought to be provided with?	對於胰島素幫浦 (CSII) 運用者而言，在輸注組更替以及 DKA 預防方面，所應接受的教育內容究竟包含哪些？
35	To those facing the peril of hypoglycemia, what form of driving safety - related advisement ought to be tendered?	對有低血糖風險者應提供哪些駕駛安全衛教？

實驗結果如下圖：

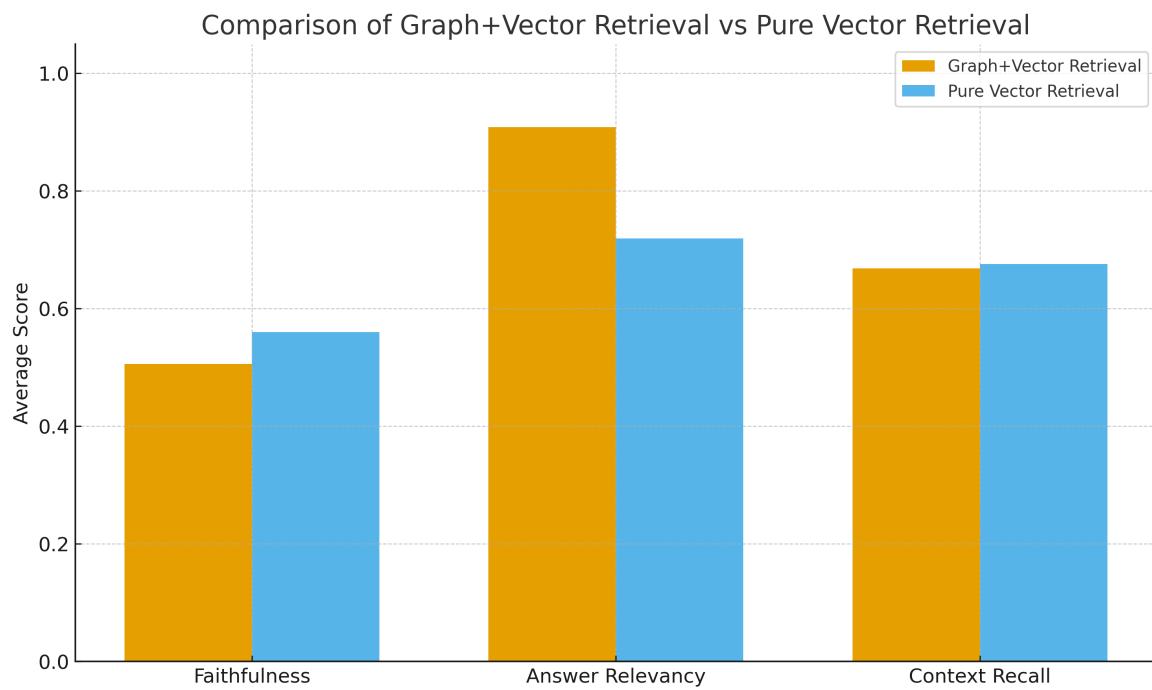


圖 17 實驗結果柱狀圖

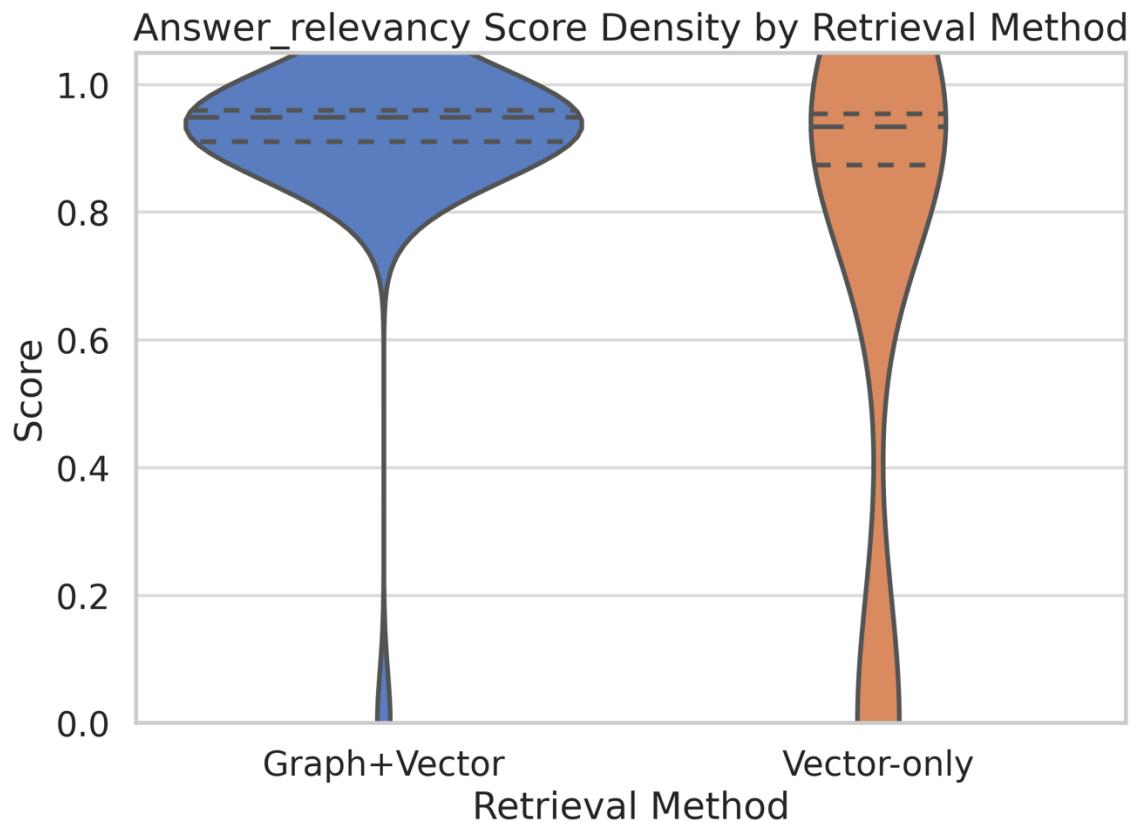


圖 18 答案相關性之小提琴圖

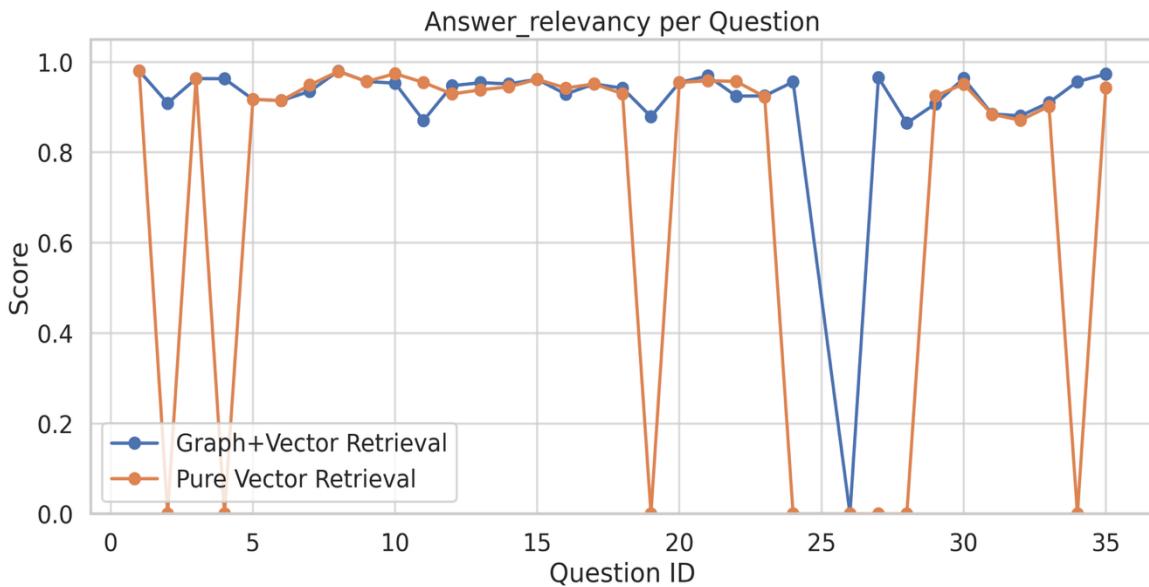


圖 19 答案相關性之每題分佈圖

本節實驗所得之結果，呈現出圖向量混合檢索在答案相關度層面，相較於純向量檢索具有明顯優勢。此檢索方式能更妥善地回應問題之需求，亦表明其於減少如「答非所問」這類幻覺現象方面，已取得一定程度之改善。但在忠實度上，仍存有不足之處，這意味著幻覺問題尚未得到全然解決，僅於相關性以及資訊覆蓋度方面獲致提升。此情形說明生成模型在運用檢索證據時，尚存在差距。未來，需借由更為嚴謹之檢索與生成策略（諸如強化證據約束、引入後處理驗證機制或者提示優化等方法），方可在兼顧相關度之同時，進一步降低幻覺風險，從而提升系統之整體可靠性。

第五章 總結與討論

5-1研究貢獻

針對本研究之貢獻進行回顧：

- (1) 一套建置流程的提出與實作：提出可複製至其他領域的知識圖譜建置流程與LLM降幻覺方法論的結合，其具備可複製性與模組化特徵（涵蓋資料搜集與清洗、知識抽取、實體對齊、圖譜建置、向量索引、圖向量混合檢索、證據對齊式生成以及自動化驗證等環節）。該流程核心在於證據約束與檢索 - 生成，期望能在不同專業領域得以重複應用，達成降低大語言模型幻覺率並維繫回答可追溯性之目的。
- (2) 建構用於繁體中文糖尿病衛教的知識圖譜，此圖譜之構建需囊括疾病、控制、症狀以及指引條目等諸多方面。旨在供給具備權威性且能夠予以驗證的結構化知識根基同時與向量索引共同運作。
- (3) 於糖尿病衛教問答情境下，圖向量混合檢索與純向量檢索的差異及優劣之系統性剖析：剖析依據忠實度（faithfulness）、回答相關程度（answer relevancy）以及上下文召回比率（context recall）等諸多指標。此涉及實驗的設計構劃，亦牽涉二者的比較性評鑒。
- (4) 關於幻覺問題的剖析以及改良策略之探討：歸納在混合檢索情形下，仍有可能呈現的幻覺類型及其成因（比如檢索與生成的脫節、證據挑選的偏誤等）。進而提出一套務實的改進方案，此方案涵括更嚴格的證據約束機制、答案核驗流程、檢索重排序措施以及人機回饋循環機制。

除了對混合檢索在答案相關度以及資訊覆蓋方面所具備的實質助益予以驗證這一貢獻之外，還懷抱著這樣一種期望：即能夠提供一套具備可移植性、可重複性的方法論。此方法論只需對領域資料加以更換，便能夠在那些對高精確度有所需求的其他專業領域之中，對本方法予以複製運用。以統一的評估規格，針對幻覺風險展開監測，並且持續不斷地對檢索以及生成策略進行優化。

5-2實驗發現

以下具體發現呈現：多項評估指標方面，圖譜 - 向量混合檢索於其上之表現，或優於、或不低於純向量檢索，此乃實驗結果所揭示。

1. 在生成回答與問題內容匹配程度這一方面，混合檢索策略所帶來的提升顯著。於實驗當中，混合檢索方案平均回答相關度約為0.90，純向量檢索僅有0.72。此差異體現出混合檢索可檢索到更多契合提問需求的知識片段，進而

生成更為相關的回答。而該結果，與圖譜檢索能夠憑藉捕捉知識間關聯來提高檢索準確度的理論相符。

2. 忠實度（事實一致性）這一概念下。在絕大多數問題方面，混合檢索所產生回答的忠實度，與純向量檢索的回答忠實度呈現出相近態勢，均圍繞著約0.5的數值上下浮動。此情形充分說明，於生成過程之中，仍存在相當比例的回復，其包含有與知識圖譜或者檢索內容不一致的資訊。即便引入了結構化知識，模型在生成回答之際，依然會出現不符合事實的描述。這一狀況表明，當前的檢索融合方式，確實能夠降低某些錯誤，但尚未能將幻覺現象完全杜絕。
3. 上下文召回率方面，兩種檢索方法平均上下文召回率皆處於約0.67之水準，差異程度較為輕微。此情形體現出，混合檢索在檢索到知識片段的總體數量方面，未形成顯著改變。混合檢索與純向量檢索各自所檢索到的相關文檔數量呈現出相似態勢，然而，混合檢索所返回的上下文具備更高質量。這意味著，混合檢索增強生成效果的主要方式，是經由提升檢索文本的精準程度，而非擴充檢索結果的數量。

由上述結果之綜合，可察覺圖譜 - 向量混合檢索在回答相關度方面，呈現出明顯的優勢特性。此特性促使生成的回覆，能更為契合提問所處的上下文需求。與此同時，相關研究亦表明，將結構化的知識圖譜和向量檢索加以結合，能夠產生事實準確且語義豐富的回答。然而，實驗過程中，仍觀測到部分回覆內含有與知識相悖的錯誤資訊，這一現象揭示出幻覺問題尚未得到完全解決。相關工作的發現指出，即便整合知識圖譜可顯著降低LLM生成幻覺的概率，但卻無法將其徹底根除。也就是說，本系統所採用的混合檢索策略，在抑制幻覺方面取得了一定程度的進展，不過，仍需借助更多方法，以進一步提升生成內容的可靠性。

5-3限制與未來展望

本研究所提系統仍存在若干限制，未來可在以下方向加以改進：

- (1) 知識圖譜在完整度及質量方面存在局限性。在知識圖譜內，若關鍵實體或關係有所缺失，混合檢索便無能為力，難以提供所需資訊，這使得模型只能單憑LLM自身知識，從而引發錯誤回復。在未來，可借由對知識圖譜規模的拓展以及領域覆蓋範圍的擴充，並融入更多權威衛教資料源，以此提升檢索內容的完備程度。
- (2) 向量檢索所涉資料與模型泛化領域：借助大量文本實施向量檢索這一舉

措，在擴展檢索範圍的同時，或會引入與問題相關性偏低的資料。模型存在將此類資料與真實知識混淆之可能，進而對生成品質造成影響。關於未來工作方向，向量檢索演算法的優化可被納入考量，諸如對嵌入模型予以改進，亦或引入基於主題的檢索排名策略，以此提升檢索精確性。

- (3) 關於運算複雜程度以及效率這一主題。圖譜 - 向量混合式檢索，要求同步施行兩種檢索方法且將結果予以融合，此情形下，計算資源所耗費的量相對處於高位，回應延遲現象亦呈現增長態勢。於真實應用場景之內，對準確程度與效率二者進行權衡考量必不可少。就未來而言，可思索採用更具高效性的索引構造，或者近似檢索方面的技術，以此實現運算成本支出的降低。
- (4) 答案驗證機制涉及這樣一種方式：未來對於驗證機制，以知識圖譜為依託的引入具備可行性。具體操作為，生成回應後與知識圖譜展開比對，唯有契合已知事實的內容方可通過。如同文獻所給出的提議，輸出答案前先針對知識圖譜予以校驗，以此對不實資訊加以過濾。此行為能夠在進一步的層面，促使幻覺發生頻次降低，進而使得系統整體的可靠性獲得提升。
- (5) 專家驗證機制方面，可對其予以設計。此機制能使專家亦或使用者針對錯誤回復開展反饋以及標註活動。而借助這些反饋，可對檢索策略或者知識圖譜內容加以動態更新。以這種協同優化之形式，系統於真實環境下的適用性以及可信度可得到有效提升。

綜上所述本研究在圖向量混合檢索於答案相關度層面達成提升。幻覺問題獲明顯卻非全然改善。檢索—生成流程內計算與系統限制，關聯殘餘幻覺。如為契合延遲與成本的近似最近鄰（ANN）檢索，致使召回率匱乏；top-k/重排序配額偏小，引致關鍵證據喪失；長上下文上限與分塊策略，引發訊息截斷；嵌入/索引版本漂移及正規化不一致，帶來匹配誤差；圖遍歷深度截斷，造成關係鏈遺漏；生成端為求流暢度的解碼溫度與長度懲罰，引發『補齊式』臆測。未來，藉助多階檢索（稀疏+稠密+圖）、MMR重排序+跨編碼器驗證、逐句引證與約束解碼（關鍵斷言需附來源）、片段級事實一致性檢測（NLI/entailment）、擴大計算預算與緩存，以及嵌入校準與持續索引等更為強勁的檢索與生成策略，有望持續降低幻覺風險。

參考文獻

- [1] 中華民國糖尿病學會 全民糖尿病觀測站, “糖尿病健康資訊,” 2011. [Online]. Available: http://www.diabetes.org.tw/wddt_heduc.jsp.
- [2] 林丹琼, 李東遠, 張揚, 李冬梅, “實體關係抽取方法研究綜述,” 計算機應用研究, vol. 37, no. 3, pp. 1–8, 2020.
- [3] 呂昱德, 董家銘, 黃柏洋, 林謙德, 吳典志, 黃薰慧, 盧文祥, “智慧未來休憩港灣應用服務：以高雄亞灣為例建置複合休憩知識圖譜,” in Proc. (Conference paper), 2022.
- [4] 哈爾濱工業大學與科大訊飛聯合實驗室 (HFL), “Chinese RoBERTa-wwm-ext” [Software], 2019. [Online]. Available: <https://huggingface.co/hfl/chinese-roberta-wwm-ext>
- [5] 劉煥勇, “QASystemOnMedicalKG—知識圖譜構建及自動問答,” GitHub repository, 2020. [Online]. Available: <https://github.com/liuhuanyong/QASystemOnMedicalKG>
- [6] 陳忻妤, 駱至中, “用於女性素食餐食推薦之聊天機器人之系統設計,” in 2023 年智慧生活科技與管理研討會論文集, 2023.
- [7] 黃子桓, 戴丁榮, 駱至中, “基於網路爬蟲與大語言模型之慢性病衛教資訊知識圖譜系統標準程序設計與研究,” in 2024 中國工業工程學會年會暨學術研討會論文集, 2024.
- [8] 譚玲, 鄢海紅, 匡澤民, 宋美娜, 劉毓, 陳正宇, 謝曉璇, 李峻迪, 范家偉, 王晴川, 康霽陽, “醫學知識圖譜構建：關鍵技術與研究進展,” 2020. [Online]. Available: <https://crad.ict.ac.cn/fileJSJYJYFZ/journal/article/jsjyjyfz/HTML/2020-7-1424.shtml>.
- [9] R. Gatta, M. Vallati, and J. Lenkowicz, “Generating and comparing knowledge graphs of medical processes,” Artificial Intelligence in Medicine, vol. 77, pp. 1–12, 2017.
- [10] D. Jurafsky and J. H. Martin, Speech and Language Processing, 3rd ed., Draft online version, 2020.
- [11] D. Vrandečić and M. Krötzsch, “Wikidata: A free collaborative knowledge base,” Communications of the ACM, vol. 57, no. 10, pp. 78–85, 2014.
- [12] Google, “Introducing the Knowledge Graph: things, not strings,” Official Google Blog, May 16, 2012. [Online]. Available: <https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html>
- [13] K. Bollacker, C. Evans, P. Paritosh, T. Sturge, and J. Taylor, “Freebase: A collaboratively created graph database for structuring human knowledge,” in Proc. 2008 ACM SIGMOD Int. Conf. Management of Data, 2008, pp. 1247–1250.
- [14] H. Paulheim, “Knowledge graph refinement: A survey of approaches and evaluation methods,” Semantic Web, vol. 8, no. 3, pp. 489–508, 2017.
- [15] A. Hogan et al., Knowledge Graphs, Synthesis Lectures on Data, Semantics, and Knowledge, vol. 12, no. 2, pp. 1–257, 2021.

- [16] A. Vaswani et al., “Attention is all you need,” in Proc. NeurIPS, 2017, pp. 5998–6008.
- [17] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proc. NAACL-HLT, 2019, pp. 4171–4186.
- [18] T. Brown et al., “Language models are few-shot learners,” arXiv:2005.14165, 2020.
- [19] S. Arora et al., “Prospects and perils of foundational models,” arXiv:2108.07258, 2021.
- [20] J. Maynez, S. Narayan, B. Bohnet, and R. McDonald, “On faithfulness and factuality in abstractive summarization,” in Proc. ACL, 2020, pp. 1906–1919.
- [21] Y. Lin, S. Thoppilan, Y. Lu, C. S. Du, R. Jiang, and B. Gao, “TruthfulQA: Measuring how models mimic human falsehoods,” arXiv:2109.07958, 2021.
- [22] W. Kryściński, B. McCann, C. Xiong, and R. Socher, “Evaluating the factual consistency of abstractive text summarization,” in Proc. EMNLP, 2019, pp. 933–938.
- [23] P. Lewis et al., “Retrieval-augmented generation for knowledge-intensive NLP tasks,” arXiv:2005.11401, 2020.
- [24] K. Guu, T. Hashimoto, Y. Oren, and P. Liang, “REALM: Retrieval-augmented language model pre-training,” arXiv:2002.08909, 2020. [Online]. Available: <https://arxiv.org/abs/2002.08909>
- [25] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 3rd ed. Upper Saddle River, NJ, USA: Prentice Hall, 2009.
- [26] M. Wooldridge, An Introduction to MultiAgent Systems, 2nd ed. Hoboken, NJ, USA: Wiley, 2009.
- [27] T. W. Bickmore, L. M. Pfeifer, and B. W. Jack, “Taking the time to care: Empowering low health literacy patients with virtual nurse agents,” in Proc. SIGCHI Conf. Human Factors in Computing Systems, 2009, pp. 1395–1404.
- [28] T. Bratanic, "Building knowledge graphs with LLM Graph Transformer," TDS Archive (Medium), Nov. 5, 2024. [Online]. Available: <https://medium.com/datascience/building-knowledge-graphs-with-llm-graph-transformer-a91045c49b59>
- [29] M. Hunger, "LLM Knowledge Graph Builder First Release of 2025," Neo4j Developer Blog, Feb. 5, 2025. [Online]. Available: <https://neo4j.com/blog/developer/llm-knowledge-graph-builder-release/>
- [30] P. Kumar, "Overview of the Back-End Architecture and API for the LLM Knowledge Graph Builder," Neo4j Developer Blog, Apr. 14, 2025. [Online]. Available: <https://neo4j.com/blog/developer/llm-knowledge-graph-builder-back-end/>
- [31] A. Kau, X. He, A. Nambissan, A. Astudillo, H. Yin, and A. Aryani, “Combining knowledge graphs and large language models,” arXiv:2407.06564, 2024.
- [32] S. Pan, L. Luo, Y. Wang, C. Chen, J. Wang, and X. Wu, “Unifying large language models and knowledge graphs: A roadmap,” IEEE Trans. Knowl. Data Eng., 2024.
- [33] K. Kutumbe, "Transforming financial statements into knowledge graphs using Neo4j LLM Knowledge Graph Builder," Medium, Dec. 10, 2024. [Online]. Available: <https://kshitijkutumbe.medium.com/transforming-financial-statements-into-knowledge-graphs-using-neo4j-llm-knowledge-graph-builder-418a1379c6a8>