

期末專題

簡介 (Introduction)

在期末專題中，你們將以 **3 至 4 人小組** 形式合作，將課堂上學到的方法應用到你們感興趣的真實問題上。

本次期末專題的主題範圍包含但不限於：

1. 搜尋中的個人化與推薦
2. Agentic IR (代理型資訊檢索)
3. 多模態檢索
4. 特定領域的檢索增強生成 (RAG)

以上所有主題近年來都相當熱門且受到高度關注，特別是隨著大型語言模型 (LLMs) 的引入。各主題的詳細說明列於下方「主題介紹 (Topics Introduction)」章節，可作為你們專題發想的參考與靈感。

評分 (Grading)

期末專題總成績由四個部分組成（共佔總成績的 **40%**）：

- 初始專題提案簡報：5%
- 期中成果簡報：10%
- 期末專題報告：15%
- 期末專題簡報：10%

要求 (Requirements)

期中專題成果簡報（**11/11 上課時進行**）

- 你需要在 **5 分鐘** 內報告你們調查過的相關研究、已經實作的方法，以及初步成果。
- 簡報內容需包含：

1. 介紹 (Introduction)

- 簡要描述你們要解決的問題。
- 為什麼這個問題重要？會處理哪些 IR 任務？

2. 文獻回顧 / 相關研究 (Literature Review / Related Work)

- 描述你們找到的、與任務相關的其他研究或論文。
- 僅僅提到一篇論文並不夠，你至少要簡單說明他們的方法/使用的手法，以及這些研究和你們工作的關聯性或差異。

3. 資料集 (Dataset)

- 描述你們使用的資料，包括：資料集大小、類別分布、以及你們做過的前處理。

4. 主要方法 (Main Approach)

- 提出一個模型、演算法或框架，來解決你們的任務。
- 詳細描述該模型與演算法，並使用具體範例示範其運作方式。
- 不要僅僅描述方法本身，而是要精確解釋這些方法如何應用在你們的問題上（例如：輸入/輸出是什麼、涉及哪些變數、因素、狀態等）。

5. 評估方法 (Evaluation Methods)

- 指定至少一個定義明確、數值化、自動化的評估指標，用於量化評估。
- 如果有任何質性評估的想法，也可以加以描述。

6. 初步成果 (Preliminary Results)

- 描述模型效能與初步結果分析，並檢視是否符合預期。
- 如果結果不如預期，請思考如何改善效能，並提出你們的下一步計畫。

期末專題報告（繳交期限：12/15 23:59）

- 你們必須繳交一份期末專題報告，用來展示專題成果。
- 報告可選擇性地包含一個系統 Demo（非必須）。
- 評分將依據以下各部分內容：

1. 介紹 (Introduction)

- 簡要描述問題。為什麼問題重要？會處理哪些 IR 任務？

2. 文獻回顧 / 相關研究 (Literature Review / Related Work)

- 描述你們找到的相關研究或論文，並簡述其方法及與你們工作的關聯性或差異。

3. 資料集 (Dataset)

- 描述所使用的資料：大小、類別分布、前處理等。

4. 基準模型 (Baseline)

- 描述與實作你們的基準方法。

5. 主要方法 (Main Approach)

- 提出一個模型、演算法或框架，並詳細解釋如何應用在你們的問題上（包含具體範例）。

6. 評估指標 (Evaluation Metric)

- 說明你們使用的評估指標（質性與量化），必要時附上公式。

7. 結果與分析 (Results & Analysis)

- 報告基準模型與主要方法的效能，以及你們進行的實驗。
- 如果找不到合適基準，可分析你們的結果是如何達成的。
- 總結你們的結果分析，並解釋這些結果對改進方法的啟發。

8. 錯誤分析 (Error Analysis)

- 描述一些實驗，說明系統的特性（優點與缺點）。
- 用數據、圖表或表格來支持分析。
- 說明可能的錯誤來源及原因。

9. 未來工作 (Future Work)

- 即使簡短，也要提出一些改進方向。
- 也可提到你們遇到的挑戰與可能的解決方法。

10. 程式碼 (Code)

- 提供 GitHub/Bitbucket 等程式碼連結。
- 內容需包含：任務概述、開發環境需求（如 requirements.txt）、使用方式、超參數設定、實驗結果等。

11. 小組貢獻 (Contribution of Each Member)

- 說明每位成員的貢獻比例。
- 若有成員貢獻明顯不足，成績會依實際情況調整。

12. 參考文獻 (References)

- 提供格式正確的參考文獻（格式不限）。

期末專題簡報（12/23 上課時進行）

- 你需要在 **5 分鐘** 內簡報你們的期末專題成果。
- 助教會嚴格控管時間，因此不可能涵蓋所有細節。

簡報需涵蓋以下內容：

1. **介紹 (Introduction)**：簡要說明問題與目標 IR 任務
2. **資料集 (Dataset)**：描述資料來源、大小、分布與前處理
3. **基準模型 (Baseline)**：簡要說明基準方法（不必太細）
4. **主要方法 (Main Approach)**：詳細解釋模型與演算法，並舉例示範如何應用
5. **評估指標 (Evaluation Metric)**：說明使用的質性與量化評估方法
6. **結果與分析 (Results & Analysis)**：呈現基準與主要方法的效能比較，以及實驗結果
7. **未來工作 (Future Work)**：簡述改進方向與挑戰

討論 (Discussion)

助教會在課程的 NewE3 平台上開設一個 **期末專題討論區**，你們可以在該論壇提出有關期末專題的問題。

助教將會盡快回覆。

期末專題報告繳交 (Final Project Report Submission)

你們必須繳交期末專題報告。相比之下，**初始提案簡報**、**期中簡報**與**期末簡報**僅依照簡報表現評分，不需要繳交任何檔案。

1. 期末專題報告繳交期限為 **12/15（星期一）23:59**。
2. 報告檔名需為：Project_Team{TEAM_ID}_report.pdf，內容應包含完整細節與程式碼連結。
3. **逾期不受理**，不接受任何遲交。
4. 只接受 **單一 PDF 檔案**，若格式或檔名錯誤，將扣 **10 分**。
5. 只需由 **一位組員** 在 NewE3 上繳交即可。
6. （選填）我們強烈鼓勵每組團隊將專題整理成論文，並投稿至頂尖會議（如 ACL 2026、SIGIR 2026、KDD 2026）。

◦ 這會成為你履歷上的強力加分項。

- 歡迎聯繫助教 **Yu-Chien Tang** 和 **Prof. Yen** 討論投稿事宜，我們非常樂意與你們合作。

7. 如果有任何問題，請隨時透過 email 聯繫：tommytyc.cs13@nycu.edu.tw

主題介紹 (Topics Introduction)

一、搜尋中的個人化與推薦 (Personalization and Recommendations in Search)

隨著使用者對更豐富、更直觀的搜尋體驗需求日益增加，**個人化與推薦** 在現代資訊檢索中扮演了核心角色。搜尋系統正在從傳統的「靜態查詢—文件比對」模式，逐漸演變為 **動態、以使用者為中心** 的服務，納入了 **情境 (context)**、**偏好 (preferences)**、**行為訊號 (behavioral signals)**、以及**多模態線索 (multimodal cues)**。

大型語言模型 (LLMs)、生成式推薦、多模態表示學習以及強化學習的整合，已經徹底改變了個人化搜尋。這些進展讓系統能更好地捕捉使用者意圖、減少冷啟動問題、支持跨領域學習，並提供 **對話式、互動式** 的推薦體驗。

然而，仍存在一些關鍵挑戰：

- 在個人化與 **公平性、隱私** 之間取得平衡
- 減少生成式推薦中的 **幻覺與偏差**
- 確保系統在 **網路規模** 下仍具備可擴展性與即時回應能力

要解決這些問題，將是下一代個人化搜尋系統的關鍵任務。

相關研究主題包括但不限於：

- 個人化搜尋與推薦模型
- 搜尋中的生成式推薦
- 對話式與情境感知的推薦系統
- 搜尋中的序列推薦與使用者行為建模
- 跨領域個人化與遷移學習
- 多模態個人化（文本、圖像、影片、圖表）
- 知識圖譜強化的個人化與推薦
- 公平性、隱私與安全性挑戰
- 使用者互動與可解釋性（透明性與信任）
- 即時、大規模個人化搜尋系統設計

- 對抗式攻擊與防禦
- 評估指標：多樣性、新穎性、覆蓋率、使用者滿意度
- 人機協作與互動設計

二、代理型資訊檢索 (Agentic IR)

隨著人工智慧技術的整合，資訊檢索領域已經發生了顯著的轉變。**AI 代理 (AI agents)**，特別是結合 **大型語言模型 (LLMs)** 與強大運算能力的代理，徹底革新了資訊檢索、處理與呈現的方式。

LLM 代理具備 **工具調用 (tool-call)**、**進階記憶 (advanced memory)**、**推理與規劃能力 (reasoning and planning)**，因此能夠：

- 執行複雜任務
- 進行連貫對話
- 提供個人化的回應

儘管進展顯著，但仍存在一些挑戰，例如：

- 確保檢索結果的 **相關性與準確性**
- 減少 **偏差 (biases)**
- 提供 **即時回應**
- 維護 **資料安全**

相關研究主題包括但不限於：

- 代理型檢索系統 (Agentic Retrieval System)
- 代理型對話式推薦 (Agentic Conversational Recommendation)
- AI 代理應用於個人化 (AI Agent for Personalization)
- AI 代理應用於序列推薦 (Sequential Recommendation)
- AI 代理應用於情境化資訊檢索 (Contextual IR)
- 跨語言與多模態檢索的 AI 代理
- 檢索增強生成 (RAG) 搜尋系統
- AI 代理檢索模型的最佳化
- 偏差緩解 (Bias Mitigation)
- 可解釋的生成式檢索系統

- 對抗式攻擊與防禦
- 代理型資訊檢索的倫理考量
- 人機協作於資訊檢索中的應用
- 代理型資訊檢索的評估
- AI 代理檢索的可擴展性與效率

三、多模態檢索 (Multimodal Retrieval)

隨著數位內容持續多樣化，使用者越來越需要能夠跨越 文字、影像、影片、音訊、表格與圖表 進行搜尋。多模態檢索的目標是統一不同模態的表示與理解，讓搜尋系統能夠提供更精確、自然且互動式的資訊存取。

近期的進展，如：

- 大型語言模型 (LLMs)
- 多模態預訓練模型（例如 **CLIP**、**BLIP**、**影片語言模型**）
- 基於擴散模型的生成技術

這些突破大幅推動了多模態檢索的發展，使得系統能將異質模態（heterogeneous modalities）對齊並嵌入共享語意空間，支援以下任務：

- 文本檢索影像 (text-to-image retrieval)
- 語音檢索影片 (speech-to-video retrieval)
- 文本檢索圖表 (text-to-chart retrieval)

然而，多模態檢索仍面臨重大挑戰：

- 跨模態的語意對齊與融合
- 細粒度檢索與可解釋性（例如：定位影像/影片的相關片段）
- 多語言與跨文化理解
- 大規模系統的效率

由於多模態檢索在教育、醫療、娛樂、電商、社群平台等領域有越來越多應用，如何在效能、透明度與公平性之間取得平衡，已成為迫切且重要的研究方向。

相關研究主題包括但不限於：

- 多模態檢索模型設計（文字-圖像、文字-影片、音訊-視覺、表格/圖表）
- LLM 與多模態檢索的整合

- 生成式多模態檢索
- 細粒度檢索與可解釋性（例如：定位相關的影像/影片片段）
- 多語言與跨文化的多模態檢索
- 對話式與互動式的多模態檢索
- 即時、大規模的多模態檢索系統設計
- 多模態模型中的對抗式攻擊與魯棒性
- 多模態檢索的評估基準（正確率、多樣性、可解釋性）
- 多模態檢索應用中的人機協作

四、領域專用的檢索增強生成 (Domain-Specialized RAG)

檢索增強生成（RAG）已在開放領域應用中展現出強大潛力，但在 **臨床醫學** 和 **法律** 等專業領域中，仍面臨更嚴峻的挑戰，例如：**準確性、可解釋性與合規性**。

在這些高風險領域中，錯誤的容忍度極低，任何幻覺（hallucination）或不精確的回答，都可能導致嚴重後果。因此，針對特定領域需求調整 RAG 系統，是未來研究的重要方向。

- **在臨床應用中：**RAG 系統必須能整合醫學知識庫、電子病歷、臨床指引與學術文獻，提供基於證據的建議與解釋，同時保護病患隱私與資料安全。
- **在法律應用中：**RAG 系統需要處理判例法、法規、契約與訴訟文件，並具備高精確度，確保引用的真實性與可追溯性，以協助律師與研究人員在複雜情境下進行工作。

主要挑戰包括：

- 建立高品質的領域語料庫
- 減少幻覺並提升引用的真實性
- 整合領域專用術語與知識圖譜
- 支援案例法與醫學文獻的結構化檢索與情境建模
- 在確保即時回應的同時，維持法規遵循

隨著醫療與法律領域對 AI 輔助系統的需求日增，領域專用 RAG 在研究與實務應用上將愈加重要。

相關研究主題包括但不限於：

- 臨床與法律領域的專用 RAG 系統設計
- 高保真度的檢索與引用生成

- 整合領域術語與知識圖譜於 RAG 系統中
- 案例法與醫學文獻的結構化檢索與情境建模
- 專用問答系統中的幻覺偵測與抑制
- 可解釋性與可追溯性
- 多語言臨床/法律 RAG 的挑戰與機會
- 人機協作：醫師與律師如何與 RAG 系統協同工作
- 專業領域 RAG 的對抗性風險與錯誤資訊
- 實務案例：臨床決策支持與法律研究輔助
- 專用 RAG 的基準與評估指標（準確性、可靠性、可解釋性）

四、領域專用的檢索增強生成 (Domain-Specialized RAG)

檢索增強生成（RAG）已在開放領域應用中展現出強大潛力，但在 **臨床醫學** 和 **法律** 等專業領域中，仍面臨更嚴峻的挑戰，例如：**準確性、可解釋性與合規性**。

在這些高風險領域中，錯誤的容忍度極低，任何幻覺（hallucination）或不精確的回答，都可能導致嚴重後果。因此，針對特定領域需求調整 RAG 系統，是未來研究的重要方向。

- **在臨床應用中：**RAG 系統必須能整合醫學知識庫、電子病歷、臨床指引與學術文獻，提供基於證據的建議與解釋，同時保護病患隱私與資料安全。
- **在法律應用中：**RAG 系統需要處理判例法、法規、契約與訴訟文件，並具備高精確度，確保引用的真實性與可追溯性，以協助律師與研究人員在複雜情境下進行工作。

主要挑戰包括：

- 建立高品質的領域語料庫
- 減少幻覺並提升引用的真實性
- 整合領域專用術語與知識圖譜
- 支援案例法與醫學文獻的結構化檢索與情境建模
- 在確保即時回應的同時，維持法規遵循

隨著醫療與法律領域對 AI 輔助系統的需求日增，領域專用 RAG 在研究與實務應用上將愈加重要。

相關研究主題包括但不限於：

- 臨床與法律領域的專用 RAG 系統設計

- 高保真度的檢索與引用生成
- 整合領域術語與知識圖譜於 RAG 系統中
- 案例法與醫學文獻的結構化檢索與情境建模
- 專用問答系統中的幻覺偵測與抑制
- 可解釋性與可追溯性
- 多語言臨床/法律 RAG 的挑戰與機會
- 人機協作：醫師與律師如何與 RAG 系統協同工作
- 專業領域 RAG 的對抗性風險與錯誤資訊
- 實務案例：臨床決策支持與法律研究輔助
- 專用 RAG 的基準與評估指標（準確性、可靠性、可解釋性）

此專題需求的部分靈感直接來自 **Stanford CS224N** 的專題說明。

參考連結：<https://web.stanford.edu/class/cs224n/project/project-proposal-instructions-spr2024-updated.pdf>