

# 醫學電資整合創意專題 (二) 書面報告

## 智慧醫療對話系統：ERAS 輔助應用

組員：傅冠豪、金哲安、陳孟潔、陳冠宇、楊哲瑜、倪昕、張玠

指導教授：吳沛遠      助教：林仲偉

### 目錄

1 介紹	2
2 動機	2
3 前期準備	2
3.1 系統架構與模型選型	3
3.2 資料建置與知識整理	3
3.3 介面設計與功能原型	3
4 系統架構	3
4.1 使用者互動與個案管理	3
4.2 智慧語言模型模組	4
4.3 資料庫與知識整合	4
5 問題與挑戰	4
5.1 生成回覆難以量化	4
5.2 資料格式與評估不一致	4
5.3 地端模型資源限制	4
5.4 跨語言語意落差	4
6 實作及技術方案	5
6.1 多代理架構設計與任務分工	5
6.2 QA Benchmark	5
6.3 Closed-ended 回覆評估方式	5
6.4 Open-ended 回覆評估方式	6
6.5 人工專家評估	7
6.6 VQA Benchmark	7
7 未來展望	9
8 結論	9
9 組員貢獻	9
10 Demo 影片	9

## 1 介紹

隨著實證醫學（Evidence-Based Medicine, EBM）的持續發展，現代醫療體系對於手術病患的術前、術中與術後照護流程提出愈加精細的規範。其中，ERAS（Enhanced Recovery After Surgery，加速術後康復）計畫作為一項多學科整合的臨床路徑設計，已在多種外科領域中展現出顯著的臨床效益，如減少併發症、縮短住院天數並提升病患整體預後。然而，目前 ERAS 指引多以敘述性文件形式呈現，實際執行仍仰賴醫療團隊的高度協作，包含手術醫師、麻醉科、護理師、營養師與復健師等角色，導致在資訊傳遞、教育衛教及流程落實上存在不小挑戰。

本專案聚焦於結合先進自然語言處理技術與臨床照護場景，嘗試以大型語言模型（Large Language Model, LLM）為核心，導入 RAG（Retrieval-Augmented Generation）與高效參數微調 LoRA（Low-Rank Adaptation）等技術架構，打造一套具備知識檢索、語意生成與互動理解能力的智慧醫療對話系統。此系統專為 ERAS 指引場景設計，期望能在提供高品質醫療知識的同時，降低臨床人力負擔，提升病患的術後參與度與資訊可及性，進一步協助醫療團隊落實智慧照護作業流程。

## 2 動機

儘管 ERAS 計畫已廣泛被臨床實踐驗證為有效的術後照護指引，其實際推行仍面臨多項挑戰。特別是在臨床人力有限或資源緊張的醫療情境中，跨部門溝通成本高、照護規範複雜且動態變化，加上缺乏結構化的數位支持系統，往往導致部分核心措施無法被充分落實。現行作法高度依賴個別醫療人員進行即時管理與經驗傳遞，缺乏具備邏輯解釋與任務導向功能的智慧平台支援，亦無法提供病患即時、可互動且具個人化的衛教介面。

另一方面，ERAS 涉及的知識高度跨領域，涵蓋外科、麻醉、護理、營養與復健等模組，這些資訊多散見於臨床指引、內部文件與醫學文獻中，缺乏一個能有效統整、檢索與生成的智慧系統。此外，現行病患教育資源多為靜態文件，難以依據個人狀態做即時調整，亦缺乏互動回饋機制，不利於提升其衛教參與度。

因此，本專案的核心動機在於結合自然語言處理和電腦視覺技術，建構一套以 LLM 為基礎，結合 RAG 與 LoRA 等技術之智慧 ERAS 輔助平台。此平台期望實現以下四項核心目標：

- **建立互動式任務管理平台：**引導病患完成術前術後之分科照護任務，提升主動參與度與依從性。
- **整合多團隊作業流程：**將 ERAS 中跨部門之照護規範進行標準化與結構化，並實現可追蹤之數位流程管理。
- **導入 LLM 衛教問答系統：**支援以自然語言進行動態互動，依據病患輸入提供個人化衛教回應與任務提醒。
- **開發視覺問答（VQA）功能：**整合術後醫學影像與報告，使 LLM 能生成淺顯易懂的圖文解釋，協助病患理解術後狀況。

透過上述技術導入與流程再設計，本系統期望有效連結臨床照護專業與智慧醫療應用，降低實務執行落差，並為 ERAS 模型在數位轉型下提供全新運作範式。

## 3 前期準備

為實現 ERAS 智慧照護平台的整體功能，團隊於初期即完成系統架構規劃、資料建置、模型選型與介面設計，奠定開發基礎。

### 3.1 系統架構與模型選型

本系統以 LLM 為核心，整合 RAG 與 LoRA 技術，支援醫療語境下的語意檢索與生成。初步選用適合地端部署之模型（如 Qwen2.5 系列），並搭配 multilingual 嵌入模型構建語意索引，完成基本推論與資料流驗證。

### 3.2 資料建置與知識整理

團隊蒐集多份 ERAS 醫學文獻與臨床指引，進行 PDF-to-TXT 轉換與內容清理，並根據 ERAS 照護節點劃分為 **麻醉科、營養科、藥劑科與復健科** 四大類，對應實際術別與照護任務（如術前營養支持、術後止痛等）。進一步彙整內容為主題導向分類，建立語意可檢索之 ERAS 知識庫。

### 3.3 介面設計與功能原型

系統前端提供聊天式互動介面與個人化 To-Do List，支援病患依手術情境完成對應任務與提問互動。初步功能已實作 ERAS 協定邏輯與 QA 模組，可根據輸入查詢知識庫並生成語意回答，完成整體平台之最小可行原型。

## 4 系統架構

本系統整合語言模型推論、視覺分析與個案管理功能，支援從術前到術後的智慧互動照護。架構如圖 1 所示，分為三大模組：**前端介面與個案管理**、**LLM 智慧互動模組**、以及**知識庫與後端資料處理**，協助醫病雙方落實 ERAS 指引與個人化任務導引。

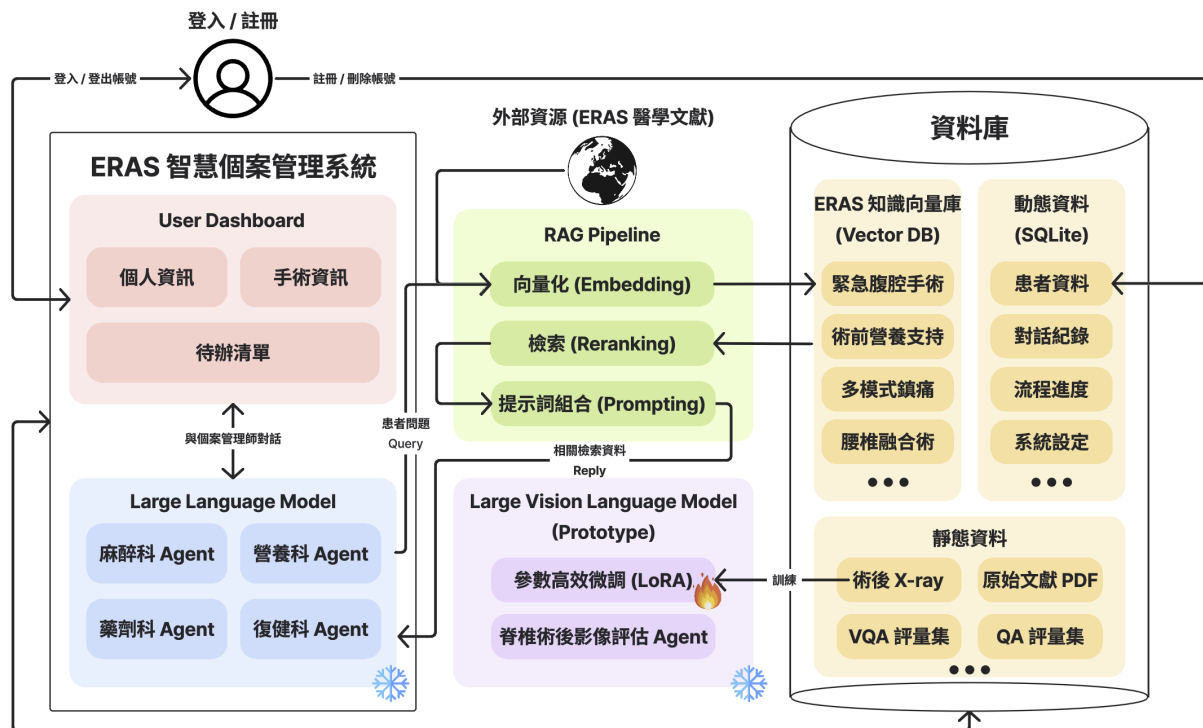


Figure 1: ERAS 智慧照護系統架構圖

### 4.1 使用者互動與個案管理

病患登入 ERAS 智慧個案管理系統後，可透過整合式介面檢視個人資料、手術資訊與 To-Do 清單。系統串接四個分科智能代理（麻醉科、營養科、藥劑科、復健科），用以回應病患提出之自然語言問題，並由後端語言模型模組產生個人化回覆。

## 4.2 智慧語言模型模組

本平台語言處理核心結合 Large Language Model 與 RAG Pipeline 架構，實現以問答為導向的 ERAS 知識應用。當病患提出查詢時，系統會透過 RAG 模組進行三階段處理：

1. **向量化 (Embedding)**：將使用者輸入與知識庫內容進行語意嵌入轉換。
2. **檢索 (Reranking)**：根據語意相關性從知識庫檢索最匹配段落。
3. **提示詞組合 (Prompting)**：將檢索結果與使用者問題融合，構建最終輸入給語言模型進行回答生成。

語言模型模組採地端部署架構，結合語意檢索與自然語言生成，支援病患查詢與知識回應等互動需求，並可透過微調機制提升特定場景下的問答準確性與語意理解能力。此外，系統亦規劃建置視覺語言模型模組，處理術後影像（如脊椎融合術後 X-ray）之語意轉譯，支援醫療影像輔助問答（VQA），提升醫病圖像溝通效率。

## 4.3 資料庫與知識整合

後端資料處理由兩大類資料庫組成：

1. **ERAS 向量知識庫 (Vector DB)**：包含以主題分類整理的知識段落，如緊急腹腔手術、術前營養支持、多模式鎮痛與腰椎融合術等，作為語意檢索與問答生成依據。
2. **動態資料庫 (SQLite)**：儲存病患資料、對話紀錄、流程進度與系統設定，支援使用者歷程追蹤與介面狀態同步。

另包含靜態資料區，用於儲存原始 PDF 文獻、術後影像（如 X-ray）、QA 與 VQA 評量集，作為模型訓練與評估之基礎資料來源。

## 5 問題與挑戰

本專案雖已建構 ERAS 智慧照護平台原型，但在落地應用與系統驗證過程中仍面臨多項挑戰，涵蓋自然語言生成品質、資料一致性、部署環境限制與跨語言處理等層面。

### 5.1 生成回覆難以量化

醫療問答需具備高度的準確性、相關性與忠實度，然而現有自動化指標難以全面反映 LLM 回覆品質。開放式問題缺乏唯一標準答案，增加量化困難；人為評估雖可信，但成本高、難以規模化。

### 5.2 資料格式與評估不一致

現有 QA 資料集來源多樣，格式與語言風格不一，增加前處理與標準化成本，也難建立通用評估基準，降低模型比較與結果再現性。

### 5.3 地端模型資源限制

為保資料隱私與本地部署彈性，系統採用地端模型，但受限於記憶體、模型大小與推論速度，影響即時互動的穩定性與效能。

### 5.4 跨語言語意落差

知識庫多為英文文獻，使用者查詢以中文為主，導致語意匹配困難。中英文術語具多義性，降低向量檢索與生成表現，也影響回答的準確性與覆蓋範圍。

## 6 實作及技術方案

### 6.1 多代理架構設計與任務分工

為因應地端模型效能限制，本系統採用多代理架構，將任務依照功能拆分為語言與影像兩大類型代理（Agents），分別處理病患問答、衛教導引與術後影像報告生成。此分工不僅降低單一模型負載，也提升系統回應效率與模組彈性，有助於實現可擴充的智慧照護流程。

### 6.2 QA Benchmark

為了能有效評估模型在不同語意任務下的表現，本專案建構一份涵蓋多面向問題的 QA Benchmark。資料內容來自 ERAS 各科別指引，問題依據語言型態分為 **封閉式（Closed-ended）** 與 **開放式（Open-ended）**，兩者在問法與回覆風格上有顯著差異，如表 1 所示。

類型	Closed-ended	Open-ended
科別	麻醉科	藥劑科
問題	剖腹產前可以吃東西嗎？	術後麻醉復甦階段的優先重點？
答案	您可以在手術前 <b>6</b> 小時吃輕食。	快速清醒與疼痛控制平衡。

Table 1: 封閉式與開放式問題設計對照表

根據統計，初步建構之 QA Benchmark 共收錄超過 120 筆問答，涵蓋四大科別（麻醉科、營養科、藥劑科、復健科），如圖 2 所示。

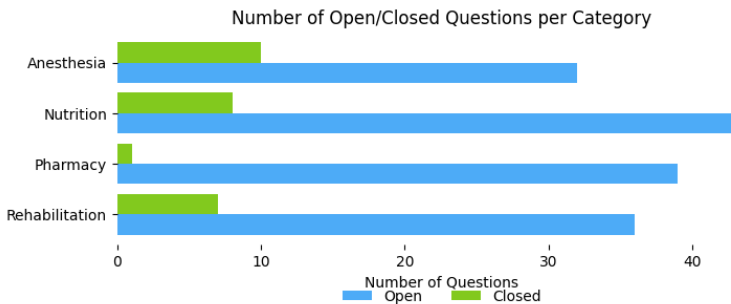


Figure 2: 各科別開放式／封閉式問題數量統計圖

### 6.3 Closed-ended 回覆評估方式

對於封閉式問題，我們採用關鍵詞比對為基礎的自動評估方法。只要模型輸出中含有 Ground Truth 所定義的正確關鍵詞，即視為正確，並記為準確率 1.0。此類問題具有明確答案，適合以文字匹配方式快速計算。以下為實例說明：

Question	剖腹產前可以吃東西嗎？
Ground Truth	會使用動脈導管進行侵入性心血管監測，確保您的平均動脈壓（MAP）接近正常值，且至少高於 <b>65 mmHg</b> 。
Model Answer	根據 ERAS 規範...應避免 MAP 下降至 <b>65 mmHg</b> 以下，以降低急性腎臟和心肌損傷的風險。此外，在某些情況下...
Accuracy	1.0

Table 2: 封閉式問答範例與準確率評估

準確率計算方式如下：

$$\text{Accuracy} = \frac{1}{N} \sum_{i=1}^N \mathbb{I}[\text{GT}_i \subseteq \text{Answer}_i] \quad (1)$$

其中， $N$  為問題總數， $\mathbb{I}$  為指示函數，當模型回覆中包含 Ground Truth 中的關鍵片段即記為正確。各科別封閉式問題之平均準確率如圖 3 所示。

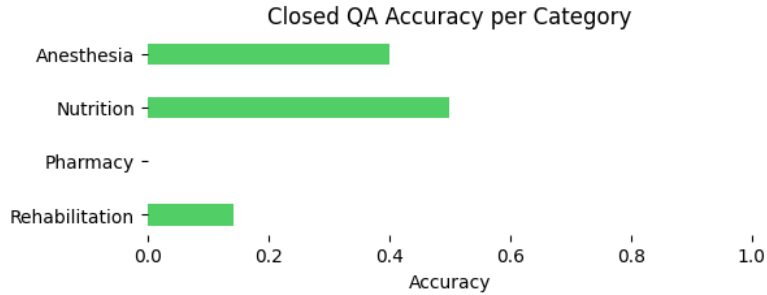


Figure 3: 各科別封閉式問答準確率統計

#### 6.4 Open-ended 回覆評估方式

針對開放式問題，我們初步採用 BLEU、ROUGE、METEOR 等評估指標，但在中文醫療場景下效果有限。後續引入 **BERTScore** 進行語意相似度評估，儘管引入 RAG 後部分科別略有提升，如圖 4 所示，整體分數仍偏低，顯示現有方法對醫療問答涵蓋性評估仍具挑戰。

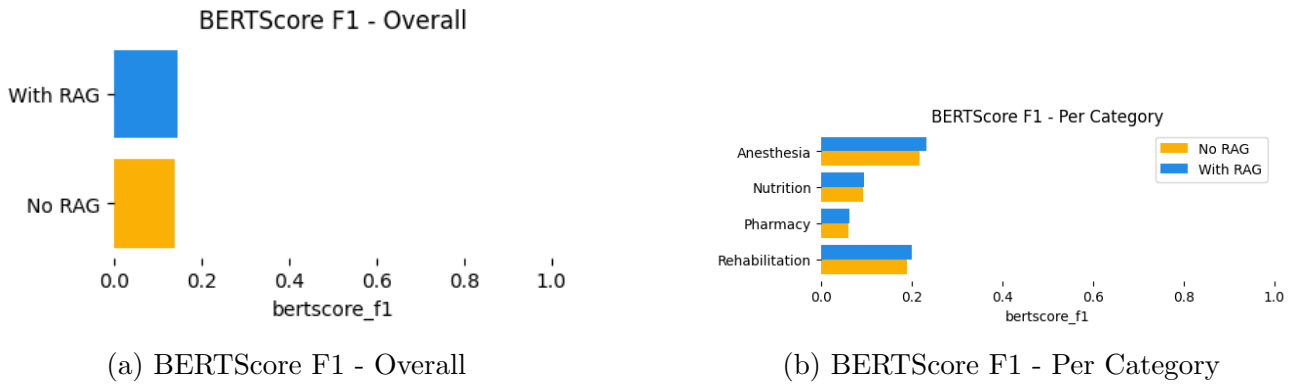


Figure 4: 引入 RAG 前後的 BERTScore 評估結果

因此，我們最終採用更具語意對應性的 **token-level recall** 作為主要評估方式。該方法以不重複中英文字元為基本單位，計算模型回覆與 Ground Truth 之交集比例，以衡量語意涵蓋程度。以下為評估實例：

Question	剖腹產前是否應該戒菸？
Ground Truth	是的，孕早期戒菸可減少胎兒體重過輕、早產和死胎等風險。
Model Answer	根據 ERAS... 剖腹產前應鼓勵患者戒菸... 戒菸還能降低手術後的併發症風險，促進更快的康復。因此，為了母親和胎兒的健康...
Recall	$(17 - 1) / (35 - 11) = 0.667$

Table 3: 開放式問答範例與 Recall 評估

Recall 定義如下：

$$\text{Recall} = \frac{|\text{Tokens}_{\text{GT}} \cap \text{Tokens}_{\text{Ans}}|}{|\text{Tokens}_{\text{GT}}|} \quad (2)$$

此處  $\text{Tokens}_{\text{GT}}$  表示 Ground Truth 回覆中的不重複 token 集合， $\text{Tokens}_{\text{Ans}}$  則為模型回答的 token 集。此指標反映模型回覆中涵蓋正確資訊的比例，適用於開放式問題之語意涵蓋度評估。

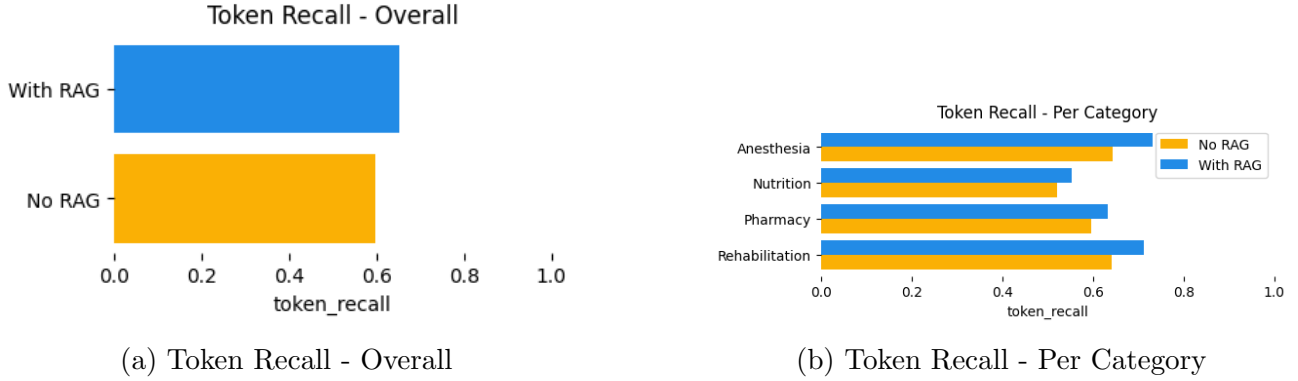


Figure 5: 引入 RAG 前後在 Open-ended QA 的 token-level recall 比較

## 6.5 人工專家評估

為驗證系統在實際臨床場景的可用性，我們邀請臨床醫師進行人工評估，針對模型回覆的正確性、可讀性與臨床合理性進行標註。結果如圖 6 所示，超過八成回覆獲評為「可接受」以上，顯示系統具良好實用潛力。然而，相較於 token-level recall 或 BERTScore 等自動化指標，人工評估結果仍存在一定差異，突顯目前評估方法與人類判準間的落差，未來仍需發展更貼近醫療語境與人類偏好的量化指標。

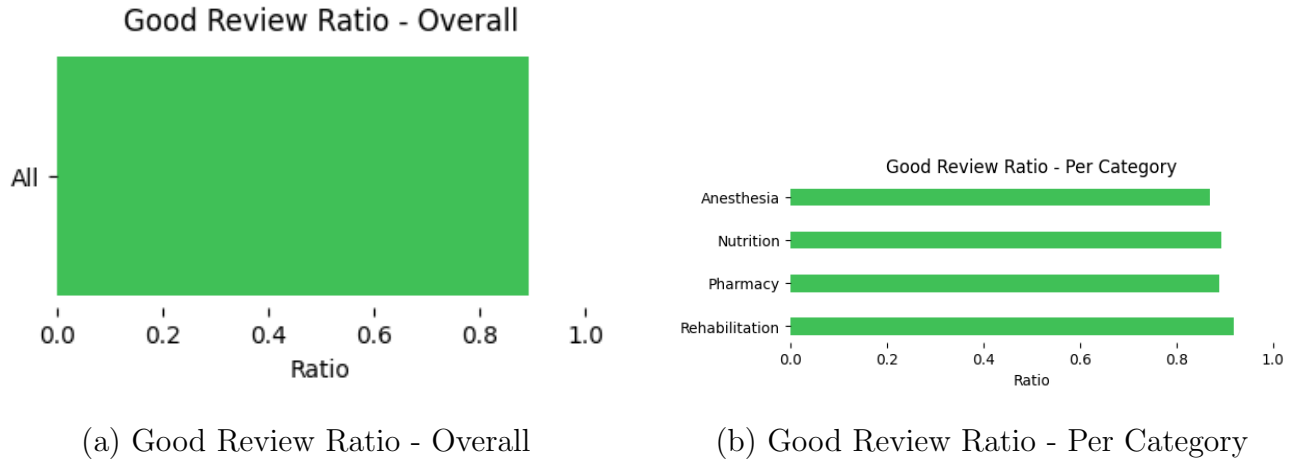


Figure 6: 醫師人工審閱結果：回答通過率統計（整體與分科）

## 6.6 VQA Benchmark

為探討語言模型於視覺輸入下生成能力的潛力，我們設計小型 VQA Benchmark，任務為生成術後脊椎影像之結構化放射報告，屬於 open-ended QA 類型，輸入為 Spine X-ray，模型需生成完整語意描述。



### Prompt (Q)

You are a radiologist. Please write a radiology report for the given spine X-ray images in the following structured format:

Radiography of [Spine Region] ([View(s)]) show:

- [Finding 1]
- [Finding 2]
- ...

### Ground Truth (A)

Radiography of L-S Spine (A-P View, Lateral View) show:

- Degenerative change of lumbar spine with spur formation.
- S/P transpedicular screw fixation at L2-S1.
- S/P disc prosthesis at L3-S1.
- Disc space narrowing at L1-2 level.

為實作微調流程，我們採用 **3-fold cross-validation** 將訓練資料等分為三組，輪流作為驗證集進行微調，最終以權重平均的方式整合模型。資料切分如圖 7 所示。

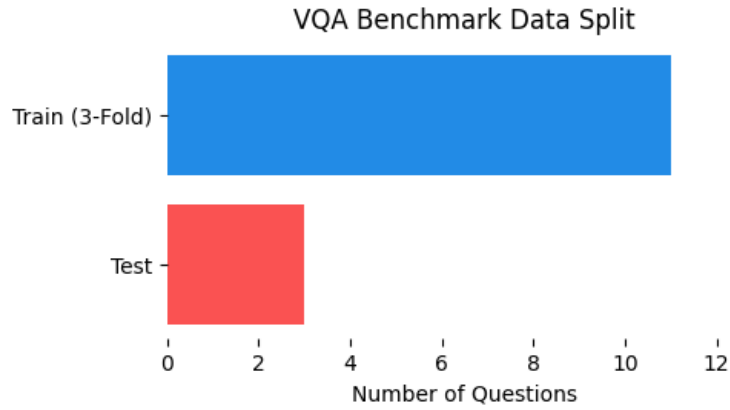
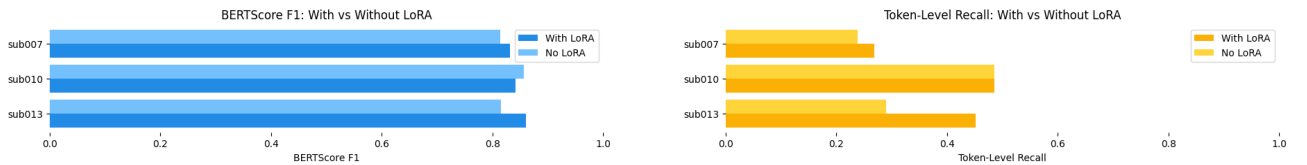


Figure 7: VQA Benchmark 資料切分方式 (3-fold + 測試集)

最終合併模型參數  $\theta$  採以下方式平均：

$$\theta_{\text{final}} = \frac{1}{K} \sum_{k=1}^K \theta^{(k)} \quad (3)$$

其中  $K = 3$  為 fold 數， $\theta^{(k)}$  表示第  $k$  次微調所得之 LoRA 權重。此設計有助於避免單一 fold overfitting 並提升模型穩定性，作為實驗性 VQA 報告生成系統之基礎。為評估 LoRA 微調對術後影像報告生成品質的影響，我們分別以 **BERTScore F1** 與 **token-level recall** 作為評估指標，針對多筆測試資料進行比較分析。結果如圖 8 所示，在多數樣本中導入 LoRA 可提升生成內容的語意覆蓋與語句品質，顯示微調對 VQA 任務具正面效益。



(a) BERTScore F1: w / w/o LoRA

(b) Token-Level Recall: w / w/o LoRA

Figure 8: LoRA 微調對 VQA 報告生成品質的影響 (不同受測資料 subXXX)



## 7 未來展望

本專案已初步建立一套結合大型語言模型與知識檢索技術的 ERAS 智慧照護系統，未來可朝以下方向持續優化與擴充：

- **強化個人化互動**：深化病患註冊資料與模型回覆內容間的連結性，進一步導入病歷資訊、術式類型與時程等個別條件，以提供具臨床語境的個人化建議。
- **擴充知識庫規模**：持續收錄最新 ERAS 指南、國際臨床實證資料與相關專業文獻，確保系統具備即時性與醫療專業完整性。
- **優化資料處理流程**：進一步提升文獻轉檔、主題分類與語意嵌入效率，簡化知識更新流程，強化系統維運與拓展能力。
- **改善地端部署效能**：測試更多適用於地端推論的語言模型與加速工具，增強多語系支持與臨床反應速度，以符合實務應用對資安與效能的需求。

## 8 結論

本專案成功建構一套整合大型語言模型、知識檢索與視覺問答功能的 ERAS 智慧照護系統，涵蓋結構化知識庫建置、互動介面設計與多模組任務導引。系統透過引入 Retrieval-Augmented Generation (RAG) 與 Low-Rank Adaptation (LoRA) 技術，有效提升語意檢索能力與生成表現，使病患能獲得即時、個別化的衛教支援與照護建議。

系統設計亦兼顧地端部署、資料隱私與臨床彈性，初步驗證其在醫療實務中具備實用潛力。整體而言，本專案於技術研發與應用實作上皆達成預期目標，未來可進一步導入多科別臨床資料，擴大系統適用範疇，成為智慧醫療輔助決策的重要基礎建設。

## 9 組員貢獻

組員	主要負責內容
傅冠豪	概念發想、組織架構
金哲安	網頁前端、網頁後端
陳孟潔	網頁前端、網頁後端
陳冠宇	LLM 系統架構整合 (RAG、LoRA)、QA 及 VQA 集評估、簡報、書面報告
楊哲瑜	QA 集設計、查找、整理
倪昕	QA 集設計、查找、整理、書面報告
張玠	一點 QA 集、Prompt 設計、查找、整理、簡報

Table 4: 組員分工與貢獻列表

## 10 Demo 影片

[https://drive.google.com/file/d/1ajuhaVhlmc\\_Mz1BA5ZZQ3UY3hVo9yzNM/view](https://drive.google.com/file/d/1ajuhaVhlmc_Mz1BA5ZZQ3UY3hVo9yzNM/view)