

出國報告(出國類別:開會-參加 AMIA Annual Symposium)

參加 AMIA Annual Symposium

服務機關：臺中榮民總醫院數位醫學部人工智慧科

姓名職稱：吳晨豪/契約高級資訊工程師

派赴國家/地區：美國/亞特蘭大

出國期間：114年11月13日至114年11月21日

報告日期：000年0月0日

摘要

此次前往美國亞特蘭大參加 AMIA 年會，讓我對國際醫療資訊與生成式 AI 的發展有更全面的理解。研討會內容涵蓋大型語言模型、資料治理、FHIR 標準、多代理架構與臨床應用等議題，許多分享都與我目前在院內進行的 AI 開發與資料處理工作高度相關。透過多場技術與研究成果展示，我深刻感受到國外在模型透明度、流程設計與資料標準化上的成熟度，尤其強調模型可控性、可稽核性以及與臨床工作流程的緊密整合，這與國內現階段仍在強化的部分形成鮮明對比。

期間也與國際研究者進行交流，其中包含討論大型語言模型在醫療領域的可靠度與偏誤問題，這些互動讓我重新思考 AI 在臨床場景中的安全性與使用者信任度。同時，在海報展區看見台中榮總的研究成果展出，更讓我感到驕傲，也意識到台灣在醫療資料與 AI 研究上的深度與潛力。

整體而言，本次 AMIA 行程讓我獲得技術、觀念與視野上的多重收穫，並對未來在院內推動 AI 落地、資料治理與跨領域合作有了更明確的方向。

關鍵字：(至少一組)

資訊學、臨床資訊學、醫療健康資料、醫療人工智慧、數位健康

目次

一、 目的	1
二、 過程	2
三、 心得	13
四、 建議事項	14
(一) 強化院內資料治理與語意標準化概念的導入	
(二) 規劃 LLM (SFT/LoRA/RAG) 的小規模驗證與實作	
(三) 逐步導入模型解釋性 (Explainability) 與多代理 (Multi-Agent) 流程概念	
(四) 鼓勵持續參與國際研討會以提升團隊視野	
五、 附錄	15

一、目的

在全球醫療科技迅速變動的今日，醫療資訊學（Medical Informatics）已經成為連結臨床、資料科學與醫療政策的關鍵橋樑。尤其在大型語言模型（Large Language Models, LLM）、臨床決策支援（CDS）、電子病歷（EHR）與多模態醫療資料迅速擴張的背景下，如何將科技有效、安全且可靠地導入臨床流程，逐漸成為全球醫療系統的核心課題。

AMIA（American Medical Informatics Association, AMIA）Annual Symposium 自 1977 年創辦以來，被視為全球醫療資訊領域最具代表性且歷史悠久的學術盛會。其特色不僅在於議題的廣度，更在於它結合了醫療機構、政府單位、AI 研究者、醫師、護理、產業端與政策制定者，形成真正跨領域的核心論壇。AMIA 的精神，是希望透過醫療資訊系統的整合、資料標準化與新興人工智慧的發展，讓醫療照護更安全、更有效、也更可持續。我此次以臺中榮民總醫院契約高級資訊工程師（AI/LLM、疾病預測模型開發）的身分參與 AMIA。

在全球醫療 AI 加速發展的背景下，掌握國際醫療人工智慧的最新動向顯得格外重要。隨著生成式 AI 與大型語言模型（Large Language Models, LLM）技術快速突破，國際醫療界已逐漸從「研究概念驗證」走向「真實臨床部署」。透過此次 AMIA 的參與，我觀察到許多大型電子病歷（Electronic Health Records, EHR）供應商，如 Epic 與 Oracle Health，已將 AI 直接融入醫師看診、護理紀錄、風險預警、病歷自動生成與族群健康管理等實際工作流程。在這樣的趨勢下，我希望藉由 AMIA 全面了解國外醫院是如何克服技術、資料與流程方面的挑戰，讓 AI 得以順利落地；同時也希望理解哪些應用已經通過實證並量化效益，醫護人員又是如何在繁忙的臨床環境中與新科技協同工作，及其成功因素與可能的瓶頸。這些觀察對台灣的醫療 AI 推動具有重要參考意義，也能協助我思考未來院內 AI 發展的方向。

除了掌握趨勢之外，我也著重於強化自身對大型語言模型應用技術的理解與實作能力。目前我在院內負責多項人工智慧專案，內容涵蓋疾病預測模型之建立與驗證，例如深靜脈栓塞（Deep Vein Thrombosis, DVT）、急性心肌梗塞（Acute Myocardial Infarction, AMI）、敗血症（Sepsis）等；同時也負責醫療資料清洗、欄位標準化、特徵工程，以及大型語言模型的微調與醫療語料處理。此外，我也參與 AI 系統的全流程建置，包括從資料處理、模型建立，到模型部署與後續效能監測等階段。因此，本次 AMIA 的實作課程對我特別重要，例如 W13 中針對大型語言模型微調的教學，讓我能實際操作低秩適應技術

（Low-Rank Adaptation, LoRA）、監督式微調（Supervised Fine-Tuning, SFT）、檢索增強生成（Retrieval-Augmented Generation, RAG）與人類回饋強化學習（Reinforcement Learning from Human Feedback, RLHF）等核心技術。這些內容不僅補強了我的技術能力，也讓我對如何在醫院建構屬於自己的 LLM 架構與應用生態系有了更清晰的藍圖。

在 AI 技術持續演進的同時，我也深刻感受到資料標準化與資料治理的重要性。AI 若要真正落地臨床，最大的關鍵並非模型效能，而是資料品質與資料架構是否完善。國外醫療體系之所以能讓 AI 穩定運作，關鍵在於全面採用 FHIR（Fast Healthcare Interoperability Resources）資料標準、建立語意資料層（Semantic Layer）、以及具備完善的資料治理框架（Data Governance）。這些措施讓病歷資料、影像資料、生命徵象甚至基因資料得以互相

整合，並成為 AI 訓練與部署的穩固基礎。透過 AMIA，我希望更深入了解美國醫療機構如何處理如此多樣化且高複雜度的資料，也希望理解像 Oracle Health 這類企業如何透過資料平台整合 AI，並思考台灣是否能逐步建立類似的資料基礎架構，縮短與國際間的差距。

此外，此次 AMIA 也讓我有機會與許多來自國際的研究者討論醫療資料平台的應用差異。國外研究界普遍使用 All of Us (AoU) 資料平台，而台灣醫院目前多採用 TriNetX。兩者各具特色：AoU 的 Cohort 工具成熟，能直接生成研究所需的圖表，且資料面向涵蓋基因、穿戴式裝置與問卷資料，更加多元；相較之下，TriNetX 在臨床查詢上較有優勢，但研究工具功能相對簡略。透過深入比較，我更加理解兩者在定位上的不同，也更能思考未來院內資料平台如何進一步擴充、整合，讓研究與臨床場景都能獲得最大效益。綜合上述目的而言，此次 AMIA 的參與對我而言不僅是接觸最新技術的機會，更是深入理解國際醫療 AI 生態、觀察真實落地案例、並反思如何把這些經驗帶回台灣與院內的重要契機。我希望能以此為起點，逐步協助院內建立更成熟、可靠且可持續運作的 AI 應用架構，為未來臨床 AI 的推廣奠定更穩固的基礎。

二、過程

(一) 臺灣桃園機場出發與美國亞特蘭大抵達-至現場報到

1. 長途旅程與轉機經驗

此次 AMIA 研討會的行程從 11 月 13 日正式展開，由於地點位於美國喬治亞州亞特蘭大，必須搭乘長程航班才能抵達。從台灣出發後，第一段航程約 13 小時，目的地為舊金山。長時間飛行不僅容易造成疲勞，也需要同時面對時差的調整、機艙環境的不舒適與入境前的心理壓力。

抵達舊金山後，由於航班因為鳳凰颱風氣流導致日本上空航線暫時停飛所以延誤至一小時，原先預留的轉機時間被大幅壓縮，使得必須在有限時間內迅速完成所有轉機程序，包括通過入境查驗、重新托運行李與再次安檢。過程雖然緊湊，但也讓我更加感受到國際旅程的不可控性。這段奔波也提醒我，前往國際研討會並非單純參與課程會議，而是需要在旅途中保持彈性與冷靜，以確保能順利抵達會場。

2. 抵達美國亞特蘭大的首要任務-適應氣候、環境以及文化

抵達亞特蘭大時已經是當地時間下午 16:35，下機後最明顯的感受是氣候與台灣截然不同。當地氣溫約 6 - 18 度，早晚溫差大，必須即時增添衣物以防受寒。從亞特蘭大機場前往住宿地點 Hotel Indigo Atlanta Downtown by IHG。這間飯店位於市中心，並非研討會主場地，但距離 AMIA 的會議地點 Atlanta Marriott Marquis 十分接近，交通便利。抵達飯店時，已接近當地傍晚，旅途的疲憊感隨時間累積逐漸浮現。儘管如此，我仍利用短暫的休息時間整理行李、熟悉房間與周遭環境，並再次確認隔天 AMIA 報到的流程與選課安排。

在異地使用 Uber 也讓我重新適應美國的交通模式與文化差異，包括小費制度、駕駛風格、道路禮儀與整體生活節奏，與台灣截然不同。



從桃園國際機場出發



抵達亞特蘭大機場



前往會場報到之街道景色(有城市森林美名)



至 AMIA Annual Symposium 會場報到

(二) AMIA Annual Symposium 參與與小組討論

此次研討會的會議課程安排緊湊，除了抵達隔日至會場 Atlanta Marriott Marquis 進行報到外，每天自早上至傍晚都有多軌議程同時進行並且同一堂會議課程又包含 5-6 項小主題。面對琳瑯滿目的主題，我依照自己在院內的 AI 工作範疇（疾病預測模型、LLM 微調、資料處理、RAG、RLHF、FHIR 架構等）精心挑選最具實用性與前瞻性的課程。



Atlanta Marriott Marquis 外觀



Atlanta Marriott Marquis 內觀

1. W10：開源大型語言模型（LLM）於健康領域的應用：KDDM Working Group—Open-Source LLMs for Health Applications

W10 是一堂偏向技術底層、面向醫療資料科學者與工程師的研討課程，主要聚焦於開源大型語言模型（LLM）在醫療場域中的應用潛力。講者首先從模型架構層面解析目前較受學界與產業重視的開源 LLM，例如 LLaMA、Mistral、MedAlpaca 等，並說明它們與商業模型（如 GPT-4 或 Claude）的差異。課堂中強調，醫療領域最重視的並不是模型規模，而是模型能否「被醫院掌控、可稽核、可持續訓練」這三大核心特性。講者展示了多模態醫療模型的整合方式，例如如何將影像（Radiology Images）、生命徵象（Vitals）、結構化 EHR 資料（如 Lab Values）、以及臨床文本（Clinical Notes）整合成一個統一的知識架構，進而提升模型在臨床推論與醫療問答上的能力。另一大重點則是 RAG（檢索增強生成）在醫療 AI 中扮演的角色。講者指出，在醫療領域中，模型永遠不能只依靠「內建參數」回答問題，必須透過 RAG 結合外部來源，例如臨床指引（Guidelines）、建議用藥資訊、醫療 SOP 等，才能確保回答根據最新、可查證的醫療知識。

課程會議後半段也觸及模型部署與資料安全的核心議題。講者指出，開源模型能直接運行於醫院的封閉環境，不需將敏感資料外傳至雲端，因此更能符合法規要求，也因此成為醫療領域近年快速興起的重要原因之一。在此基礎上，我更加理解到，醫療模型的成熟並不取決於參數規模，而是是否具備可控性與可稽核性；唯有能追蹤、監測與調整模型行為，醫療機構才有足夠信任度讓其介入臨床流程。同時，檢索增強生成（RAG）被視為醫療 LLM 的必備架構。醫療模型無法僅依賴內部參數，而必須透過 RAG 引入可查證、可追溯且最新的資料以避免 hallucination，確保輸出內容符合臨床邏輯。此外，課程也強調多模態資料整合的重要性，未來模型若能同時理解影像、病歷文本、生命徵象與實驗室檢驗數據，其臨床價值將大幅提升。

整體而言，這會議讓我體認到開源大型語言模型（LLM）將是台灣醫院未來部署自家 AI 系統的最佳策略。開源模型在客製化、資安控制與臨床適配度上都有明顯優勢，更適合作為建立醫療生成式 AI 生態系的基礎。



Session 1_KDDM Working Group



Session 2_KDDM Working Group

2. W13：臨床大型語言模型的監督式微調（SFT）、低秩適應（LoRA）與人類回饋強化學習（RLHF）實作課程：A Hands-on Tutorial on Supervised Fine-Tuning and Reinforcement Learning for Clinical Applications

W13 是本次 AMIA 研討會中最貼近我現有工作內容的一堂課，課程採取完整的實作導向，從資料準備、模型微調到安全性強化的流程皆逐步示範，讓參與者能親自走過打造臨床 LLM 所需的每一項核心技術。課程的架構相當完整，講者以臨床文本為核心案例，展示如何將醫療語料整理成可用於模型學習的格式，並解釋資料品質對模型表現的影響。對於目前負責 LLM 微調與醫療資料處理的我而言，這堂課提供的操作細節與概念整合帶來極大收穫。

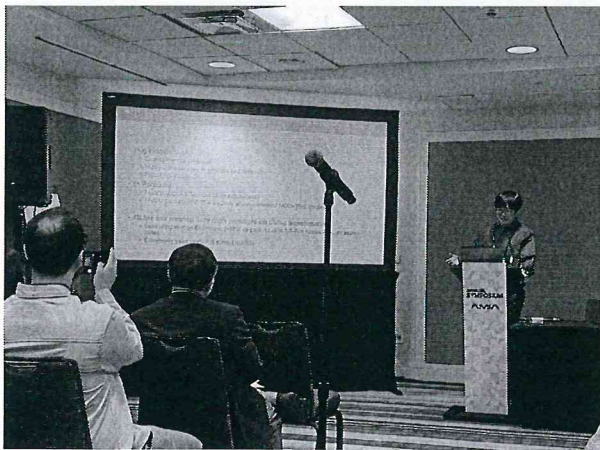
課程的第一部分聚焦於監督式微調（Supervised Fine-Tuning, SFT）。講者示範如何將原始醫療資料轉換成 instruction-based dataset，包括消除醫療縮寫的不一一致、處理格式錯誤的病歷紀錄、以及將重要臨床資訊重新組織為模型可理解的指令與回答格式。講者特別強調，在醫療領域中，SFT 的成功與否很大程度取決於資料是否具備臨床語境一致性（Clinical Context Consistency）與是否能代表實際使用情境。這讓我更加理解資料前處理的重要性，也讓我反思目前院內資料清洗方式仍有優化空間。

課程第二部分深入介紹低秩適應技術（Low-Rank Adaptation, LoRA），這是許多醫療機構能在有限硬體資源下仍然進行 LLM 客製化的關鍵方法。講者以模型架構視角說明 LoRA 如何透過「低秩矩陣」插入方式達到參數高效更新，並示範如何選擇 target modules（例如 attention 模組中的 Q/K/V matrices）以提升微調效果。透過課堂示例，我更清楚理解 LoRA 為何能同時達成高效、可控與成本親民三者兼具的優勢，並且感受到未來醫院自行訓練模型不再是不可能的事。課程的最後一部分聚焦於人類回饋強化學習（Reinforcement Learning from Human Feedback, RLHF）。講者說明 RLHF 的目的不是提升模型的語言流暢度，而是建立模型的「臨床安全性」。講者展示如何建立 reward model、如何讓醫師針對模型回答提供偏好排序（preference ranking）、以及如何透過獎勵與懲罰機制降低 hallucination。這讓我清楚認識到，SFT 能讓模型學習醫療語言的基本表達方式，而 RLHF 才是讓模型真正能在臨床回答中兼具正確性、風險敏感度與專業性的核心步驟。

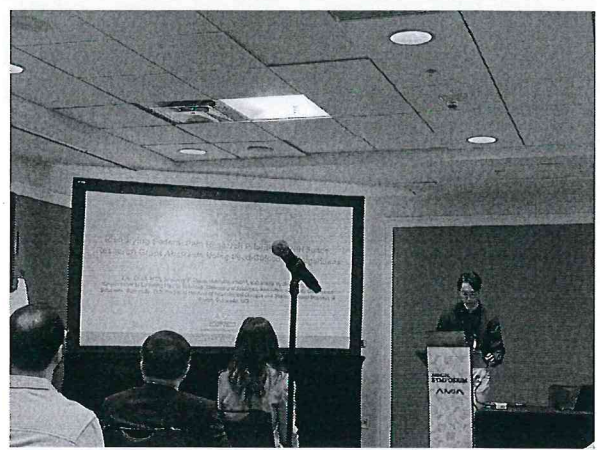
綜合本堂會議內容，我深刻體會到這三項技術並非互相獨立，而是共同構成臨床級 LLM 的基礎架構。SFT 提供語言能力、LoRA 提供可負擔的訓練途徑，而 RLHF 則提供醫療安全性保障。對我而言，W13 不僅強化了技術層面的理解，也讓我更具體地看見如何在院內建立屬於我們的 LLM 環境與微調流程，未來亦能協助更多醫療情境中的 AI 導入與落地。

讓模型符合臨床倫理與風險框架仍是不可忽視的重點。

對於本會議讓我深刻感受到國際醫療 NLP 的研究速度之快與涵蓋領域之廣。無論是資料抽取、自動摘要、多模態整合或偏誤治理，都能看到學界的努力逐漸向臨床實務靠攏。這堂課特別讓我反思，台灣醫療 AI 若要提升競爭力，不僅需要模型應用能力，更需要持續與國際研究接軌，理解 NLP 在各專科領域的演進方向。透過這場課程，我不僅學到前沿技術，也更清楚看到 NLP 在未來醫療環境中的定位與發展潛力。



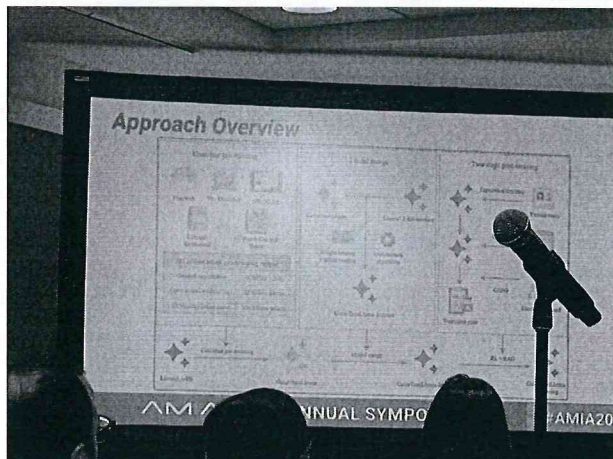
Session 1_NLP Working Group



Session 2_NLP Working Group



Session 3_NLP Working Group



Session 3_新式架構串接

4. IPS01：Epic — AI 在電子病歷（EHR）中的實際落地與臨床成效

IPS01 是本次研討會中最能讓人真切感受到「醫療 AI 已走入臨床」的一堂課。不同於以研究為主的課程，Epic 提供的內容完全以臨床實證與真實案例為核心，從醫師使用情境、護理紀錄流程到系統整合架構，都以實際運作的平台與成效數據進行說明。這種「非理論、非模擬、而是真實部署」的分享，讓我首次清晰看到美國大型醫療系統如何把生成式 AI 與 NLP 相關技術融入現場工作流程，並且確實改善照護品質。

課程首先介紹了 Ambient Documentation，也就是透過生成式 AI 協助醫師完成看診文書的服務。講者示範如何在醫師與病患對話發生的同時，由系統自動產出結構化病歷，並在醫師確認後立即匯入電子病歷（EHR）。此功能能讓醫師在



Session 1_臨床大型語言模型的 SFT、
LoRA 與 RLHF 實作課程



Session 2_臨床大型語言模型的 SFT、
LoRA 與 RLHF 實作課程

3. W33：NLP Working Group—醫療 NLP 前沿技術與研究趨勢

W33 是一場以學術研究與技術發展為主的工作坊，由多位博士生、研究者與產業專家共同分享最新的醫療自然語言處理 (Natural Language Processing, NLP) 研究成果。不同於 W10 與 W13 著重於模型應用與實作，W33 更像是醫療 NLP 的「國際研究近況總覽」，讓參與者能夠一次看到各領域的技術突破與未來趨勢。這堂課的內容範圍涵蓋臨床文本理解、病歷資訊抽取、病理與放射報告處理、生成式 AI 在醫療中的可能性，以及偏誤與安全品質等議題，形成非常完整的 NLP 研究拼圖。

課程首先聚焦於醫療命名實體辨識 (Biomedical Named Entity Recognition, Biomedical NER) 與臨床資訊抽取 (Information Extraction, IE)。多位講者展示了如何使用深度學習方法從臨床病歷 (Clinical Notes) 中自動抽取疾病名稱、檢驗值、用藥資訊等關鍵內容，並分析這些技術如何協助醫療機構建立疾病追蹤與風險預測系統。這部分讓我聯想到院內正在推動的結構化病歷專案與自動化資料提取流程，未來若能搭配成熟的 NLP 模型，醫護人員在資料紀錄與分析上將獲得更有效率的支援。

接著，講者介紹病歷摘要 (Clinical Summarization) 與多文件整合模型如何幫助醫護快速掌握病患狀況。講者示範利用生成式模型整合長篇幅的病歷內容，包括門診紀錄、急診紀錄與住院病程，產出可供醫師快速參考的摘要。雖然仍存在 hallucination 與一致性問題，但其臨床潛力相當明顯。另一方面，放射報告與病理報告的資訊抽取 (Radiology IE / Pathology IE) 也是本場重點，研究者展示如何讓模型自動從影像報告中萃取異常描述、結節大小或病理特徵，協助追蹤病灶或加速診斷工作流程。

課程也討論生成式 AI (尤其是 GPT 類模型) 在臨床推論、問答與醫療教育中的可能應用。研究者展示了 GPT 在不同臨床任務中的表現，並對比其優勢與限制，包括推論深度、臨床語境理解能力、以及模型偏誤 (Bias) 帶來的風險。值得注意的是，多位講者都提到「NLP 模型的偏誤控制與安全治理」是未來幾年醫療 NLP 推動的核心課題。這讓我意識到，即使模型能力持續提升，如何

看診後不再花費大量時間撰寫紀錄，使醫療人力得以回到真正的臨床服務上。根據簡報內容（檔案 IPS01_Epic_2025.pptx），Epic 指出此技術已成功提升病歷一致性，並降低醫師行政負擔，成效明確且持續擴大。

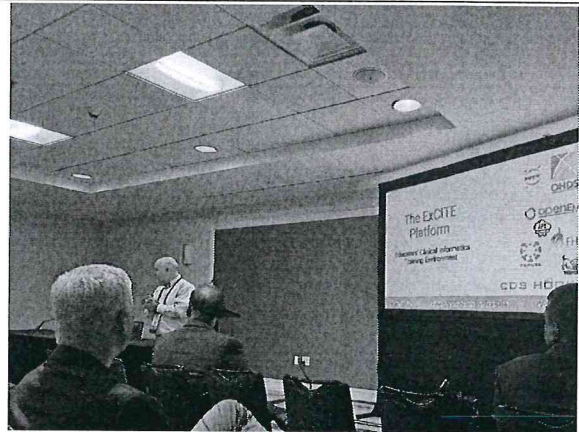
另一項令人震撼的案例是 CONCERN EWS (Early Warning Score)，一套利用 NLP 與多變量特徵分析來預測病患惡化風險的早期預警系統。Epic 分享的數據顯示，該系統能在平均兩天前預測病患狀況可能惡化，其導入後的院內成效更是非常亮眼，包括降低住院死亡率 35.6%、降低敗血症 7.5%，並協助護理團隊提早介入高風險病患。這是少數已被大型醫健康體系「證實能改善臨床結局」的 AI 系統，也再次說明資料品質、NLP 技術與臨床流程整合的重要性。

Epic 也展示了 AI 在放射科追蹤工作流程中的應用。例如利用 NLP 自動解析放射報告內容，偵測是否存在肺結節，並建立追蹤清單。系統能自動分辨結節大小、位置與變化，並提示臨床是否需要後續追蹤。這不僅降低漏追風險，也讓醫療團隊能以標準化方式管理大量影像資訊，提升整體照護安全。

綜合來說，IPS01 讓我強烈感受到 AI 不再只是「概念」或「研究原型」，而是真正在美國醫院大規模落地、並產生可量化改善的成熟技術。透過這堂課，我對未來台灣醫院導入 AI 的潛在方向有了更清楚的想像，例如：語音看診的應用可降低醫護文書負荷、EWS 可提升病患風險管理能力、NLP 可協助改善追蹤流程。這堂課讓我更深刻理解，醫院導入 AI 的關鍵在於「技術 × 流程 × 使用者體驗」的協作，而非單純追求模型效能。



SESSION 1_ Oracle Health



SESSION 2_ Oracle Health



FINAL_DISCUSS

5. IPS05：Oracle Health — AI 與資料平台整合策略（FHIR、Semantic Layer、Population Health）

IPS05 是一堂以醫療資訊架構與資料治理為核心的課程，講者來自 Oracle Health，分享他們在大型醫療資訊系統中整合 AI、資料標準與語意層（Semantic Layer）的策略。不同於偏向臨床應用的 Epic 場次，本課程更多著重於「AI 能否在醫療體系穩定運作」的根本基礎：資料標準化、資料平台架構與跨系統整合能力。對目前在醫院推動 AI 的我而言，這堂課提供了釐清整體資料策略的視角，也讓我意識到 AI 的成功並非取決於單一模型，而是整個資料生態系統的成熟程度。

課程首先介紹 FHIR(Fast Healthcare Interoperability Resources)在 Oracle Health 系統中的核心地位。講者說明，FHIR Projection 是讓異質性的 EHR 資料能自動對應至 FHIR 結構的關鍵技術，能將來自不同醫療機構、不同系統、甚至不同資料格式的內容統一轉成標準化格式。透過這種自動映射（auto-mapping），可大幅降低資料整理成本，並確保資料能被 NLP、預測模型或 Decision Support 系統順利使用。這使我重新意識到：台灣若能加速 FHIR 的普及，將能大幅提升醫療 AI 開發速度與跨院協作的的能力。

接著，課程進入 Semantic Data Platform 的介紹。講者指出，僅靠 FHIR 的格式標準化仍不足以支持 AI 的全域運作，醫院還需要建立「語意層」（Semantic Layer），將資料定義、欄位意義、生命週期與版本控制系統化管理。只有在語意層完善的情況下，AI 才能確保解讀到的資料具有一致含義。例如，血壓、血氧、心率等資料若未經語意標準化處理，模型可能會因來源差異而出現誤判。這讓我理解到語意資料管理不僅是一項技術任務，更是確保模型安全與可靠性的基礎工程。

課程的後半段聚焦於 Population Health（族群健康）的資料治理策略。講者展示 Oracle Health 如何利用 AI 分析大型族群資料、辨識高風險族群與預測公共衛生趨勢。這些功能必須依賴跨機構資料整合與大型資料平台，而非單一醫院資料來源。同時，講者也提到 AI Agents 的應用，例如自動提醒異常檢驗值、主動偵測疾病群聚、協助進行慢性病追蹤等，象徵醫療 AI 未來不僅是工具，更將成為醫療工作流程中的主動協作者（Proactive AI Collaborator）。

綜合本堂課內容，我深刻感受到 AI 在醫療中的成功落地，取決於資料基礎建設是否完善。AI 模型本身並非最複雜的部分，真正困難的是資料跨系統整合、標準化、語意管理與治理架構的建立。Oracle Health 的分享讓我認識到，一間醫院若要真正發展 AI 不僅需要模型能力，更需要資料平台能力。台灣醫療體系若能在 FHIR、語意層（Semantic Layer）與跨院資料整合上加強投入，未來將能更有效推動 AI 在臨床的價值，也能縮短與國際醫療資訊系統的差距。



6. S61: Using LLMs to Interpret Arterial Blood Gases – Comparison of a Math Scratchpad with Different Prompting Methods

S61 會議課程是我此次在 AMIA 中印象特別深刻的一堂技術導向課程，由 Dana-Farber Cancer Institute 與 Amazon Web Services (AWS) 共同報告。課程主題聚焦於利用大型語言模型 (LLM) 解讀動脈血氣 (Arterial Blood Gas, ABG)，並比較不同提示方法 (prompting methods) 以及是否加入「數學推論草稿 (math scratchpad)」對模型準確度的影響。由於 ABG 是急性照護中極為關鍵的檢查項目，判讀上的失誤可能直接影響臨床決策，因此這堂課不僅是技術展示，也呈現了 AI 在高風險臨床任務中需要面對的限制與挑戰。

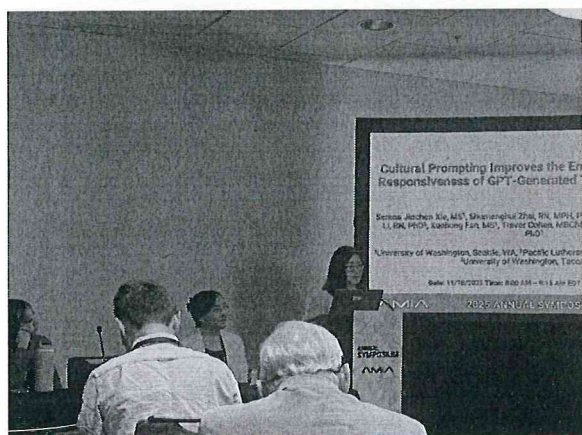
講者首先介紹 ABG 在臨床中的重要性，說明醫師與護理人員在高工作量環境下，常需在短時間內完成複雜的計算與推論，例如代謝性酸中毒、呼吸性酸中毒的判別、酸鹼平衡的補償機制等。因此團隊嘗試以 LLM 協助完成這類高認知負荷的任務。他們比較三種讓 LLM 執行 ABG 計算與解讀的方法：第一種是「Zero-shot prompting」，直接要求模型解讀 ABG 結果；第二種是加入知識檢索 (RAG) 與 prompt engineering 的提示鏈提升模型可靠性；第三種則是結合「數學 scratchpad」，讓模型在回答前先寫出推論步驟與公式計算，形成可追蹤的數學推理流程。

結果相當具有指標性：Zero-shot 方法僅有 48% 準確度，加入 RAG 與提示鏈後上升到 78%，而納入數學 scratchpad 的版本則提升至 86% 的準確度。這項發現強調一個重要觀點——臨床 AI 的關鍵不在於模型多大、多強，而在於是否能讓模型的推論過程具備「可稽核性」(auditability) 與「可追蹤性」

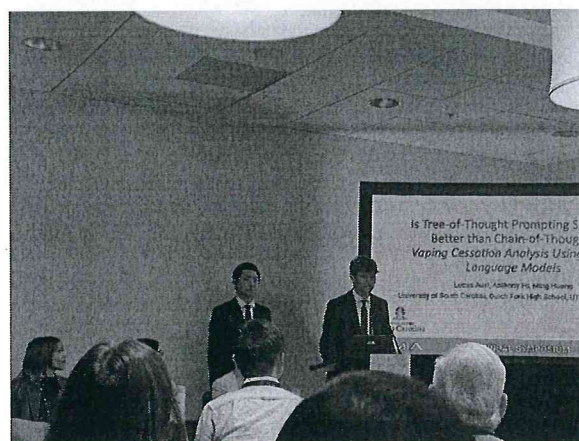
(traceability)。尤其是在 ABG 這類數值敏感的任務，LLM 若無明確計算步驟，往往容易出現 hallucination 或錯誤推論。Scratchpad 方法的價值在於強迫模型逐步思考，並讓每一次中間運算能被檢查、驗證，降低臨床應用上的風險。在課程的後半段，講者分享其在臨床整合方面的心得。他們指出，即使 LLM 的準確度逐漸提升，若無法清楚呈現推論邏輯，臨床人員仍會對模型缺乏信任感。數學 scratchpad 正是一種能提高模型透明度的方式，讓醫護人員能看到模型是如何計算、如何得出診斷建議，也讓 AI 能真正融入臨床工作流程之中。

對我而言，這堂課帶來的啟發非常實際。身為醫院 AI 工程師，我常接觸疾病預測、臨床資料處理與 LLM 微調等任務，而這堂課讓我更加確信「可控性」

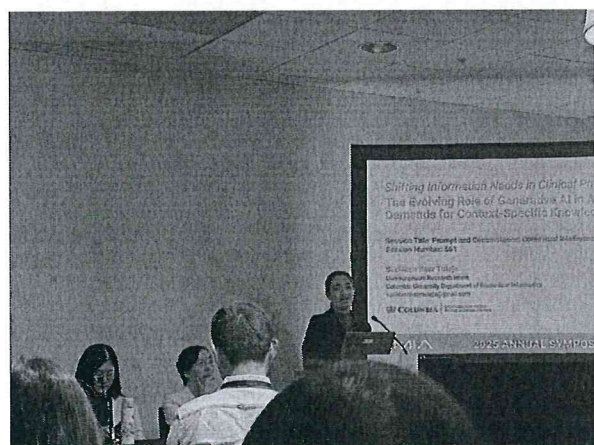
與「可解釋性」才是醫療 LLM 成功落地的核心。與國外相比，我們現階段的流程多半仍停留在模型效能 (accuracy/AUROC) 的比較，但 AMIA 的討論已經走到模型治理、推論透明度、提示鏈設計與多代理架構等更高層次的方向。總結來說，S61 會議課程讓我看到 LLM 在臨床應用上的極大潛力，也讓我意識到 AI 若要真正協助醫療決策，必須超越一般語言生成任務，而進入「可解釋、可控、可驗證」的成熟模式。這堂課提供的思考將對我未來在本院推動 LLM 應用與模型落地策略產生深遠的影響。



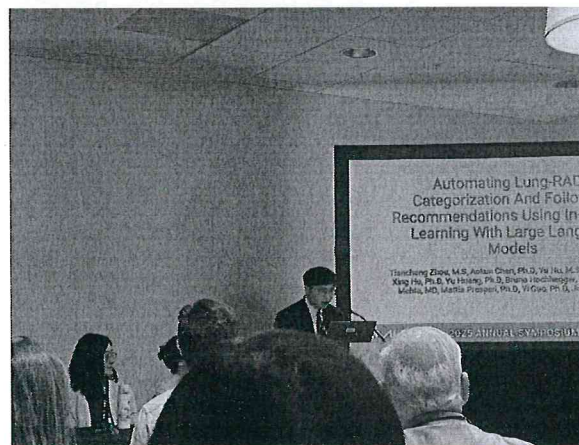
SESSION 1_Prompt and Circumstance:
Contextual Intelligence in the Age of LLMs



SESSION 2_Prompt and Circumstance:
Contextual Intelligence in the Age of LLMs



SESSION 3_Prompt and Circumstance:
Contextual Intelligence in the Age of LLMs



SESSION 4_Prompt and Circumstance:
Contextual Intelligence in the Age of LLMs

7. AMIA 研討會後的亞特蘭大文化參訪

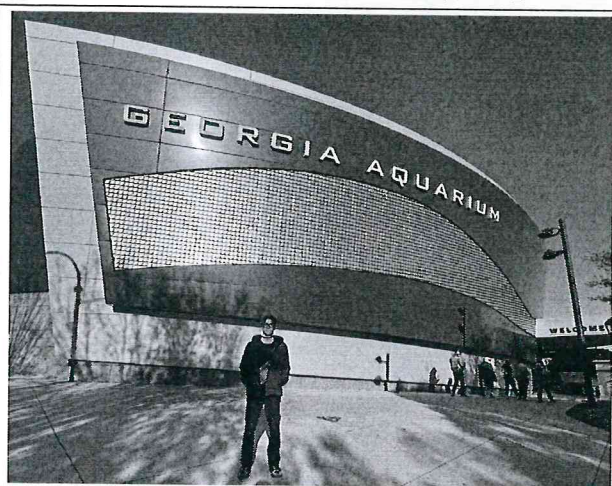
在 AMIA 研討會結束後，我利用空檔前往亞特蘭大數個具有代表性的景點，作為這趟學術行程的延伸體驗。首先造訪的是 可口可樂世界 (World of Coca-Cola)，透過品牌歷史展示及沉浸式展演，我得以看見一個企業如何將產品文化化、全球化。這樣的呈現方式也讓我思考，醫療 AI 若要真正被廣泛採用，其技術溝通與使用者教育也同樣重要。

接著，我前往市中心的 百年奧運公園 (Centennial Olympic Park)。夜間的五環雕塑與周遭燈光營造出具有象徵性的氛圍，使我能在相對安靜的環境中重新整理

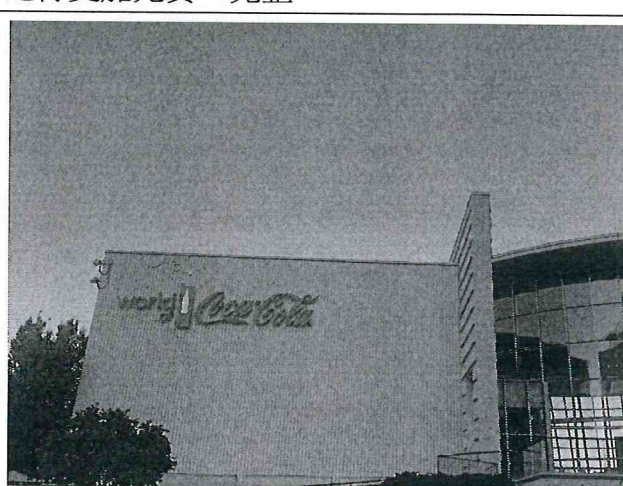
研討會期間所吸收的資訊，也自然反思 AI 與人、科技與臨床間的「連結」這個核心議題。

最後參觀的 喬治亞水族館 (Georgia Aquarium) 以其規模與生態展示著稱。館內完整的動線設計與視覺呈現，使我深刻感受到良好設計如何提升整體體驗。這也讓我聯想到醫療資訊系統中，使用者介面 (UI/UX) 與流程設計與 AI 是否能被真正採用有密切關聯。

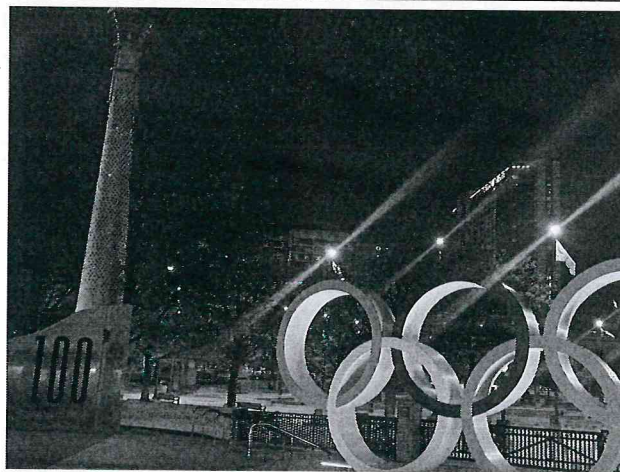
整體來說，這些行程雖然屬於休閒性質，但卻讓我得以在不同情境下重新回望研討會的內容，也讓整趟 AMIA 之行更加充實、完整。



喬治亞水族館 (Georgia Aquarium)



可口可樂世界 (World of Coca-Cola)



百年奧運公園 (Centennial Olympic Park)

(三) 海報展區-發現研究部臺中榮民總醫院海報

1. 在研討會的後半段，我特地保留時間參加 AMIA 的 Poster Session。這場展示安排在晚上 17:30 至 19:00，是許多研究者在一天的課程結束後仍會留在會場交流的重要時段。來自世界各地的研究者站在自己的海報前與參觀者互動，整個空間充滿討論聲與學術交流的氛圍。

在沿著展板逐一瀏覽海報的過程中，我意外地注意到其中一張熟悉的署名——來自 Taichung Veterans General Hospital, Taiwan。在國際醫療資訊年會中遇到來自自己醫院的研究成果，讓我感到格外驕傲，也使整個海報參訪變得更有連結感。這張海報編號為 P16，題為：「The Role of Digital Biomarkers for Insulin Resistance in

Predicting Fatty Liver and Type 2 Diabetes: Evidence from the Taiwan Biobank」。

能在 AMIA 這樣的國際會議現場看到臺中榮總的名字，讓我倍感鼓舞，期待未來能有機會親自投稿 Poster 或 Oral Presentation，把自己在醫院的 AI 研究帶到國際舞台上展示。



三、心得

此次參加 AMIA 的經驗，讓我對國際醫療資訊學的發展速度與深度有了更具體的感受，也讓我重新檢視自己的能力、視野與未來方向。其中最讓我強烈衝擊的，是多堂與大型語言模型（LLM）相關課程中所呈現的「多 Agent 架構」思維。以往我在院內較多接觸的是單模型或單流程的 AI 設計，而在 AMIA 中，許多研究者展示的是由多個 Agent 共同組成、具備前後串聯能力的複雜系統。最吸引我的地方是，這些多 Agent 並不僅限於單純的專業任務，例如「醫療問答 Agent」、「臨床推論 Agent」，還包含前處理層級的 Agent，如「資料確認 Agent」、「資訊檢索 Agent」、甚至「資料品質檢查 Agent」。這種帶有自動化前置流程的架構，讓我深刻感受到：未來的醫療 AI 不再只是單一模型的表現，而是整體系統協作的成熟度。這樣的觀念對我而言，是一次思維上的躍遷，也是本次 AMIA 帶給我最重要的衝擊之一。

在展區的交流中，我也遇到一位參與 AMIA 的 OpenAI 與會者。她在與多位研究者討論時，對一個問題表現出濃厚興趣——「GPT-3 在醫療領域的專業理解度，究竟是否足以被視為醫療模型？」這樣的提問背後透露出國際研究者對模型可靠性的高度關注。她解釋，許多大型模型雖然在語言流暢度上已達到相當成熟的水準，但不同版本的模型往往由不同工程團隊訓練，其訓練語料、偏誤、醫療知識深度也有所不同。她也提出疑問：學術界是否真的理解模型是如何學習醫療知識？是否足以信任模型背後的推論過程？這段對話讓我深刻感受到，美國研究界在模型安全性與可稽核性的討論遠比我原先預期得更前面，而這些議題也正是台灣在推動臨床 AI 時會遇到的核心挑戰。我從這段交流中獲得許多反思，也更加理解資訊透明度與模型治理的重要性。

從工作內容的角度來看，本次 AMIA 與我現職的差距也非常明顯。身為醫院的 AI 工程師，我習慣處理模型建置、資料前處理與臨床協作等任務，但在國外的報告與課程中，

我看到的是更細緻且完整的架構流程，包括資料語意層 (Semantic Layer)、FHIR-based data modeling、跨平台模型監控系統、強化模型安全性的機制、甚至在資料流程的不同節點設置專門 Agent 進行驗證。這些精緻化的架構讓模型輸出更可靠、更能避免臨床風險，而這些正是目前台灣醫療體系仍然需要加強的部分。這也是我希望未來能強化的能力——更完整的架構設計視野，以及更深層的資料治理與系統整合能力。

這次研討會帶給我最大的反思，是必須保持不斷學習的態度。在國際舞台上，每一個細節都值得借鏡，不論是研究方法、資料流程、系統設計，甚至是與研究者互動的方式，都展現出與台灣不同的文化與思維深度。這種強烈的落差並不是負面的感受，而是一種提醒：在醫療 AI 的浪潮中，我們永遠都需要把自己放在學習者的位置，不斷吸收新的知識並檢視自身的不足。

最後，參加 AMIA 也讓我重新思考未來的工作方向。我希望自己不僅具備實作能力，更能理解國外在研究設計、思維深度與語意資料層等領域的操作邏輯。我也期待自己未來能更深入參與多 Agent 架構、RAG、LLM 安全性與醫療資料平台的建置，並逐步累積足夠的能力與經驗，站在國際舞台上，展示屬於臺灣、屬於本院的 AI 成果。

四、建議事項

(一) 強化院內資料治理與語意標準化概念的導入

此次 AMIA 多次強調 FHIR、Semantic Layer 與資料標準化的重要性。建議院內可在既有資料流程上，逐步導入資料一致性與欄位語意對齊的概念，為未來 AI 導入奠定基礎。

改進作法：可先從「常用欄位」著手（如生命徵象、檢驗、用藥），逐步建立欄位定義表或簡易語意說明，讓團隊有一致的資料理解方式。

(二) 規劃 LLM (SFT/LoRA/RAG) 的小規模驗證與實作

多堂課提到臨床 LLM 的「可控性」與「可稽核性」。建議院內未來可選定部分臨床場景進行小規模 PoC (Proof of Concept)，逐步測試 LLM 微調與檢索增強生成的可行性。

改進作法：可先選擇單一、固定格式的任務（如病歷摘要、固定規則的 QA）進行模型微調與 RAG 驗證。

(三) 逐步導入模型解釋性 (Explainability) 與多代理 (Multi-Agent) 流程概念

國外醫療 AI 的重點正從「模型效能」轉向「模型透明度與流程安全性」。建議未來 AI 開發可逐步加入基本的推論檢查流程（如推理步驟、輸出驗證 Agent），提升臨床信任度。

改進作法：從現有模型開始加入簡單的「結果檢查」或「提示鏈」流程，而非一次導入完整 multi-agent 系統。

(四) 鼓勵持續參與國際研討會以提升團隊視野

此次 AMIA 讓我深刻感受到國際醫療 AI 與資料科學的最新發展。建議院內可持續鼓勵同仁參與國際研討會，以掌握最新趨勢並累積國際合作可能性。

改進作法：可於年度教育訓練規劃中加入「國際會議參與報告分享」機制，由參與者回院分享學習成果。

五、 附錄

(一) AMIA Annual Symposium 研討會地點：

<https://amia.org/education-events/amia-2025-annual-symposium/venue>

(二) AMIA Annual Symposium 會議資訊：

<https://amia.org/education-events/amia-2025-annual-symposium>