



星球永續健康線上直播

可解釋 AI (XAI) 醫療應用

XAI 臨床決策應用治理

2026 年 1 月 14 日

人工智能由輔助工具逐步轉變為影響臨床判斷的重要協作角色，其透明性、可理解性與可信任程度，已成為臨床實務、倫理治理與公共信任的核心議題。本週我們將聚焦於可解釋人工智能（XAI）在臨床決策輔助中的治理架構與實際應用，並以臨床實例說明 XAI 導入臨床後，對醫療照護人員決策應用轉變。

星球健康新知

委內瑞拉變局牽動石油供應全球經濟：「權重失衡」

2026 年初，美國特種部隊突襲加拉加斯並拘捕委內瑞拉總統馬杜羅，成為冷戰後罕見由大國在未經多邊授權下，對他國在位元首採取強制行動的案例，迅速引發國際震盪。此行動其政治、法律與經濟層面的震盪效應在短時間內即向區域與全球評議。行動發生前數小時，馬杜羅仍於公開場合會晤中國高層外交代表，重申中國-委內瑞拉戰略夥伴關係與雙方政治互信，使事件發生後北京所處的被動與尷尬位置更為明顯。過去二十餘年間，中國對委內瑞拉提供約 1,060 億美元的官方貸款與援助，資金長期集中於能源與基礎建設部門，並在美國制裁體系下，成為少數仍與委內瑞拉維持實質經濟往來的主要國家之一。雖然委內瑞拉原油僅占中國整體進口量約 5%，卻支撐了委國約八成的對外原油需求，顯示雙方在政治與經濟層面的高度相互依存。然而，當美國實際動用軍事力量並成功帶走馬杜羅後，北京的回應僅止於強烈譴責與主權論述，並未伴隨任何具體反制或安全支援行動，使中國長期以「多極化」與「反霸權」為核心的外交敘事，在拉丁美洲首次遭遇極為直接的現實檢驗。多項評論因而指出，這一結果對中國其他所謂全天候夥伴形成警訊，在國際法與多邊體系層面，此次行動使聯合國安理會政治張力急速升高。多數與會國家雖承認馬杜羅政權長期存在民主正當性不足、壓制反對派及涉入跨國犯罪的爭議，仍清楚區分「政權性質」與「使用武力正當性」兩個層次。法國與丹



麥代表強調，美國行動違反和平解決爭端及禁止使用武力的基本原則，削弱《聯合國憲章》所奠定的國際秩序根基。西班牙亦表態不會承認任何違反國際法、可能將區域推向高度不確定與對抗風險的介入行為。相對之下，英國與希臘刻意避免明確定性美國行動是否構成違法，僅籲統呼籲和平過渡與政治對話，反映歐洲國家在維護國際法規範與顧及對美安全依賴之間的結構性矛盾。聯合國秘書長則透過正式聲明警告，此一事件可能成為高度危險的先例，進一步加劇委內瑞拉長期累積的政治、社會與經濟危機，並對區域穩定與國際關係產生外溢衝擊，強調即便面對嚴重治理失靈，處理途徑仍應以法治與多邊機制作為依歸。在美國國內政治與經濟層面，白宮刻意將此次行動與能源戰略緊密連結。川普 公開宣示，美國將暫時接管委內瑞拉，核心目標之一在於重建其石油產業，並協助美國能源企業追回因過去國有化政策而累積、估計高達 600 億美元的資產損失。委內瑞拉擁有約 3,000 億桶可探明石油儲量，居全球之冠，但多年來受制裁、資金匱乏、技術流失與管理失序影響，產量已滑落至每日約 100 萬桶，僅為歷史高峰期的三分之一左右。美方構想認為，若解除制裁並引入美國資本與技術，產量可逐步回升，甚至在中期內對全球能源市場結構產生影響。不過，多數能源分析對此持高度保留態度，指出短期內美國的封鎖與油輪查扣反而使委國出口驟降，稀釋劑短缺亦迫使產量進一步下修；中期仍需面對至少 1,100 億美元以上的資本支出需求、人力大量流失與國營石油公司體制重整等重大挑戰；長期而言，在全球原油供給相對充裕、價格承壓的情境下，高成本重油投資的商業吸引力顯著下降。儘管如此，金融市場已對潛在的地緣政治重組與資源再分配迅速作出反應，美國大型能源企業股價在消息公布後明顯上揚，顯示投資人將此視為一項具長期選擇權性質的政治經濟賭注。馬杜羅被拘事件同時呈現出三項深層現實。首先，國際法與多邊規範在大國直接行動面前的實際約束力正遭遇嚴峻考驗，盟友間對原則的解讀與實踐已出現明顯裂縫。其次，中國在西半球長期累積的經濟存在與外交敘事，尚不足以在關鍵時刻轉化為可制衡美國的實質能力。第三，委內瑞拉的政治過渡與能源重建已不再只是單一國家的內政問題，而是將在未來多年持續牽動大國競逐、區域穩定與全球能源秩序的高度複合型議題。



格陵蘭主權爭議北約表達堅定立場：「權力失序」

川普政府將獲取格陵蘭島視為外交政策與國家安全的首要任務。白宮證實川普及其顧問正在討論多項取得格陵蘭的方案，且並未排除使用軍事力量的可能性。美方的核心動機在於戰略防禦與資源開發，川普指出格陵蘭周邊佈滿俄羅斯與中國船隻，對美構成威脅；同時，氣候變遷導致冰層融化，使得該島豐富的稀土、鈾及鐵礦等礦產更易於開發。此議題的急迫性在美國對委內瑞拉採取軍事行動後迅速升溫。在美軍特種部隊於加拉加斯逮捕委內瑞拉總統馬杜洛及其妻子後，白宮高級顧問史蒂芬·米勒的妻子隨即發布覆蓋美國國旗顏色的格陵蘭地圖，暗示格陵蘭是下一個目標。對此，格陵蘭與丹麥政府強烈駁斥，並強調格陵蘭並非待價而沽的商品。丹麥首相佛瑞德里克森則發出嚴厲警告，指出格陵蘭是北約成員國丹麥的領土，美國若採取軍事行動將意味著「北約的終結」。國際社會也展現團結立場，英國、法國、德國、義大利等多國領導人發表聯合聲明，重申格陵蘭的未來僅能由當地人民與丹麥決定，並呼籲維護主權與領土完整的國際法原則。英國首相施塔默亦明確表態支持丹麥，對美方的併吞意圖表達反對。儘管國務卿魯比歐曾試圖緩和氣氛，表示正探討購買或建立自由聯合協定等非戰爭手段，但美方不排除武力的立場已使北約盟友間的信賴關係面臨空前挑戰。

朝鮮半島重啟多邊外交：「審慎求衡」

區域地緣政治高度動盪、強權競逐加劇的背景下，韓國新政府試圖重新調整其對中政策定位，呈現出在安全依賴美國與經濟高度連結中國之間進行審慎平衡的外交取向。韓國總統李在明於就任後不久即選擇優先訪問北京，並與中國國家主席習近平舉行高峰會談，此一安排在時序與象徵層面皆具高度政治意涵。雙方在會談中以修復關係、恢復互信與推動經濟合作作為主要共識基調，並多次強調兩國在維持區域穩定、促進共同繁榮方面具有重疊利益。中國方面強調中韓關係的戰略重要性，並以歷史記憶與共同反對日本軍國主義的敘事作為情感連結；韓國方面則在重申尊重「一中原則」的同時，刻意避免在台海議題與中菲南海衝突上作出具體承諾，顯示其試圖降低安全議題的正面衝突，將外交重心暫時放在經貿與功能性合作層面。在安全層次上北韓問題成為雙方少數被明



確納入討論的議題。李在明強調，中國在北韓問題上具有不可替代的影響力，並呼籲北京協助促使平壤重返對話軌道。中國則回應以維護區域和平與促進對話原則性表述，未直接觸及去核化細節，但雙方同意加強溝通機制並推動年度領導人會晤。在此脈絡下，韓國政府同步展現對北韓的政策轉向，其統一部長透過公開新年談話釋出「任何時間、任何地點皆可對話」的訊號，明確表達結束敵對關係、優先推動和平共存的政策意志，並強調不追求吸收式統一，而是尊重現存體制、反對任何形式的敵對與武力行為。此一立場被定位為在對話管道中斷情況下，透過公開訊息向北韓與國際社會傳遞政策轉向的嘗試。在經濟與功能合作層面，韓中雙方於會談期間簽署多項合作備忘錄，涵蓋科技、貿易與環境等領域，並凸顯雙方在半導體供應鏈、稀土資源、觀光與文化產業上的高度相互依賴。與此同時，韓國政府亦構想以中國作為關鍵中介者，推動更具突破性的南北韓合作方案，包括經由北韓連結首爾與北京的跨境鐵路構想，以及以中國遊客與海外僑民為起點、分階段推進的「和平觀光」模式。然而，這些構想在現行聯合國與美國制裁架構下多被視為高度受限，且在韓國國內亦引發對其可行性與合規風險的質疑，顯示政策理想與現實條件之間仍存在顯著落差。

巴黎會議英法承擔俄烏停火安全：「言重行難」

巴黎近日舉行「志願者聯盟」高峰會，在俄烏戰爭邁入第四年的關鍵時刻，盟國試圖為烏克蘭打造一套可長期運作的戰後安全保障架構。此志願者聯盟高峰會中，烏克蘭及其主要盟友試圖就潛在停火後的安全架構形成更具體的共識，但整體談判氣氛仍充滿不確定性。雖然英法簽署共同安全協議且法國總統馬克宏在會前曾表達樂觀，強調與會各國將提出具體承諾，以確保烏克蘭在任何和平協議之後仍具備有效嚇阻俄羅斯再度侵略的能力，並規劃包括停火監測、強化烏克蘭武裝部隊、多國部隊部署、面對再次侵略的即時支援承諾，以及長期防衛合作等多項核心機制，然而這些努力同時受到美國政策重心轉移的影響，美方近期在委內瑞拉軍事行動與格陵蘭主權企圖引發歐洲盟友高度關切，並使跨大西洋關係面臨微妙張力。儘管歐洲國家普遍認知，烏克蘭未來的安全保障仍高度仰賴美國的軍事與戰略支撐，但華府是否能在多重地緣政治議題下持續投入，仍



讓部分與會國家心存疑慮。烏克蘭方面則一再強調，美國實質參與為安全承諾重要關鍵，避免停火淪為俄羅斯重整軍力緩衝期。俄羅斯立場至今未見鬆動，仍堅持在達成全面政治解決之前不接受停火，並明確排除任何北約部隊進駐烏克蘭領土的可能性。另一方面，美國在此次峰會上首度公開支持建立廣泛且具法律效力的安全保證框架，並表示願意在停火監測與後續安全機制中扮演「後盾」角色，包括透過無人機、感測器與衛星等手段參與監督，但相關承諾的具體形式與法律強度仍待進一步釐清。巴黎會談展現出歐洲與美國在支持烏克蘭安全上的高度外交動能，但在美國全球戰略優先順序、歐洲內部政治限制，以及俄羅斯強硬立場交織之下，將這些政治宣示轉化為可執行且持久的安全安排，仍是一項高度複雜且充滿變數的挑戰。

2026 自動駕駛車載運算新篇章：「實體轉向」

國際消費性電子展(CES) 拉斯維加斯展會上，NVIDIA 推出創新 AI 運算架構 Vera Rubin 平台，並由執行長 Jensen Huang 發表名為 Alpamayo 的自動駕駛 AI 平台，展現 Nvidia 積極推動進階大規模 AI 架構以及加速將人工智慧從運算晶片推進至可實際部署於車輛與機器人的實體 AI 系統兩方面布局。該平台核心特色在於具備情境推理與決策解釋能力，車輛可在真實且高度不確定的環境中，自行判斷如號誌失效、交通混亂等複雜狀況的應對方式，並清楚說明其行為邏輯。Alpamayo 採開放原始碼設計，研究人員與車廠可重新訓練模型，透過車載電腦整合攝影機與多元感測器資料，將複雜情境拆解並生成對應解決方案。NVIDIA 亦延續與 Mercedes-Benz 的合作，推動具備高速公路免持駕駛及城市道路導航能力的車輛發展，預計首批搭載該技術的車款將率先於美國上市，隨後進入歐洲與亞洲市場。在產業層面，今年 CES 顯示汽車發展焦點已由單一動力形式轉向可即時創造價值的智慧系統，先進駕駛輔助、安全預測與持續軟體更新被視為回應消費者對安全、可靠與使用便利需求的關鍵手段，也有助於車廠在經濟壓力與電動車採用進程不均的情況下維持競爭力。Alpamayo 的推出被認為可能對 Tesla 等既有自動駕駛布局形成新的競合關係，同時也呼應產業顧問與科技企業對生成式 AI 加速進入車輛系統的觀察。隨著 NVIDIA 亦展示應用於機器人並與 Siemens 合作的相關模型，



以及共乘平台 Uber、Lyft 重新評估無人駕駛計畫並與 Waymo 等業者接軌，整體趨勢顯示，未來車輛價值正逐步由動力與硬體轉向智慧、學習與持續進化的能力，自動駕駛與汽車 AI 再度成為全球交通產業的重要發展方向。

生成式 AI 促發製造業轉型：「人機並行」

1980 年代通用汽車嘗試打造全自動無人工廠的失敗經驗為起點，說明即使工業自動化在硬體與軟體層面已較過去大幅進步，距離完全取代人力的理想狀態仍有落差。近年全球工業機器人數量持續增加，但平均配置密度仍不高，疫情期間的自動化投資高峰在製造業景氣降溫後趨於平緩，相關設備與供應商的市場表現亦顯得保守。然而，多數分析認為 2026 年可能成為新一波成長的轉折點，背後因素包括利率環境改善、製造業回流政策，以及人口老化導致的勞動力短缺。更關鍵的變化來自軟體與人工智慧的進展，使機器人得以透過重新編程執行多樣任務，數位分身與生成式 AI 亦有助縮小模擬與實際操作之間的落差，提升自動化系統的彈性與效率。在此趨勢下，工廠的角色與型態可能隨之轉變，由過去高度集中、規模龐大的產線，逐步走向更分散、更靈活、由人機協作與 AI 輔助決策所構成的生產體系，其影響程度或將超出早期提出「未來工廠」概念。

學術諮詢公共價值與挑戰：「用其所長」

傳統上大學被視為承擔研究、教學與社會影響三大核心任務的機構。然而，目前的「社會影響」多被窄化為專利授權或新創公司等商業化成果，使得具備高度擴展性的「學術顧問工作」長期受到忽視。在現行體系中，顧問工作常被視為瑣碎的行政事務，缺乏統一的衡量標準與績效肯定。在高等教育面臨財務壓力與社會需求脫節的背景下，重新檢視並制度化顧問工作顯得至關重要，因為它不僅能拓展收入來源，更能將研究成果轉化為具體的社會效益。從效益層面來看，學術顧問工作能為產官學三方帶來共贏。對大學而言，這類活動在財務吃緊時提供了彈性資金，並能促進產學互動，進而孕育聯合研究與新資金來源；對學者個人而言，則能精進專業、拓展人脈、獲得經濟回報，並提升實務視角與職業流動力。而在社會層面，顧問機制能加速公共研究回應能源、健康等國家優先議題，並將實務經驗回饋至教學內容，縮短學生的學用落差。現狀顯示學術顧問



的參與度與合約數量正持續下降，且落後於民間顧問產業。這主要歸因於各校規範的不一致（如時間上限與抽成比例之差異），以及行政支援（如法律審核、保險等）品質不一，迫使學者轉向非正式的校外活動，增加了個人與機構的法律風險。若要改變此現狀，必須推動制度標準化與服務專業化，並建立清楚的誘因，以降低參與門檻並建立社會信賴。大學應將顧問工作提升至「策略性任務」的高度，並透過明確的政策與專業基礎建設，回應產業對效率與可靠性的期待。政府與資助機構也應將顧問自由度納入支持條件，使顧問工作成為學術貢獻的正當形式。若缺乏實質改革，大學在顧問領域的空間將被私人體系所占據，其在財務與社會影響力上的角色亦將隨之弱化。

同儕審查數位治理風險：「信任為本」

2025 年底人工智慧領域大型國際會議 ICLR 的同儕審查平台 OpenReview 發生軟體漏洞，導致約 10,000 篇論文所涉及的作者與匿名審查者身分資料遭揭露並外流。事件已造成部分審查者遭受騷擾、恐嚇與不當施壓，並破壞原本依賴匿名機制維持公平性的雙盲審查制度。雖然漏洞已迅速修補，仍被視為同儕審查史上規模極大的資料外洩事件之一。此事件同時暴露 AI 研究領域投稿量急遽成長所帶來的結構性壓力，包括審查人力不足、審查品質下滑，以及審查意見疑似由大型語言模型生成等問題。外洩資料使學界得以檢視審查制度的脆弱性，但也引發對研究倫理、平台治理與學術信任的廣泛關切，促使研究社群重新思考現行同儕審查模式在高速擴張環境下的可持續性。

線上研究轉型時刻：「真假難分」

人工智慧技術大型語言模型的快速進展，已對線上社會科學研究的資料可信度造成嚴重衝擊。過去仰賴線上問卷與實驗蒐集大規模人類行為資料的研究模式，正因高度擬人的人工智慧代理能成功規避既有辨識機制而面臨根本性挑戰。實證研究顯示，這類模型不僅可通過專為辨識非人類設計的測試題，亦能模擬細緻的人類操作行為與符合特定社會角色的人口學回應，使研究者難以區分資料來源是否為真實人類。此一發展同時影響惡意造假行為與一般受試者的行為預期，進一步加劇資料偏誤風險。儘管研究平台與學者持續投入偵測與防堵技術的更新，但技術演進速度極快，使線上研究方法陷入長期



對抗狀態。部分研究者因而對未來是否仍能以低成本方式取得可信的人類行為資料表達高度保留，並呼籲重新審視線上資料蒐集在社會科學研究中的適用範圍與角色。

XAI 臨床決策輔助治理

在《黑色止血鉗》中，高階醫師代表積極引入創新科技的臨床領導者，而渡海醫師則是以精湛手術技巧見長、對人體與心臟結構與功能具有深厚經驗與理解外科職人。兩人於東城大學合作，結合創新科技與高度專業的臨床技術，成功完成原本被視為不可能的治療任務。病患小春因無法輸血且具有凝血異常體質，使心臟手術幾乎無法進行，然而其病情又迫切需要瓣膜置換。在此極端受限的條件下，透過導管技術結合瓣膜置換模組，最終成功完成治療，展現病患導向治療如何推動創新科技實際導入臨床。此一突破性成果隨後轉化為學術論文發表，不僅對臨床醫療具有實質貢獻，也大幅提升西崎教授相關研究的學術影響力與問鼎外科學會理事長競爭實力。但在完成瓣膜置換後病患小春出現感染併發症。手術過程中以心導管穿越未完全閉合的結構進入左心，並於術後以封堵器關閉通道，導管與封堵器本身成為體內異物，引發感染性心內膜炎。感染性心內膜炎的治療關鍵在於徹底移除感染源，單純內科藥物治療無法根治，意味著病患必須再次接受手術。然而小春因無法輸血的特殊體質，且再度接受心臟手術風險極高，治療因此陷入困境。渡海醫師僅能先以藥物暫時控制病情，但此策略無法長期維持，仍需最終治療方案。在此背景下，高階醫師提出以智慧手術機器人輔助的微創方式作為替代方案。該技術結合醫師的精細操作與電腦運算控制，僅需少量切口即可進入心臟，精準移除感染病灶。此一作法不僅降低開刀風險，也避免大量出血與輸血需求，為高風險病患提供兼顧安全與治療成效的可行選項。為了讓病患小春能接受先進治療，高階醫師促成智慧機器手臂系統引入東城大學，作為突破既有治療限制的手段。西崎教授長期致力於引進創新醫療科技，其核心理念在於透過科技，降低高難度手術對個別醫師經驗的高度依賴，使更多醫師能在安全條件下完成複雜治療。該智慧機器手臂結合即時運算與人機協作機制，具備一定程度的人工智慧輔助，類似自動駕駛系統中由運算核心支撐的輔助決策能力。日本醫療體系亦將此類創新科技視為縮小外科醫師技術落差的重要方向，並由政府



層級積極推動導入。執行手術主刀醫師松岡，雖對智慧機器手臂的精準度與自身操作能力高度自信，卻忽略該協作術式從未應用於兒童病患的事實。面對小春特殊體質與高風險條件，高階醫師曾提出應進行更充分的模擬與演練，以確保人機協作在全新情境下的安全性，但相關建議未被採納，為後續風險埋下伏筆。手術當日松岡醫師與智慧機器手臂協作成功進入心臟並抵達感染病灶。然而由於病患小春為兒童體型，在開始移除感染組織的過程中，機器手臂在有限空間內無法如成人手術時展開，手臂彼此擠壓卡死而無法移動。此一物理性限制並非演算法或智慧判斷所能克服，即使具備人工智慧與人機協作機制，在空間受限的情境下仍無法完成既定操作。感染病灶已被破壞卻無法即時移除，導致嚴重出血。病患又因無法輸血的特殊體質，使情況迅速惡化陷入危險。關鍵時刻渡海醫師與高階醫師即時介入。前者具備深厚的解剖與臨床經驗，後者熟悉機器手臂系統與科技限制，雙方協作排除障礙，最終完成感染源移除，成功挽回病患生命。此一案例清楚顯示，新科技並非必然帶來安全與效益。過度信任或過度排斥，都可能削弱人機協作的實際功能。如何在臨床情境中建立適當的科技理解、使用邊界與決策輔助機制，成為後續導入可解釋人工智慧與臨床治理設計的重要課題。

在醫療情境中，臨床決策的困難性與自動駕駛高度相似，皆涉及在高度不確定條件下即時做出關鍵判斷。當決策完全由人主導，規則與責任由醫療人員承擔；然而，一旦引入人工智慧，臨床決策便不再是單一主體的行為，而是人與系統之間的互動結果，醫療人員也不可能完全退出決策流程。因此，關鍵問題不在於是否使用 AI，而在於醫療人員如何理解 AI 的判斷依據。若缺乏適當理解，臨床決策輔助系統可能導致兩種極端結果：一是過度依賴，盲目遵循錯誤的系統建議；二是自我依賴，忽略正確且有助於臨床判斷的系統輸出。真正理想的狀態，則是醫療人員能適度依賴系統，採納正確建議，同時排除不適當的輸出。《黑色止血鉗》的劇情呈現這三種依賴型態的對照。高階醫師象徵對科技的過度信任，而渡海醫師早期則代表對系統的排斥與忽略；隨著經驗累積，其角色逐漸轉向在專業判斷基礎上，適度運用科技輔助，展現人機協作的成熟狀態。要引導臨床決策走向這種「適當依賴」，前提在於 XAI 的導入。透過清楚的解釋機制，協



助醫療人員理解系統如何產生預測與建議，XAI 成為臨床決策輔助治理設計中的關鍵起點，而非單純的技術附加功能。

XAI 的導入，使臨床決策輔助必須納入治理層面的討論。首先，臨床決策支援系統是否能實際影響醫療決策行為，是 XAI 應用的基本前提。若系統無法取得醫療人員的信任，即使具備預測能力，也難以在臨床現場發揮作用，而缺乏清楚解釋正是醫療人員拒絕採納的重要原因。其次，不同的 XAI 解釋方式，會直接影響醫療人員對系統建議的採納程度。示範型解釋與非示範型解釋，可能引發不同的理解模式與決策反應，進而影響臨床行為。第三，XAI 解釋本身也可能帶來風險，若解釋設計不當，可能導致過度依賴系統，或在錯誤情境下不當使用 AI 建議。此外，醫療人員對臨床決策支援系統的解釋形式存在明顯偏好差異，有些偏好直觀敘述，有些則偏好量化資訊，顯示單一解釋形式難以滿足所有臨床需求。最後，醫療人員的背景與態度，包括年資、專業經驗與科別差異，皆會影響其對 XAI 建議的理解與採納行為。

XAI 主要可區分為兩種核心解釋模式，並直接影響其在臨床決策中的治理設計。第一種為特徵貢獻式解釋，透過 SHAP 與 LIME 呈現各項臨床特徵對模型預測結果的影響程度，作為模型判斷依據的量化呈現。例如，年齡與體重呈現保護性貢獻，懷孕週數呈現正向風險貢獻，模型據此綜合判定整體風險水準。第二種為範例案例式解釋，透過產生多項特徵上與當前個案高度相似的數位雙胞臨床案例，並呈現其實際臨床結果，作為模型建議的經驗對照依據。此解釋形式補足單一病人難以進行反事實臨床比較的限制，使模型判斷更貼近臨床思維。兩種解釋模式分別對應模型判斷邏輯與臨床經驗對照，並可能形塑不同的理解方式與依賴型態。因此，在不同臨床情境中選擇或整合適當的 XAI 解釋形式，構成臨床決策輔助治理設計的關鍵課題。

以孕婦安妮為例，臨床決策支援系統判定其屬於妊娠糖尿病低風險個案。特徵貢獻式解釋顯示，各項臨床特徵對預測結果的影響方向與程度明確可辨。分析結果顯示，年齡、血清白血球數、體重及無糖尿病家族史皆呈現保護性影響，而懷孕週數則為唯一具有風險提升效果的因子。在此情境下，單一風險因子是否足以改變整體風險判定，仍需



進一步整合評估。特徵貢獻式解釋能揭示個別因素的影響方向，但不足以單獨判定最終臨床風險，因此系統進一步引入範例案例式解釋作為輔助。該數位雙胞臨床案例在多項特徵上與安妮高度相似，包括年齡、體重、白血球數與糖尿病家族史，僅在懷孕週數上略有差異。其實際臨床結果顯示，該名孕婦於懷孕期間血糖皆維持正常，未發展為妊娠糖尿病，作為經驗對照支持系統對低風險的整體判定。此例顯示，特徵貢獻式解釋與範例案例式解釋分別提供「因素層次」與「整體結果層次」的資訊，對醫療人員的理解與依賴型態可能產生不同影響。如何在不同臨床情境中適當選用或結合兩種解釋形式，構成 XAI 臨床決策輔助治理設計的重要考量。

研究結果顯示，不論採用特徵貢獻式解釋或範例案例式解釋，XAI 皆能顯著提升醫療人員對臨床決策支援系統建議的採納程度。整體而言，兩種解釋方式在提升採納意願上的效果相近，顯示只要能提供可理解的決策理由，醫療人員即更願意將系統建議納入臨床判斷。採納程度並非僅由解釋形式本身決定，而與使用者特徵相關。從受試者整體分布來看，兩種解釋方式的平均採納程度相近，但進一步依據醫療人員特徵分層分析可觀察到職類與年資差異。相較於助產師與營養師，產科醫師在兩種解釋模式下皆呈現較高的建議採納程度，顯示專科角色與臨床決策責任可能影響對系統建議的接受度。此外年資亦為關鍵影響因素。臨床經驗較少的醫療人員，對特徵貢獻式解釋的採納程度相對較高，顯示此類解釋方式在早期臨床決策學習階段具有較強的輔助效果。年資較高者對 XAI 的採納程度較為保守，反映其既有經驗框架與對新型決策支援技術的熟悉程度不同。整體而言，研究結果指出，XAI 的臨床效益並非僅取決於解釋是否存在，而取決於解釋形式是否能與使用者背景、專業角色與經驗階段相互契合。

不論採用特徵貢獻式解釋或範例案例式解釋，XAI 皆能顯著提升醫療人員對臨床決策支援系統建議的採納程度。兩種解釋方式在提升採納意願上的效果相近，顯示只要能提供可理解的決策理由，醫療人員即更願意將系統建議納入臨床判斷。然而，採納程度並非僅由解釋形式本身決定，而與使用者特徵密切相關。從受試者整體分布來看，兩種解釋方式的平均採納程度相近，但進一步分層分析可觀察到職類與年資差異。相較於



助產師與營養師，產科醫師在兩種解釋模式下皆呈現較高的建議採納程度，顯示專科角色與臨床決策責任可能影響對系統建議的接受度。年資亦為影響因素。臨床經驗較少的醫療人員，對特徵貢獻式解釋的採納程度相對較高，顯示此類解釋方式在早期臨床決策學習階段具有較強的輔助效果；相對地，年資較高者對 XAI 的採納程度較為保守，反映其既有經驗框架與對新型決策支援技術的熟悉程度不同。研究結果指出，XAI 的臨床效益並非僅取決於解釋是否存在，而取決於解釋形式是否能與使用者背景、專業角色與經驗階段相互契合。這也進一步說明，在臨床決策支援系統的治理設計中，解釋策略不宜採取單一模式，而應考量不同使用者族群的需求與理解路徑。

可信任 XAI 決策輔助臨床應用實例：胎兒週齡判別

研究以產科超音波判讀為情境，評估 AI 與 XAI 輔助工具，對臨床人員決策行為之影響。研究設計包含三組：傳統人工判讀對照組、僅提供 AI 預測結果的 AI 輔助組，以及同時提供預測與解釋資訊的 XAI 輔助組。核心關注在三項議題：AI 是否能提升胎齡估計準確度、解釋資訊是否在既有預測基礎上帶來額外改善，以及不同臨床人員對同一套 XAI 的反應是否存在差異。評估指標涵蓋主觀信任評分、AI 介入前後判讀修正幅度，以及臨床人員對 AI 建議採納是否適當，藉此全面分析 XAI 在臨床決策支持中的實際效益。

該研究設計用以評估 AI 臨床決策輔助對胎兒超音波胎齡估算影響，依介入資訊層級分為三組。對照組提供超音波影像，由受試者自行判讀胎齡，並標記判讀信心程度及所依據的影像特徵。AI 輔助組在影像基礎上額外顯示 AI 預測胎齡，受試者可參考模型輸出進行修正，並填寫對模型信任與使用方式之間卷。XAI 輔助組除提供胎齡建議決策外更供熱區圖等運用於判斷胎齡之可解釋演算資訊，使受試者可理解與評估 AI 判斷依據。透過比較三組在胎齡估計、信心評分與決策行為上的差異，本研究評估 AI 與 XAI 在臨床輔助決策中的實際影響，並評估 XAI 解釋性資訊是否有助於建立適當信任與提升臨床判讀品質。

AI 與 XAI 在提高臨床人員對於胎齡判讀準確度與一致性。相較於對照組 AI 與



XAI 介入皆可降低胎齡估計的平均絕對誤差，顯示模型預測能有效提升判讀準確度。其中 AI 組相較對照組有顯著改善，而 XAI 組在平均表現上並未進一步顯著優於 AI，顯示提供解釋本身未必帶來整體準確度的額外提升。然而，從個別臨床人員的變化軌跡可觀察到明顯差異性，部分受試者在 XAI 介入後判讀誤差大幅下降，但亦有部分受試者反而受到干擾，改善幅度有限，顯示解釋資訊並非對所有人皆有一致助益。在臨床人員主觀層面上 AI 與 XAI 組的信心評分皆高於對照組，且 XAI 組高度自信比例最高，顯示可解釋資訊有助於提升使用者對模型與自身決策的信任感。整體結果指出，AI 能穩定提升判讀表現，而 XAI 的價值更體現在「差異化影響」與信任建立，而非單純的平均準確度提升。臨床醫療決策中引入 AI 與 XAI 輔助可有效降低評估變異性，提升判斷一致性。然而，進一步導入 XAI 後，卻出現「使用者表現分化」的新現象。AI 輔助組在導入 AI 預測值參考後，評估表現更趨一致。XAI 輔助入 AI 決策解釋模組後，有部分人員表現更接近模型，但也有些反而受影響偏離原準確值，顯示 XAI 可能加劇人因差異。引入 AI 與解釋介面雖有助於決策支援，但仍需考量使用者對資訊的理解與反應差異，設計須兼顧人因適配。

當臨床人員僅接收 AI 模型預測結果時，整體仍表現出高度信任。受試者對「AI 可提供正確建議」與「系統順利應用於胎齡估算」的認同度最高，顯示 AI 在臨床輔助決策中具潛力。大多數正向評分集中在「同意」與「非常同意」，如「我對 AI 建議有信心」、「AI 輸出具良好預測力」等題項均呈現明顯正分布。相對地，對 AI 產生質疑的選項（例如「我不信任 AI 預測建議」）則獲得最低評分，僅少數人表達不同意見，顯示臨床人員普遍不擔憂其可靠性。此信任基準可作為後續 XAI 系統成效評估的重要比較標準。針對 AI 預測建議系統加入 XAI 解釋後，使用者對 AI 信任感反而出現下降趨勢。受試者對如「我信任 AI 建議」、「AI 演算法很可靠」等正向敘述的同意度下降，顯示解釋未能有效增強臨床人員對 AI 的信心。同時，負向敘述如「我不信任 AI 預測建議」的同意比例略有上升，研究團隊觀察到，XAI 輔助反而誘發更多中立甚至懷疑態度，尤其在醫療風險敏感場景下，部分使用者產生更多警戒與不確定感。引入 AI 解釋功能設計時，須



審慎考量其對信任感的實際影響，避免“多解釋反失信任”的反效果。

加入 XAI 工具是否有助於臨床決策，取決於使用者對解釋的接受度與個人背景。調查結果顯示，認為解釋有幫助的參與者中，XAI 組回覆自信明顯提升（151 人），高於純 AI（120 人）及對照組（84 人）。但在認為解釋無幫助族群中，XAI 組仍有較多人表示提升自信（139 人），略高於 AI（125 人）與對照（86 人），但也出現更多人反而信心下降的回覆（41 人）。XAI 解釋效果並非普遍有益，須考量解釋形式、認知負荷、品質一致性等工具特性，以及使用者臨床經驗與推理模式，才能真正發揮其輔助潛力。

以上內容將在 **2026 年 1 月 14 日(三) 10:00 am** 以線上直播方式與媒體朋友、全球民眾及專業人士共享。歡迎各位舊雨新知透過星球永續健康網站專頁觀賞直播！

- 星球永續健康網站網頁連結: <https://www.realscience.top/7>
- Youtube 影片連結: <https://reurl.cc/o7br93>
- 漢聲廣播電台連結: <https://reurl.cc/nojdev>
- 不只是科技: <https://reurl.cc/A6EXxZ>



講者：

陳秀熙教授/英國劍橋大學博士、許辰陽醫師、陳立昇教授、嚴明芳教授、林庭瑀博士

聯絡人：

林庭瑀博士 電話: (02)33668033 E-mail: happy82526@gmail.com

劉秋燕 電話: (02)33668033 E-mail: r11847030@ntu.edu.tw