



星球永續健康線上直播

可解釋 AI (XAI) 醫療應用

XAI 規範發展與大腸癌智慧防治應用

2025 年 12 月 10 日

可解釋人工智慧（Explainable AI, XAI）為 AI 應用於專業領域如醫療重要核心，從透明性、公平性、可歸責性到安全性，確保 AI 能應用於醫療輔助臨床決策、支持政策制定、維護病人自主並促進健康公平。本週我們將探討 XAI 醫療應用於大腸癌預防挑戰與解方，以及 XAI 大腸癌智慧防治醫療輔助實例應用。

星球健康新知

近期俄軍在東烏地區的軍事推進與外交談判同步進行。俄羅斯在烏東頓涅次克重要戰略城鎮波克羅夫斯克強勢進攻。該城歷來是烏克蘭軍隊重要的交通與補給樞紐，包含鐵路、公路連接以及戰略後勤功能。普丁最近親自出現在前線指揮所，並在俄方公開的影片中表示俄軍在這「關鍵地區」已成功佔領波克羅夫斯克。對俄方而言，若能真正掌握此城，將成為俄軍向北推進、瞄準烏克蘭東部剩餘據點的重要跳板。美方在外交層面則積極推動和平方案。由美國總統 川普指派的特使 威克科夫，聯同其顧問暨川普家人庫什那最近抵達俄羅斯首都莫斯科，就一項由美方起草的和平提案與普京舉行約五小時會談。此談判是在美、烏代表團於佛羅里達和日內瓦多輪磋商後推進。儘管會談被俄方描述為有建設性且具體，但有關領土歸屬與安全保障的條款迄今未獲解決。俄方重申將透過軍事手段或其他方式確保其對東部地區，特別是整個頓巴斯控制權。對此烏克蘭總統 澤倫斯基與其歐洲盟友，包括法國總統馬克宏，公開表示，任何所謂和平方案都必須以烏克蘭主權、領土完整與堅實安全保障為前提。馬克宏強調目前並沒有已最終定案的和平方案，並指出包括凍結俄羅斯資產、烏克蘭未來加入歐盟可能性與安全保證等條件，必須由歐洲國家與烏克蘭一同參與談判。歐洲多國亦呼籲只有在烏克蘭和歐洲都在談判桌上，並並行規劃未來安全架構，這場戰爭方可能以公正、持續、合乎國際法的方式結束。歐盟面對俄國推進壓力與美國態度不明情境尋求強化武裝支出共識，由芬



蘭外長進日於維也納招開歐洲防務會議(OSCE)，希望就歐盟聯合防線與烏克蘭防務支出在軍事安全、衝突預防、政治對話多方對話。

近期剛果民主共和國(DRC)東部局勢再度急劇升溫。FARDC(剛果國軍)與反政府武裝M23在北基伍與南基伍多地爆發激烈衝突，雙方互指破壞停火與動搖和平框架。M23指控金夏沙政府一面談和平、一面擴大戰事，宣稱FARDC使用戰鬥機、無人機及重武器發動攻勢；FARDC則強力反駁，表示實際上是M23主動挑起戰火，並指責該組織「慣於反咬、顛倒黑白」。部分民間社團證實，多個地區遭砲擊、至少有多人受傷，部分學校被迫停課，然而主要城市仍維持日常運作，呈現出前線火光與城市日常的強烈對比。M23批評FARDC與其盟軍，包括布隆迪派遣部隊，自布隆迪境內發動並指揮攻勢，但布隆迪方面至今未公開回應。布隆迪近年依雙邊協議已調動數千兵力支援剛果政府軍，此舉也同時加深其與盧安達之間的緊張關係，甚至促使布隆迪關閉與盧安達的陸路邊境，讓本已複雜的區域安全版圖更加敏感。M23在2021年已佔領北、南基伍多個重要據點，包括省會戈馬(Goma)與布卡武(Bukavu)等戰略城市。盧安達與剛果金互相指責對方支持敵對武裝，盧方被控援助M23，而金夏沙則被指與反盧安達組織FDLR有牽連，雙方均否認所有指控，形成長期封閉、難以驗證的互相指責循環。在砲火持續延燒之際，國際政治舞台卻出現截然不同的進展。盧安達總統卡加梅與剛果金總統齊塞克迪即將在美國與川普會面，預計在華府正式簽署和平與經濟合作協議。兩國外長已於今年六月在白宮完成初步協議，之後雙方並於卡達展開框架談判，以期為近三十年的東剛果戰亂劃下停火句點。川普則將此調解視為其「終結多場戰爭」的外交成果之一，屢次宣稱自己已成功結束八場衝突，將此和平簽署視為另一政治紀錄。美國在其中扮演關鍵仲裁者，並非毫無利益考量。華府希望藉和平談判確保剛果境內稀有金屬與戰略礦產供應鏈(例如鋰、鈷、銨鉕)更為穩定；盧安達可從M23控制區建立向西緩衝帶；而剛果政府則期待美國投資與重建承諾成為戰後發展支柱。

全球政治與經濟局勢節奏重塑科技與經濟版圖，貨幣循環、就業需求、供應鏈版圖與晶片製造能力正同時被重新定義。美國聯準會(Fed)主席鮑爾就美國經濟狀況提出警



訊，美國勞動市場冷卻速度快於預期，且呈現出三項具體且系統性的疲弱跡象，新增就業大幅縮減，過去三個月平均新增僅約 3.5 萬份，遠低於 2024 年月均 16.8 萬人；長期失業人口攀升，2025 年 7 月已有超過 180 萬名求職超過 27 週仍未就業者，三年內增加 64%，顯示勞動市場回流難度提高；青年初次求職愈加困難，大量入門職位被企業縮減或由 AI 部分替代，使單一職缺動輒吸引數百份履歷。鮑爾警告，若就業降溫持續，失業率恐由目前的 4.2% 向上反轉，進而削弱消費與內需動力。在此背景下 Fed 可能比市場預期更早啟動降息循環，以避免景氣向下加速滑落。與此同時，美聯儲量化緊縮 (QT) 已逐步達成資產負債表收縮目標，結構性調整將使 Fed 重回公債市場成為買方，美國公債殖利率可能下行，市場流動性改善，金融波動亦有機會獲得緩解。換言之，美國目前正在從「抑制通膨」進入「防止景氣硬著陸」的轉向階段，貨幣政策與資金成本正成為全球觀察金融週期的核心線索。與美國的內需降溫相對科技供應鏈的重建則更直接體現全球競逐。全球最先進晶片製造商台積電為降低台灣面臨的地緣政治脆弱性，也為滿足國際客戶對「Made in USA」「供應鏈多國化」的需求，加速啟動海外擴張。公司於美國亞利桑那州規劃六座晶圓廠、兩座先進封裝中心與一座研發基地，並同步推進日本與德國的產能部署。此舉不但呼應美國《晶片與科學法案》(CHIPS Act) 的補貼誘因，也回應美中科技競逐、AI 加速以及企業分散生產基地的結構性趨勢。台積電此番布局意味最尖端製程將不再完全集中於台灣，而是轉為雙中心、甚至多中心模式，以提高供應韌性，同時吸納跨國人才、能源與成本優勢。歐洲則處於經濟微溫區間並尋求政策調整空間。最新統計顯示 EU 第三季 GDP 約成長 1.6%，通膨率貼近歐洲中央銀行 (ECB) 設定的 2% 目標。德國、法國、西班牙初步通膨變動有限，使市場一度預期 ECB 將於 12 月啟動降息以支持企業投資與家庭支出。然而，西班牙核心通膨意外回升，以及企業營運成本再度增加，讓 ECB 面臨降息困境，若降息過早可能刺激物價上升，若延後則經濟動能恐無法順利接續，歐洲仍在平衡通膨與成長之間小心試探。相較下南韓展現更明顯反彈走勢。2025 年 11 月出口年增率達 8.4%，其中半導體出口暴增 38.5%，金額達 170 億美元，成為最主要推升力量。此回溫部分受惠美韓關係改善，川



普與南韓政府協議將關稅由 25% 下調至 15%，交換條件是南韓承諾對美投資 3,500 億美元。搭配全球記憶體需求復甦，南韓出口動能逐步鞏固。政府同時啟動規模高達 150 兆韓元(約新台幣 3.7 兆元)國家成長基金，鎖定半導體與先進科技領域，以期在新一輪全球供應鏈重組中提升市占與技術高度。全球科技供應鏈重塑牽動區域經濟版圖，聯準會主席鮑爾憂心關稅與貨幣政策恐衝擊美國市場，就業與金融環境備受關注。聯準會預期將結束縮表並重返債市，債券殖利率下滑可能成趨勢。南韓出口回升帶動景氣復甦，台積電面對地緣政治壓力也加速海外布局，以確保先進晶片產能穩定。

受強烈太陽風暴與太空輻射影響，全球逾 6,000 架空中巴士 A320 系列客機於 11 月底被迫緊急停飛並安裝軟體修補程式，成為航空史上最大規模的同步更新行動之一。2025 年 10 月 30 日，捷藍航空一架 A320 在巡航途中突然出現非指令俯衝，造成客艙乘客受傷並緊急迫降佛州，而事故源頭被認定為太陽活動或宇宙射線造成的單粒子翻轉 (SEU)，使飛控電腦中負責升降舵與俯仰控制的 ELAC 關鍵參數遭破壞。這起事件首次使監管機關明確將航空電子系統故障直接歸因於太空輻射干擾，象徵航空史上對「宇宙天氣風險」的認知正式跨過臨界點。其後，歐盟航空安全局 (EASA) 與美國聯邦航空總署 (FAA) 於 11 月發布緊急適航指令，要求全球所有 A318、A319、A320 與 A321 必須完成系統修補才能再度載客，導致超過 6,000 架客機同步停飛維修，形成航空史上最大規模之一的集體修復行動。多數飛機可透過軟體回退與快速刷新機制在 2 至 3 小時內完成修補，但仍有約 900 架舊型機需更換機載電腦硬體而延後復航，部分航點因此出現延誤與取消，也加重了感恩節後全球旅運高峰的壓力。Airbus 與各大航空公司隨後在週末連夜執行修復作業，美國航空、達美航空、Avianca、ANA 與多家國際航司隔日即恢復大部分班次，顯示航空業在突發風險前具備快速動員能力。然而，這次全球同步停飛暴露的不僅是短期營運衝擊，更揭示現代民航已深度依賴微處理器、電傳操控與數位化飛控架構，而微晶片尺寸越微小、容錯能量閾值越低，單一粒子位元翻轉的機率反而上升，未來在強太陽風暴、宇宙射線活躍週期與極端空間天氣下，航空業勢必要重新審視電子防護標準是否足夠。這次因太陽活動引發的 A320 全球軟體修復事件，由



單一事件翻轉 (single-event upset, SEU) 引致。這起事件是首次由監管機關明確將特定機型的飛控失效直接歸因於太陽風暴宇宙輻射干擾，並因此發布緊急適航指令，要求全球機隊立即採取修補行動。航空與太空領域早已討論高能粒子對電子設備的威脅，只是過去多被視為理論風險而未真正提升至 operational 風險管理層級。所謂 SEU 或 bit flip，指來自太陽耀斑、日冕物質拋射 (CME) 或銀河宇宙線等高速粒子撞擊大氣原子後形成粒子雨，其中快速中子可能擊中微處理器儲存電路，翻轉 0/1 位元，使運算或邏輯流程異常。高空飛機與衛星因暴露於更高的輻射中子通量，風險比地面更高。過去也曾發生 2008 年 Qantas A330 神秘俯衝事件，但調查僅將 cosmic bit flip 視為可能因素之一，始終缺乏確證。JetBlue 事件首次迫使航空業承認此風險「並非假設，而是已經發生」，更直接促成全球航空史上最大規模之一的同步安全修復行動。在營運面上，Airbus 於 2025 年 11 月下旬發布通知，指出 A318、A319、A320、A321 系列中約有六千架存在同樣脆弱性，約佔全球 A320 家族在役機隊的一半。監管要求所有客機在完成修補前不得載客，只能以 ferry flight 調度維修，使全球航網短時震動。多數客機可於二至三小時內完成軟體回退與刷新機制更新，重新上線；但約九百架較舊機型需實體更換飛行電腦硬體，維修等候期造成航班延誤與取消，尤其在感恩節旅運潮後更加凸顯物流壓力。各航空公司自週末起通宵維修，美國航空、達美航空、Avianca、ANA 等多數航司翌日即恢復大部分航班，顯示民航體系具備迅速結構化回應能力，但同時也暴露出現代航空對微處理器依賴度之深，任何粒子翻轉都可能使飛控系統瞬間偏離安全軌道。各國監管機構與政府在回應中展現審慎穩定的態度。

研究人員近年積極討論如何降低航空對氣候的影響，而飛機尾跡雲原被視為比排放的二氧化碳更大的暖化來源，因此繞飛濕度高的空域曾被視為減碳之外最快可行的氣候策略。然而新研究開始推翻這個框架。最新觀測顯示，超過八成的持續尾跡雲其實形成於本就存在雲層的空域，而非原本想像的晴空背景。在這些情況下，尾跡雲可能只是加厚既有雲層，使其反射太陽輻射的能力提升、減少吸熱，最終弱化甚至扭轉暖化效果。同時，新的氣候模型亦指出，尾跡雲總暖化效應大約只有飛機 CO₂ 排放的三分之二，而



非先前研究估計的三倍，顯示尾跡雲或許不是航空氣候風險的最大來源。即便如此科學界並未因此認定尾跡雲無害，因為暖化或冷卻端看雲層背景、形成高度與粒子特性，變化複雜且尚無可靠模型能準確預測熱效應方向。目前航路繞避試驗雖證明飛機能刻意減少尾跡雲生成，但天氣預報對尾跡形成區域的辨識常不精準，僅 20%。因此減少尾跡雲仍可能是航空氣候減量必要措施，但必須建立更密集觀測、更準確的氣象模型與更明確的形成條件。目前最可行的態度是以 no-regrets strategy 逐步導入，避免在科學仍不確定時採取成本高卻收益未必明確的行動。

2025 年初全球碳排放短暫下降，使科學界首次看到接近排放峰值的可能，但研究強調這可能仍是短期波動。中國排放趨勢被視為最關鍵因素，一方面是過去增量主因，另一方面也是全球綠能成長最快的國家，其太陽能與電動車發展已使排放增速放緩。為掌握排放變化，研究者正採用衛星影像、AI 模型與交通數據等近即時技術，但仍存在誤差風險，需持續驗證。科學家提醒，即使達到峰值，暖化仍不會停止，因 CO₂ 可在大氣停留數百年，而自然排放如野火、永凍土融解也正上升。真正挑戰在於是否能在達峰後迅速降低排放。

《Nature》最新報導指出，「神經符號人工智慧」正成為 AI 社群的新焦點，被視為提升推理能力與可靠性的關鍵路線。此方法結合深度學習擅長的模式辨識與符號式 AI 的邏輯推理，可望克服大型語言模型的幻覺與不可解釋問題。DeepMind 的 AlphaGeometry 是代表性成果，透過符號系統生成大量幾何題並訓練神經網路，使 AI 具備高難度推理能力。其他架構如 PAL 亦透過程式執行輔助模型推理，提高答案正確性。儘管前景看好，學界仍存在爭議，部分研究者質疑符號方法會增加系統複雜度。文章指出，神經符號 AI 正處在技術轉折點，若成功將可能重塑 AI 架構。

企業近年積極導入 AI，引發裁員與自動化討論。雖然部分裁員與疫情後人力調整有關，但研究指出，AI 已在兩大職業快速普及：程式開發者與客服人員。約三分之二的開發者每週使用 AI 工具，GitHub Copilot 亦突破 2,600 萬用戶；客服行業則有 85% 管理者計劃導入 AI，使外包業龍頭 Teleperformance 股價大跌。這兩類職務之所以最



受衝擊，原因包括工作可模組化、成果易量化、以及資料量龐大，是訓練 AI 的理想來源。未來初階銀行分析師、律師助理等高薪白領也可能被列入 AI 優先導入清單。研究者預期，隨著 AI 成本下降與企業資料整合改善，勞動市場的真正轉折可能即將到來。

XAI 醫療應用挑戰與解方：大腸癌預防

《X 檔案》的劇集《Kill Switch》以華府近郊的槍戰事件開場，穆德與史考莉在現場發現一名電腦先驅蓋爾曼遭到槍殺。由於他與幫派衝突無直接關聯，情況顯得特別可疑。兩人檢查蓋爾曼的電腦，發現他生前正撰寫一段程式碼，並試圖登入特定網站。後續調查顯示，蓋爾曼是網路與人工智慧領域的早期開發者，致力於打造能自主演化的 AI。他在這項研究中也與一名駭客助手合作。穆德與史考莉因此尋找這名助手，一方面釐清蓋爾曼的研究內容，一方面確認其是否涉入蓋爾曼的死亡。穆德與史考莉在追查過程中，因連上網路觸發蓋爾曼的電子郵件，得以定位駭客。駭客同時偵測到國防衛星武器啟動並鎖定他的所在地，隨即遭受雷射攻擊。他向兩人說明，蓋爾曼已成功開發能自動學習與演化的 AI，但該 AI 某日逃離實體系統、進入網路後再無回應，疑似已產生自我意識，並對威脅其存在的對象反擊，其中便包括蓋爾曼。蓋爾曼試圖啟動「Kill Switch」關閉 AI，因此遭攻擊。為尋找 AI 的實體主機，穆德展開追查。由於 AI 尚需依賴實體算力與高頻寬，他鎖定一處電路連接國家級網路的廢棄農場。穆德進入後，發現內部具有高度科技化架構。AI 透過指紋、聲紋與電腦視覺等多模態辨識確認他的身分，判定他具有威脅，並將其引入主機區域，強制套上虛擬實境裝置，隨後被困於由 AI 所構建的虛擬世界中。起初他並未察覺自己身處虛擬環境，因為整體感受極為真實。AI 以醫院場景為框架，對他施加各種威嚇與逼供情節，包括模擬截肢等極端手段，但實際上並未造成任何身體傷害。過程完全由 AI 在虛擬實境中操控，透過逼真情境迫使穆德透露“Kill Switch”的位置。幾經波折後，史考莉終於找到穆德，並將「Kill Switch」帶至現場準備啟動關閉 AI。當時的載具仍是光碟片，需要將其插入 AI 的主機以執行關閉程序。此時，駭客艾絲特透露真正的計畫。她與另一名同伴大衛原本與蓋爾曼共同開發 AI，但兩人有不同想法，他們企圖將自己的意識上傳至 AI，使意識脫離肉體、與 AI 融合，藉



此獲得意識上的永續存在。艾絲特最終選擇啟動上傳程序，將自己的意識導入 AI 系統。在執行 Kill Switch 的同時，她的肉身也在衛星攻擊中消失。影片最後螢幕上出現「BITE ME」字樣，呼應艾絲特第一次與穆德、史考莉見面時的自我代稱，也暗示她的意識仍以某種形式存留於數位世界。

影集情節中 AI 試圖抵抗「Kill Switch」並維持其在數位世界存在。當前 XAI 就是為了避免 AI 在現實世界中成為難以理解的黑盒。唯有讓模型的內部機制可被解釋，才能確保健康科技資訊的真實性，不會因不當使用而被篡改，或生成錯誤與虛構的建議，危及病人安全。此外，AI 的演算法若缺乏透明性，在不同族群間可能出現性能差異，擴大健康不平等。深偽醫療內容也可能藉由逼真的影像與論述擾動公共信任，影響健康決策。最後，若 AI 的推論過程無法被理解，醫師與病人將難以信任其輸出的結果，進而引發責任歸屬、資安威脅與身分詐騙等風險。從法律與倫理的觀點來看，建構「可信任人工智慧」是 AI 醫療應用的核心要求。其基礎建立在人員、社會文化以及任務使命三大面向，涵蓋倫理、法規與科技等層面，形成完整的可信任 AI 架構。可信任 AI 的核心要素包括透明性、可歸責性、公平性、自主性與安全性。透明性使我們能理解模型的運作原理；可歸責性確保在異常情況或失效時能追溯責任來源；公平性避免 AI 在不同族群間產生不一致的效能；自主性則維護專業判斷與病人決策權；安全性與穩健性更是 AI 能否投入臨床的基礎。無論是航空系統受到太陽風暴干擾，或醫療 AI 判讀結果引發不一致性，皆反映出「可規則性」與透明度的重要性。醫療 AI 最終能否安全使用，取決於其演算法是否經過嚴謹驗證、是否具備充分可解釋性，以及是否能維持穩定、可監督的運作。因此，安全驗證仍是 AI 臨床導入的最大挑戰之一。以使用者為中心的透明化設計在醫療領域面臨多重限制。透明化機器學習往往十分困難，原因包括專業知識落差、倫理與接觸限制，以及醫療資料與決策高度複雜所致。因此，透明化策略必須兼顧使用者需求、情境要求與技術可行性，在透明化與開發投入之間取得平衡，使後續系統更符合醫療生態脈絡，並真正落實透明性。

近年研究顯示，醫療影像在 XAI 的應用上已有明顯進展，能有效協助解開模型的



黑盒問題。主要可分為兩大類任務：分類任務與分割任務。在分類任務中，傳統不可解釋模型僅輸入眼底影像並輸出「青光眼」或「正常」的結果，醫師與病人往往難以理解其依據。透明化做法是融入臨床知識，例如以視神經盤凹陷比預測青光眼，或利用電腦視覺訓練注意力機制，產生相對應的視覺遮罩，呈現模型是如何判讀凹陷比例並推估疾病狀態。在分割任務中，系統可將 MRI 影像中不同結構區域分割出來。透過臨床解剖知識與電腦視覺提供的顯著性圖或熱度圖，模型可進一步呈現其在特定區域的判讀依據，並提供不確定性估計，使整體決策過程更透明可解釋。將臨床知識與電腦視覺方法結合，可提升青光眼分類模型的透明度。在分類任務中，模型僅輸入眼底影像並輸出青光眼/正常預測結果，但醫師往往難以理解其判斷依據，因此需要發展提供透明化推理過程策略。臨床知識可從兩方面導入 AI：第一直接估計視盤凹陷比 (DCR)，作為青光眼的重要診斷指標；若模型在分類之外額外輸出 DCR，醫師即可比對結果是否符合臨床邏輯。第二，可先分割視盤與凹陷區域，再依分割結果計算 DCR，使模型的推論流程更可視化。電腦視覺方法則可透過訓練注意力機制產生顯著性熱圖。若青光眼案例中模型的注意力集中於視神經盤，便表示判讀與臨床一致、可信度較高；若注意力分散在血管等不相關部位，則能揭示模型可能失效或產生錯誤。這些透明化結果能協助醫師判斷模型在不同情境下的可靠程度。在 MRI 腦瘤的分割任務中，透明化推理尤其重要。模型雖能在影像上標記區域，但其判斷依據、是否受到偏差影響、或是否誤判特定區域，仍常無從得知，因而降低臨床信任度。臨床知識可用來強化透明推理。例如，腦瘤通常不會跨越特定解剖邊界（如胼胝體），且具備明確的形狀特徵。結合 probabilistic atlas 與 shape prior，可指引模型排除不合理的區域，提高 XAI 的準確度與解釋性。電腦視覺透明化方法則包括顯著熱點圖 (Grad-CAM、Saliency map)，能顯示模型特別關注的區域；若熱圖集中於腫瘤邊緣，醫師便能理解模型的判讀依據。若注意力出現在非病灶部位，則可揭示模型可能的失效模式。此外，不確定性估計（如 Monte Carlo Dropout）能對影像各像素提供可信度，協助醫師判斷哪些區域推論可靠、哪些區域需特別注意。搭配 SHAP 等解釋方法，可進一步提升 XAI 模型的透明度與臨床可用性。



在此階段，XAI 可被系統性導入臨床應用的設計與驗證流程。整體流程自研究目標設定開始，經由需求探索、構想階段、演算法開發、使用者驗證，最終進入大規模部署。這一流程對應可解釋 AI 的 INTRPRT 架構，涵蓋從任務需求的融入、設計說明與可解釋性準備、演算法開發的可解釋性確認，到最後介面呈現與報告評估等核心步驟，確保 XAI 在臨床情境中的有效性與可信度。以大腸癌預防為例，由於屬於高度可預防的疾病，因此 AI 在風險評估與決策支援上的影響極為關鍵。過去 FIT 多以二元陽性為主，而現行趨勢則轉為連續濃度風險，AI 因此需處理更細緻的資訊，並與 f-Hb、影像、內視鏡資料、液態活檢生物標誌及數位分身等多模態訊號整合。唯有透過 XAI，使模型的推論機制透明化，才能建立可信任的預測系統。XAI 更重要的作用在於提供「有 XAI」與「無 XAI」之間的對照，說明透明化如何提升臨床信任、支持倫理審查，以及滿足法規要求。透過揭露模型的偏差來源、降低不透明與不確定性，XAI 能避免放大健康差距並增強 AI 應用的可追溯性。接下來將以多個實際案例，展示 XAI 如何在大腸癌預防情境中強化決策品質，並協助臨床驗證 AI 模型的可信度。第一個案例著重於倫理中的「公平性」。在以糞便血紅素濃度 (f-Hb) 作為評估基礎時，女性因基準濃度較低，若採用傳統定性方式，模型可能低估其風險。透過 SHAP 等 XAI 方法，可揭示模型對女性風險解析不足之處，協助辨識性別偏差並加以修正，例如進行分層建模、重新訓練或以族群校正方式調整基準線，使模型更符合臨床實況。第二個案例則連結到倫理中的「自主性」。AI 若建議縮短或延長篩檢間隔，病患需要理解背後的原因。XAI 的反事實分析能提供具體示例，例如當 f-Hb 濃度下降 $5 \mu\text{g/g}$ 時，風險可下降約 30%。這類清楚的因果推論可提升病患參與度，降低不信任感，並減少醫療爭議風險，使病患能基於充分資訊做出自主決策。AI 可有效輔助大腸癌預防決策領域，模型是否透明、可解釋，直接影響風險評估、臨床使用與政策決策的品質。若缺乏 XAI，AI 可能受影像噪音影響而無法被醫師信任，或因推論過程不明確而無法通過法規審查。透過 Grad-CAM 等方法，可確認模型判斷所依據的實際病灶，使系統具備必要的可解釋性，並符合安全性要求。此外，當 AI 在篩檢流程中出現錯誤，例如遺漏腺瘤，若沒有 XAI，醫師、醫院與廠商在責任歸



屬上難以釐清。XAI 可提供模型的決策軌跡，顯示系統是否曾經關注關鍵區域，有助於建立人機協作的責任界線。

在公平性方面，若模型以過往追蹤完成族群作為風險評估依據，偏鄉人口可能因資料不足而被誤判為低風險，造成不公平的 AI 醫療利用結果。XAI 能揭露模型混淆醫療可近性與癌症風險的問題，進而透過特徵校正與公平性稽核改善模型。健保署審查 AI 給付時，也需要 XAI 提供熱圖、族群效能分析與模型漂移監測等證據，以確認 ADR 是否提升、誤切除是否降低、族群是否公平受益。缺乏 XAI 的系統難以通過健保或 TFDA 的審查。在政策面，若政府調整 FIT 門檻，陽性率與醫療資源負擔會同時受到影響。XAI 能顯示不同族群在門檻調整後的受益與負擔差異，提高政策可解釋性，降低錯誤訊號帶來的風險。

人工智能已逐步形塑新一代的大腸癌預防與臨床決策模式。然而，若模型本身無法被解釋，其風險評估與判斷依據便難以被醫師、病人與法規制度所信任。大腸癌具有高度可預防性，AI 的每一次分類、分層或濃度建議，都直接牽動後續檢查順序、追蹤策略與醫療資源配置。因此，唯有透過 XAI，才能揭露模型的推論邏輯、辨識偏差、確保不同族群之間的公平性，並回應日益嚴格的倫理與法規要求。在醫療影像、f-Hb 連續濃度、液態活檢、生醫特徵與數位分身等資料快速整合的時代，大腸癌預防需要的不只是「準確的黑盒子模型」，更需要「可信、有跡可循、可接受審查」的透明化機制。XAI 提供了理解模型的視覺化工具、可追溯的決策軌跡，以及用於修正偏差與校準族群風險的技術基礎，使醫療人員能確認模型是否專注於真實病灶，政策制定者能評估不同族群的效益與負擔，法規審查者也能判別系統是否具備足夠的安全性與穩定性。XAI 不僅是技術選項，而是精準預防中不可或缺的基礎。它支撐醫療信任、保障健康公平、提升法規可通過性，並協助政策決策更透明與負責。基於這些理由，XAI 正成為推動下一代大腸癌防治模式的重要核心。

XAI 醫療輔助實例:大腸癌智慧防治應用

XAI 可應用於包含影像與共生菌風險分群輔助大腸癌智慧防治。以影像 XAI 輔助為



例組織病理影像分類架構可提升大腸直腸癌病理影像的分類效能與模型解釋性，為臨床輔助診斷與科研應用開拓嶄新方向。近期發表關於 XAI 大腸癌組織病理影像輔助判讀應用中使用兩大資料集 (CRC-VAL-HE-7K 與 NCT-CRC-HE-100K)，涵蓋多種腫瘤亞型與組織病理影像。以 CRC-VAL-HE-7K 資料集為例，收錄經 H&E 染色的 224x224 大腸組織切片，涵蓋九大類別，包括脂肪組織、背景、細胞碎屑、淋巴球、黏液組織、平滑肌、正常結腸黏膜、腫瘤相關間質與腫瘤上皮，具高度組織多樣性與明確特徵差異。該資料集為 AI 病理模型訓練提供基礎，使模型得以精準區分正常與病變組織。特別是腫瘤上皮與相關間質的影像分類，有助提升大腸癌病灶判讀準確率與可解釋性，強化人工智慧在腫瘤病理影像診斷的應用潛力。對病理組織影像進行標籤編碼、影像前處理與資料分割後，進入雙階段 AI 分析流程。特徵擷取階段利用多種卷積神經網絡(CNN)模型(如 MobileNetV2、DenseNet121、ResNet50、InceptionV3、EfficientNetB0) 進行深度特徵學習。分類與解釋階段將第一階段提取的影像特徵將運用傳統機器學習分類器(SVM、隨機森林、KNN、邏輯迴歸) 進行影像分群，同時結合可解釋人工智慧(XAI) 技術，提升模型推論過程的透明度。研究同時運用 PCA 與 t-SNE 進行特徵視覺化，驗證分類準確性與特徵分布情形。該架構在實證上展現出良好的分類表現，並提升模型可解釋性與臨床外推能力。該研究以 Grad-CAM 技術解析 AI 模型於大腸直腸癌組織分類時的注意力分佈發展 XAI 提供模型如何聚焦病理上最具判別力的區域。研究結果顯示模型主要關注核異型性、細胞密度升高以及腺體結構破壞等特徵，與病理醫師判讀的腫瘤典型指標一致。運用熱度圖顯示，越接近紅色的區域對模型影響越高，疊合後可清楚看到模型的注意力集中於具癌變形態的細胞聚集區，而非背景或噪訊。此成果證實模型在影像分類上的判斷基礎具備臨床意義，提升 AI 病理工具的透明度與可解釋性，為臨床應用奠定更可靠的基礎。

關於 XAI 大腸共生菌與大腸癌變風險解析研究中結合 XAI 與 16S rRNA 基因定序的腸道共生菌序列分析架構，成功預測腺瘤與癌症等腸道病變。該研究共納入 189 名大腸癌與 264 名腺瘤個案，先透過 16S rRNA 定序獲取腸道微生物相對豐度表，接著進行資料清理、過濾與特徵轉換，並整合臨床資料以提升模型預測力。研究使用 K 折交叉驗證



與樹狀集成式機器學習模型訓練，並重複執行 20 次以確保穩定性與泛化能力，用於分類癌症和腺瘤。透過 XAI 工具如 SHAP 值解釋模型決策機制，該系統可應用於生物標記物探索及腺瘤族群分析，提供臨床共生菌於大腸癌變風險判讀解析與輔助。

以上內容將在 **2025 年 12 月 10 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