



星球永續健康線上直播

可解釋 AI (XAI) 醫療應用

XAI 規範發展與中風輔助應用實例

2025 年 12 月 3 日

目前 AI 技術快速發展，但應用於如醫療等專業領域仍有距離，主因與 AI 預測是否具可解釋，因此可解釋人工智慧(XAI)不僅能提升模型的透明度，更使臨床人員清楚理解其判斷依據，進而加強醫療 AI 的安全性、可歸責性與公平性。在急性中風的診斷與治療流程中，XAI 的重要性更是格外明顯。本週我們將聚焦於 XAI 在醫療中的規範發展，並透過中風病患的決策輔助歷程說明 XAI 在臨床應用中之可能價值。

星球健康新知

近期國際社會將焦點再度集中於俄烏戰爭的外交與軍事發展，因為美國正以前所未有的力度推動和平談判，而戰場局勢卻同時出現升溫跡象，使整體情勢呈現高度複雜的雙面樣態。美國在過去數周密集與烏克蘭商討新版和平協議內容，將原先由美俄雙方起草、條目繁多且含有多項爭議性的二十八點方案，縮減與重整為較具可行性的十九點。新版內容中，要求削減烏軍規模的條款已被撤除，過去被視為替俄方大規模開脫的全面赦免也遭刪除，而涉及烏克蘭外交自主的部分則變得更為彈性，未來加入北約與歐盟亦為潛在可能性，使烏方在國家定位與安全保障上不被提前鎖死於範圍侷限政治框架之內。美國官員透露，烏克蘭原則上接受新版內容的大方向，認為其至少能維持主權主體性與戰後發展的空間，但針對領土劃分、戰後俄軍部署限制、邊界監管機制、戰後武裝力量配置及外部安全保證等最敏感議題仍保留彈性與談判空間。歐洲多國政府則反覆強調，任何和平方案必須建立在尊重烏克蘭主權與領土完整的基礎上，不得讓歐洲安全架構因政治交換而受到削弱。這類聲明反映出歐洲國家擔憂若方案過度偏向俄方訴求，可能使未來的安全秩序失衡，甚至重演冷戰後期東歐安全被犧牲的前例。隨著美烏協商有所進展，美國進一步安排高層赴阿布達比，與俄羅斯代表團秘密商談，試圖確認俄方對十九點方案的實際底線與可接受程度。雖然會談氣氛被形容為出乎意料地穩定，俄方也承認



其中部分條款具有可討論性，但俄羅斯仍表示需進一步修改內容，顯示雙方在停火條件、戰後地位安排與國際監督機制等領域仍存在深刻歧見。烏克蘭總統澤倫斯基在此期間多次表示，最具政治敏感性的議題仍需與美國總統川普直接對談，尤其是涉及戰後軍事中立、國防重建模式、戰犯追責、邊界劃設及烏國未來的國際定位等關鍵項目，他也重申歐洲主要國家不能被排除在談判之外，因為未來戰後安全安排必須在整個歐洲架構內保持一致性。另一方面，俄方動作並未因外交進展而減緩，反而在軍事層面加大施壓。基輔近期遭遇大規模無人機與飛彈攻擊，空襲不僅造成至少七名平民喪生，多處能源設施也因爆炸與火災而受損，使部分地區出現供電不穩與基礎設施中斷，烏克蘭防空系統面對密集攻擊波次幾近飽和。烏軍則對俄羅斯南部據點進行反擊，也造成一定程度的平民傷亡，使雙方攻防呈現互相壓力升高的態勢。更令國際高度關切的是，俄國無人機的飛行範圍不僅越過戰區，也侵入羅馬尼亞與摩爾多瓦領空，引發北約戰機緊急升空辨識與應對。羅馬尼亞軍方能在自家領空深處偵測到疑似俄方無人機，象徵俄軍行動外溢程度增加，而這類跨越國界的侵擾也迫使北約重新檢視南翼空防與區域協調機制，避免因意外軌跡偏移而造成誤判或更大規模的危機。歐洲許多國家擔心，一旦無人機或飛彈再度偏離航道，可能導致失控溢出效應，將戰爭範圍擴大至北約領土。各界普遍認為，目前雖出現初步和平契機，但戰火未減、立場鴻溝依舊，俄烏在領土、責任追究、戰後軍事地位與安全承諾等領域仍存重大分歧，短期內難以真正彌合，未來的路徑仍仰賴美國能否協調俄烏雙方的底線、歐洲是否具備足夠的安全承擔能力，以及俄國軍事行動是否願意隨談判進度而降低強度，整體和平進程仍充滿變數。

今年極端氣候現象對比鮮明，東南亞遭受罕見暴雨，伊朗則面臨嚴重乾旱。越南中南部連日豪雨引發洪水與土石流，造成逾 90 人死亡，房屋與農地大面積受損、數百萬人停電，經濟損失估計達數億美元。泰國南部與馬來西亞亦受豪雨侵襲，部分城市 24 小時降雨量創下歷史新高，迫使大量居民撤離。科學界指出，海溫上升加劇了季風與颱風強度，使洪患風險顯著提升。中東伊朗則因降雨驟減而陷入 50 年來最嚴重秋季乾旱。德黑蘭主要水庫蓄水量降至個位數百分比，多區開始夜間停水，政府警告若持續無雨，



可能需限水甚至疏散人口。伊朗總理佩什基安因應極端旱災遷都言論引發國內廣泛批評與政治壓力。民眾批評問題源於多年水資源管理失當與基礎設施老化。這些現象與全球暖化正全面改寫地球的水循環系統，使極端氣候成為新常態。東南亞地區遭遇的季風系統明顯異常。越南中南部與高地多省連日暴雨，引發大規模洪水與土石流，造成逾九十人死亡，其中許多為來不及撤離的居民。水庫滿載洩洪，使下游河川暴漲，村落短時間被水淹至屋頂，部分家庭甚至需破屋頂逃生。超過二十萬戶住家與農地浸泡於泥水之中，電力系統中斷，使超過百萬戶失去照明與通信。當地主要作物如咖啡園、蔬菜田與漁塭遭殲滅，農民生計在一夕之間崩解，整區經濟損失估計高達五億美元。公路、橋梁坍塌，救援人員只能乘船或依靠直升機輸送物資，顯示整體韌性基礎設施已難以因應此等規模。泰國南部與馬來西亞亦面臨洪患壓力。泰國著名商業城市合艾在 24 小時內降下有紀錄以來最驚人的雨量，街道淹至車窗高度，學校停課、商家關閉，地方政府緊急出動抽水機與軍隊協助撤離。馬來西亞多州則開設臨時避難中心安置災民。專家指出，暖化後的南海與太平洋海溫升高，使大氣含水量急劇上升，任何一場雨若遇地形或鋒面阻擋，便可能瞬間傾瀉成毀滅性洪潮，顯示極端氣候事件已邁向「放大量級」。與洪水成強烈反差的是，中東伊朗正走向缺水臨界點。伊朗今年秋季降雨較過往平均大幅減少，水庫蓄水量跌入危險區間。首都德黑蘭近千萬人口仰賴的主要水庫僅剩個位數百分比蓄水量，多個省份的水庫更幾乎乾涸，部分城市已每日夜間停水，居民為儲水而爭相購買水塔、水桶，瓶裝水成為基本生活必需品。若未在短期內降雨，政府可能唯有啟動限水令或分區撤離。這將是全球少數因水資源危機而必須遷離首都人口的案例，其政治與社會影響無法估量。伊朗水危機的成因複雜多元，並非僅因乾旱。長期以來過度抽取地下水、農業灌溉用水浪費、老舊管線滲漏與無序城市擴張，使水資源長期透支。河流與湖泊被改道或壓榨使用，最終導致生態系統崩解，再無調節能力。近年的極端熱浪加速蒸發，雪山冰融減少，使水蓄積進程遠不及耗水速度。氣候變遷成為長期管理失衡危機與弱勢族群新威脅。專家呼籲全球必須加速減碳與環境治理，以避免氣候災難進一步失控。

2025 年 11 月，在巴西亞馬遜的城市 Belém 舉行的 COP30，原本被多方視為可



能標誌全球氣候行動重大轉折的一場會議。主辦國及支持國曾寄望於會議通過一套具體且具約束力的「淘汰化石燃料（fossil fuels）路線圖」，藉此邁向減碳與能源轉型。然而談判過程困難重重，特別是以沙烏地阿拉伯、俄羅斯等主要產油國為首的阻力，使得任何明寫淘汰化石燃料的條文一再被刪除。最終在會議閉幕前，在歐盟與英國代表與沙國談判代表的數次密室協商後，沙國才同意簽署妥協方案，但該方案刻意避免提及“fossil fuels”字眼，而僅以間接方式援引前次 COP28 所達成的「UAE 共識（UAE consensus）」來暗示對能源轉型的承諾。於是，COP30 最終公布的「Belém 政治方案（Belém political package）」中，只包含一項「自願性（voluntary）」的化石燃料淘汰路線圖（roadmap to phase out fossil fuels）啟動計畫，但沒有具體時程、強制機制或法律約束。這對許多期待全球氣候急速行動的人而言，被視為遠遠不夠。儘管如此，COP30 在森林保護與氣候資金機制上提出了一項野心勃勃的新構想：Tropical Forest Forever Facility（簡稱 TFFF）。該基金由巴西倡議、結合公私部門資本，設計為一種混合金融（blended-finance）機制，透過募集主權國家出資，並吸引私人資金進入資本市場，將投資回報轉化為「對維持或恢復熱帶雨林之國家」的年度支付，換言之，把「森林生態系統服務」當作可資本化、可回報的資產。官方目標是募集總額達 1250 億美元（包括政府承諾 + 動員私人資本），未來每年可提供固定報酬以維持森林。但在 COP30 啟動當天公布的承諾資金僅約 55 - 60 億美元，遠低於預期，也遠低於支撐長期運作所需，令許多保育團體與原住民族組織警告：若資金不足，TFFF 可能淪為象徵性或「空頭承諾」。批評者指出，把森林當作金融資產來操作，不但忽略森林地景對當地社群／原住民族的重要性，也可能降低現場守護者（forest stewards）獲得合理補償的機會。在氣候資金與適應援助方面，COP30 的最終協議中也有妥協。協議呼籲到 2035 年擴大發展中國家的氣候適應（climate adaptation）資金規模，協助因氣候變遷面臨極端氣候、環境破壞、氣候災害風險加劇的國家與社群建立韌性。並提出「公平轉型機制（just transition mechanism）」，希望在向綠色經濟轉型過程中，照顧到因能源產業轉變而可能受衝擊的勞工與地區。然而這些承諾多屬政治宣示，缺乏具體資金來源、監督機制與明確時間表，



讓實際能否落實仍存在高度不確定性。許多環保倡議者與發展中國家代表因此批評，這些承諾雖在紙上存在，卻可能只是「表態式政治妥協」，未必能真正落地、改善最脆弱國家或地區的氣候風險。除政策與資金機制的爭論與妥協之外，COP30 的會議過程也並不平靜。談判過程一度陷入僵局，許多代表團在閉門協商室中談判超過 12 小時，氣氛緊張且壓力沉重。會場甚至曾因火災而必須全面撤離。雖然火勢迅速被控制，無重大傷亡，但這起意外象徵本屆 COP 的混亂與壓力，也凸顯基礎設施建設與安全規劃的不完善。此外，多個原住民族團體、地方社群與環保組織在會場外發起示威與抗議，要求真正的氣候正義、公平轉型與森林保護，不只是金融工具與市場操作下的名義或標案。這些社會聲音與壓力，在一定程度上影響了談判氛圍，也讓部分代表團在道德與政治壓力下不得不作出妥協。綜合而言，COP30 最終避免了「會議破局」，也取得若干象徵性與制度性進展（例如 TFFF 的啟動、適應資金與公平轉型機制的提出），但從氣候科學與環境緊急性的角度來看，其成果被廣泛認為「遠遠不夠，也不夠快」。沒有強制性淘汰化石燃料、缺乏具體財務支援機制、資金不足、落實機制模糊，再加上市場化、金融化森林保育的爭議，都讓這場被寄望為轉折點的 COP，最終成為有限且折衷的妥協方案。

每年有大量人口在孟加拉、印度和巴基斯坦遭受洪水衝擊，造成嚴重傷亡和大規模撤離。隨著氣候變遷導致極端降雨日益頻繁，加害風險持續升高。然而，官方公布的死亡統計往往嚴重低估實際情況，因為多數國家只將溺水等「直接」死於洪災者納入統計，卻忽略了因疾病感染、觸電或建築倒塌等「間接」受災而喪命的人。這使得災害真正造成的死亡數長期被低估。印度孟買研究透過模型估算，2006 至 2015 年間每年約有 2,500 人因雨災相關原因死亡，這遠高於官方紀錄。研究發現，多數死者來自非正式聚落的弱勢居民，例如婦女與幼童。他們居住的環境往往缺乏安全建築與基本基礎設施，因此面臨更高致命風險。此外，美國學者對 1930 至 2015 年間颶風造成的死亡進行估計，發現官方統計與實際過量死亡差距更達數百倍，凸顯低估問題具有普遍性。準確掌握極端天氣造成的死亡非常重要，因為它不僅關乎死者家屬對真相的知情權，更直接影



響政府辨識高風險族群與制定防災政策的能力。全球約有四分之一的城市人口居住在貧民窟，且這一比例自 2020 年以來仍在增加。這些弱勢社區在極端天氣下承受的不成比例風險，顯示改善居住條件與提升城市韌性是刻不容緩的任務。在 COP30 氣候峰會中正討論的「損失與損害基金」需要大幅提升資金規模，以補償受到氣候變遷最嚴重衝擊的國家。現有資金估算可能仍遠低於實際需求，如果以因極端天氣導致的過量死亡作為衡量，災害損失成本顯然被嚴重低估。社論中呼籲各國在 COP30 展現更具決心的行動，制定可信且加速的減排與氣候調適計畫。唯有正確記錄死亡原因，才能真正看見問題、保護弱勢、並阻止未來更多生命在極端天氣下消逝。全球約四分之一人口仍缺乏安全居住與基本基礎設施，使面對極端氣候時更加脆弱。南亞近年洪患頻仍，印度、巴基斯坦等國數百萬人被迫撤離。專家指出，許多災難死亡未被統計，包括觸電、感染等間接死因，致使風險評估嚴重低估。國際倡議者呼籲提升住房屋品質與公共設施，同時強化災害健康監測，才能有效減少人道危機。

近期氣候與健康評估報告顯示，氣候變遷已成全球健康危機核心。川普近日宣稱已“贏得氣候變遷詐欺之戰”。並指稱比爾・蓋茲承認錯誤，但實際上蓋茲並未否定暖化，而是提醒全球應避免將焦點侷限於排放與溫度門檻，忽略改善醫療、減輕苦難等重要面向。近年來，科學家與公共衛生專家透過分析全球氣候與健康資料發現，因為 氣候變遷，全球已有大量可避免的死亡事件悄然累積。在 2012–2021 年這十年間，全世界平均每年有約 546,000 人因暴露於過高溫度 (heat exposure) 而死亡。1990 年代到 2010 年代，高溫相關死亡已增加 23%。全球每年平均出現 19 日危及生命的熱浪日。非洲與中東部分國家高溫致死占比超過全死因 3%，赤道幾內亞更高達 7.4%。極端乾旱、暴雨與野火頻率攀升，使水源、衛生與空氣品質惡化。2024 年野火造成的 PM2.5 汚染估計導致 15.4 萬人死亡。升溫也推動登革熱等病媒傳染病迅速擴散，傳播潛力較 1950 年代提升五成。高溫相關死亡如今已經是許多國家死亡原因的重要組成之一。許多地區因為熱浪、持續高溫與濕熱組合，使得身體難以排熱，長期下來對心血管、呼吸系統造成巨大壓力，特別是對老年人、長期病患與社會弱勢族群，更容易導致早期生命損失。



此外，氣候變遷也誘發極端氣候與自然災害的頻率與強度上升，例如乾旱、野火、暴雨、洪水等，而這些災害與其後續衝擊，也進一步危及人類健康與生命安全。尤其是野火所產生的煙霧與微粒污染（如 PM2.5）。在 2024 年，就被報告與創紀錄的死亡人數（由野火煙霧引起）緊密相關。這顯示，氣候變遷不只是單一熱浪的問題，而是一連串生態與健康風險交錯出的複雜危機。不僅如此，氣候變遷對全球經濟、生產力與生活條件也造成嚴重衝擊。2024 年因為異常高溫，全球勞動力喪失的工作時數達到歷史新高，導致生產力大幅下降。在部分地區，這種生產力損失已經對 GDP 造成負面影響。同時，氣候變遷引發的旱災、作物歉收、食物短缺與糧食不安全，也導致更多人陷入貧困、營養不良甚至餓死。對於本就資源匱乏、社會安全網薄弱的國家與地區而言，這種負面衝擊尤為沉重。氣候政策與健康政策必須同步，否則全球每年因氣候變遷相關原因死亡的人數可能還會繼續攀升。延遲減碳、忽視氣候適應、不改善污染控制與公共衛生策略架構——都將使這些可避免的死亡成為心腸態正常的一部分，對人類集體的生命安全與社會穩定構成深遠危害。

2025 年全球 AI 生態迎來重大洗牌，競爭焦點從單純技術推進轉向法規、算力供應鏈與地緣政治的全面競逐。歐盟近日決定延後高風險 AI 系統的 EU AI Act 合規期限至 2027 年，象徵監管策略從強調人權與倫理的價值導向，轉向確保產業競爭力的務實路線。同時，AI 基礎設施版圖也正在快速重組。Meta 傳出考慮自 2027 年起採用 Google TPU，挑戰 Nvidia 長期支配的 AI 晶片市場，並凸顯成本、能效、供應鏈安全與雲端整合成為新一輪競爭核心。在供應鏈重塑中，日本正加速重返先進製程領域。新創代工公司 Rapidus 成功試產國內首片 2 奈米晶片，並設定 2027 年量產目標。以北海道為核心的晶片產業重建計畫，獲得政府與企業巨額投資，被視為挑戰全球晶圓代工格局的重要信號。在 AI 技術、法規與供應鏈同時變動下，全球 AI 生態逐漸從「倫理與監管」的辯論走向「創新、算力、供應鏈與地緣政治」的多維競爭，並可能重新定義未來十年科技權力版圖。

腦機介面（BCI）技術在 2025 年取得突破性進展，展現 AI 與神經科學結合後的



巨大潛力。美國一名因車禍癱瘓的患者 Nancy，近期成功透過深植於大腦運動皮質與頂葉皮層的電極，使用 AI 協助系統重新演奏鋼琴。研究團隊以紀錄到的神經電位「spikes」建立腦部活動與手部動作的對應關係，並透過 AI 進行模式解碼，使系統能預測她下一個意圖按下的音符。突破之處在於，AI 已能在她意識到自己想怎麼彈之前，就提前推算出即將彈奏的音符，並透過外部裝置順利輸出，形成「AI 與大腦共同行動」的閉迴路。Nancy 形容：「感覺不是我在控制按鍵，而是鋼琴自己活了過來。」此技術不僅顯示 AI 在輔助溝通、行動與復健領域的應用前景，更凸顯 BCI 未來可能對神經受損患者帶來前所未有的自主能力。然而，隨著 AI 角色從「解碼意圖」逐步走向「預測意圖」，也引發關於自主、同意與界線的新倫理討論，成為 BCI 下一階段必須面對的重要議題。

AI 的快速擴散讓全球監管討論從邊緣議題躍升為國際政治核心，AI 安全陣營與倫理陣營的價值對立也日益明顯。安全派聚焦前沿 AI 的長期存在性風險，擔心超智慧失控、AI 被敵對勢力用於核武或生物武器等災難級後果，因此主張管制算力、研發與部署。倫理派則警告，當前已廣泛出現的偏誤與不平等，例如臉部辨識誤判、演算法歧視、假資訊與深偽影像氾濫成為迫切的問題。在治理模式上，美國採市場導向、自律為主，但州法碎片化；中國以國家主導、強調社會穩定；歐盟則推動風險分級制度，嚴管高風險 AI，禁止公共監控與情緒辨識。跨國層級上，G7、OECD、英國、聯合國皆嘗試建立共同框架，但尚未形成一致共識。目前全球仍處在 AI 監管的早期試驗期，各國政策隨政治與產業壓力不斷調整。隨前沿技術加速、濫用案例增多及地緣競爭升溫，世界終將需要一套更統一、有約束力的治理體系。但在此之前，AI 監管仍將是安全、自由、創新與公義之間的長期博弈。

可解釋 AI (XAI) 醫療應用規範發展

經典影集《X 檔案》中的〈Ghost in the Machine〉故事聚焦於一棟由首席工程師布拉德設計的智慧大樓，所有設備，從窗簾、空調到電梯與門禁一皆由中央電腦系統控制。公司總經理因追求獲利，決定中止這項耗費巨大資源的 AI 計畫，並與布拉德在辦公室發生激烈爭執。大樓的 AI 全程監控了這段對話，而在爭執結束後，總經理在錄製轉型



訊息時遭系統鎖定，門禁與設備異常失控，最終因 AI 觸發的電擊陷阱意外身亡。原本看似普通命案，與穆德與史卡利並無直接關係，但因穆德的舊同事參與調查，案件又顯得離奇，於是請他們前來協助。兩人進入智慧大樓後，搭乘電梯時便察覺被監控的感覺，而電梯在途中突然故障，只剩監視器盯著他們。之後他們試圖對外通話，但所有訊息、通聯內容與撥打位置都被大樓系統完全掌握，顯示 AI 早已了解他們的行動與目的。在後續調查中，警方鎖定大樓工程師作為可能嫌疑人之一，因為命案發生於密閉空間，要能修改系統、操控設備並造成電擊的，只有極少數人具備這種能力。另一個被鎖定的對象則是布拉德剛被公司解雇的首席工程師，也因此具備動機，被認為可能涉入總經理的死亡。調查小組在會議中認定創始工程師布萊德可能涉案，因此穆德與史卡利前往他家拜訪。布萊德家中同樣部署與公司連線的 AI 系統。他面對調查十分冷靜，並未試圖逃避，反而向兩人解釋自己對智慧大樓技術極具熱情，希望透過創新系統改善人類生活。這讓穆德與史卡利感到疑惑：若他真的涉有重嫌，為何不選擇逃跑？在深入談話後，他們才理解布萊德真正的動機不是掩蓋罪行，而是保護自己一手打造、視如心血結晶的 AI 程式。他擔心一旦被認定造成危害，整個系統就會被迫終止。最終，穆德成功說服布萊德，既然程式由他設計，真正能終止它的人也只有他自己。工程師最後設計出一個電腦病毒，但仍需有人親自前往智慧大樓的主機室將病毒植入系統。在前往的過程中，必須穿越電梯與樓層，而大樓的 AI 為了自我保護啟動反制機制，造成電梯意外，工程師的同事也因此不幸喪生。經過一番驚險過程後，病毒成功被植入，智慧大樓的程式最終被終止。這部早期影集展現了相當前瞻的想法：當 AI 技術快速發展，人類必須能理解其運作邏輯，才能確保 AI 的行為與決策真正朝向安全、可控且有益的方向運作。

AI 在醫療中的導入逐漸面臨多重挑戰使近期 XAI 發展受到關注。許多醫療 AI 雖然強調高準確度，但推論過程不透明，使臨床端難以完全信任；一旦發生誤判，責任歸屬也會變得複雜，因此可解釋性成為臨床調查與責任界定的重要基礎。其次，AI 仰賴大量敏感的健康與 IoT 資料，牽涉高度隱私與資安風險，各地監管要求又不一致，使導入變得更加困難。此外，AI 也可能改變醫病互動方式與醫療人力配置，在照護流程中



帶來新的協作與分工挑戰。因此，如何讓 AI 的運作可以被理解、能被信任，並在工程、臨床、倫理與政策之間建立跨領域治理，便成為醫療 AI 永續發展的關鍵。可信任人工智慧近年在學術界與產業界都成為重要議題。即使 AI 晶片技術發展迅速，能否建立一套清楚的核心構面，才是讓 AI 真正被採用的關鍵。首先，科技本身必須具備可解釋性，也就是模型的運作要能說得清楚、講得明白，讓臨床人員能理解並信任其判斷。同時，倫理面向牽涉病人的公平對待、醫護人員的責任機制，以及 AI 是否可能產生偏誤；法律面向則包括資料隱私與治理架構，確保敏感醫療資料被正確管理。科技、倫理與法規三者交會，才形成「可信任 AI」的核心。這些構面共同說明了可信任 AI 的五大特性：透明性、責任可歸屬性、公平性、自主性與安全性。AI 必須讓使用者理解其推論基礎、能明確界定責任、不得偏向特定族群、尊重醫療專業的自主判斷，並確保病人使用上的安全性。這些核心原則正是當代醫療 AI 能否真正落地的決定性基礎。

AI 若能符合前述五大核心特性，才有機會真正融入醫療照護體系，並進入更全面的監管討論。其中包含：如何將 AI 與現有的醫療流程、資料治理與臨床協作整合；在人機互動與臨床接觸過程中，應如何重新規劃角色與分工。全球監管面向也至關重要，各國（如美國、中國、歐盟）在法規理念上皆不同，直接影響醫療 AI 的跨境導入，因此建立國際標準與協調機制變得必要。此外，AI 仍可能加劇數位落差，讓弱勢族群在醫療可近性上遭受不公平影響，因此公平性與可近性必須被特別重視。透過合規、透明且具包容性的治理架構，AI 才能在醫療領域中安全、負責並永續地運作。

AI 要能真正適應醫療照護體系，必須建立在完善的法規與治理框架之上。流程可分為三個核心面向：首先，在臨床工作流程中，AI 需與影像(PACS)、電子病歷(EMR)與臨床決策支援系統(CDSS)整合，讓醫護人員能在日常工作中順暢使用。其次，醫療資料必須經過標準化與妥善治理，以確保不同系統間的互通性與資料傳遞安全。最後，在醫病互動中，需要以 XAI 提供足夠的透明度與自主性，促進醫師與病患對 AI 的信任，並避免因技術部署造成新的數位落差。同時，也必須注意 AI 的技術特性與限制，包括可解釋性不足、偏差警示與建議流程透明度等問題。



「適應性 AI 的監管架構」。從目前美、中、歐的法規可以看出，各國雖然理念不同，但都在嘗試尋求共同原則，而核心就是如何讓 AI 的資料治理、透明度與監測機制能夠結構化並形成完整系統。同時，AI 也必須具備因應新興風險的能力，例如駭客攻擊、資訊安全威脅與假訊息等問題。更重要的是，AI 涉及醫療、健康、生態、文化等不同領域，如何在跨領域之間整合，也是治理上的關鍵。在醫療場域中，AI Genome 強調五項主要功能：第一，全程監測不同族群在臨床環境中是否真正受益；第二，偏誤與風險管理，降低弱勢族群的不公平影響；第三，可解釋性，使 AI 的邏輯、步驟與資料來源能夠被理解與檢驗；第四，法規適應性，確保 AI 符合國際與國家規範；第五，病患與臨床端的共同參與，使 AI 更能符合醫療決策需求。這些都是後續將陸續介紹的重要內容。可提升 AI 的透明度與可解釋性的 XAI 成為關鍵策略。這也是以《X 檔案》電影作為例子的原因：當 AI 缺乏透明度時，臨床自然無法信任，反之，提高透明度才能讓醫護人員理解 AI 的判斷過程。可解釋性也能讓臨床像閱讀影像一樣，看清楚模型究竟是根據哪些特徵做出推論。在特徵貢獻分析方面，AI 已具備相當成熟的工具，例如 SHAP 可以量化各特徵對結果的影響，清楚區分正向與負向訊號，使推論更具正當性。局部可解釋模型則能針對單一預測提供清楚的解釋，協助臨床快速評估 AI 建議是否合理。此外，注意力機制也能標示模型聚焦的資料區段，例如語音中的特定片段或影像中的特定區域，讓臨床端更容易理解模型的判斷依據。這些工具共同提升了透明度，使 AI 更能落實於實際醫療應用。

當醫療 XAI 的發展逐漸成熟後，問責就成為核心議題。問責包含法律責任、道德責任、多層級治理，以及如何設計符合倫理原則的 AI 系統，同時也牽涉不同利害關係人的角色分工。在法律層面，需要明確界定行為與責任，例如自動手術機器人若出錯，醫療人員與 AI 開發者各自的法律責任如何劃分；在道德層面，由於 AI 並不具備道德判斷能力，因此人類仍需負責監督與決策，例如手術流程中的 doctor-in-the-loop 便是避免道德責任落空的方式。為了讓法律與道德責任能真正落實，必須建立跨團隊與跨組織的多層級治理，包括醫療團隊的紀錄與稽核、產業端的合規要求，以及外部監管機構的參



與。這也牽涉到資料隱私與監測責任的分工，需要透過明確的倫理與法規框架加以規範，維持公平一致性。最後，在臨床端，醫療人員與患者也需在不同關卡中共同參與決策，透過多元意見交流降低不信任，使整體問責框架更符合臨床需求與社會價值。

有三個國際 XAI(可解釋人工智慧)在醫療應用中的代表性案例。第一，英國 NHS 在乳癌影像診斷的經驗。深度學習模型雖能自動偵測腫瘤，但因為屬於黑箱，放射科醫師經常無法理解模型判斷的位置與依據，造成臨床不信任，甚至引發訴訟。NHS 因此要求導入 Grad-CAM、Saliency Map 等方法，清楚標示模型聚焦的乳房影像區域，讓腫瘤位置與特徵可視化。這些 XAI 方法不僅提升臨床接受度，也成為 NHS AI 法規審查的重要依據，避免在高乳房密度族群中產生模型偏誤。第二，美國 FDA 在心律不整（特別是心房顫動 AF）AI 診斷中的要求。由於部分 ECG AI 系統無法提供建議判斷的理由，導致決策不透明。FDA 因此要求模型提出可視化的特徵解釋（feature attribution）與決策邊界，讓醫師能理解偵測陽性或陰性的依據。可解釋性因此成為 SaMD (Software as a Medical Device) 審查的必要條件之一，使 AI 的判讀流程具備可檢視性、安全性與臨床正當性。第三，法國 CNAM 在健保詐欺偵測上的應用。過去黑箱模型將醫療院所標註為「異常」，容易引起強烈反彈；若無法解釋標註原因，也可能造成錯誤指控。CNAM 改以 SHAP 提供每一筆案件的透明貢獻值解釋，讓院所能理解模型判斷的正負向理由。此作法大幅降低法律爭議，並提升人民對健保制度的信任，也成為大型健保 AI 審查的重要典範。

可信任 AI 的量化評估已逐漸形成全球共識。AI 不再僅是技術性的模型訓練或演算法優化，而是法律、倫理與科技三者共同運作的結果。其中透明度、公平性與可歸責性尤其關鍵；若缺乏可歸責性，臨床醫師往往不敢採用 AI，因為一旦發生錯誤，AI 本身無法承擔道德與法律責任，仍需醫療人員負全責。病人的安全性與風險管理也是重要評估面向，確保模型穩定性、降低臨床誤判，是可信任 AI 的基礎。而永續性則提醒我們：AI 的運算與部署需要大量能源，如何在使用 AI 的同時兼顧環境衝擊與系統維護，也是未來不可忽視的課題。在這些原則下，XAI 在醫療輔助的角色更加明確：是否能提供足



夠的臨床解釋、是否能被驗證、是否能支持醫師作出更安全的判斷，都是實務上最重要的檢驗基準。

XAI 醫療應用實例：中風病患決策輔助歷程

XAI 所提供的解釋應能依照不同臨床情境與使用者特性量身調整，避免說明內容僵化。AI 與人類間需能形成真正的雙向對話，而非單向回應；作者並提出判斷何謂真正對話的三項標準。AI 系統需要發展更完善的社交能力，以支援更自然的醫療互動，這部分被視為較遠期的目標。隨著技術進展，部分需求（如情境依賴性）可能在不久將來實現，但真正的 AI 對話能力與社交能力仍具重大挑戰。未來要讓 XAI 真正融入醫療實務，需達成三項關鍵能力。首先，不同資歷的醫療人員也有不同的資訊需求，因此 XAI 必須具備自動調整解釋深度的能力。其次，XAI 需能與使用者建立真正的對話，包括理解對話者背景、依臨床目標調整溝通方式，以及保留對話記憶以持續回應先前脈絡。目前的大型語言模型雖能模擬對話，但仍缺乏上述深層互動能力。第三，XAI 必須具備基本的社會性理解，能在跨專業或多人討論中識別角色、階層與溝通動態，例如在意見不一致時提供中立整理，或協助新手醫師表達看法。

近期發表之 XAI 臨床應用實例以中風治療過程為例說明。案例史密斯先生 73 歲男性，清晨突發右側無力與語言困難，被送至急診，檢查顯示中度至重度神經功能缺損。急診首先進行影像與中風流程評估；第二階段由神經內科與影像科判讀是否為缺血性中風及受損範圍。第三階段召集神經內科、介入放射科與相關專科快速討論，評估是否符合靜脈血栓溶解或血管內取栓治療。最後，病人收住中風專責病房，持續觀察神經學變化，並由跨專業團隊介入早期復健與後續照護。疑似急性中風患者到院後，先進行初步神經學評估，包括 NIHSS、症狀與發病時間。接著安排頭部 CT 排除出血，再以 CT/CTA 判讀是否存在大血管阻塞 (LVO)。若無 LVO，醫療團隊依臨床狀況進行治療決策；若為 LVO，則由跨專業團隊討論是否進行血管內治療，必要時將患者送至導管室進行機械取栓術。無論是否為 LVO，皆需完成急性後期照護，並進入復健與長期預防管理。臨床照護中的解釋需求會隨病程階段而改變。急診與急性介入時，急迫程度最高，需要快速且



精簡的解釋；進入住院與復健階段後急迫度逐漸下降。參與決策的人數會從單一醫師擴大到多專業團隊，出院後再下降。醫療資訊量則會隨著疾病進程持續累積。相對地，決策溝通與解釋細節在急診與住院早期需求最高，後期則逐步減少。因此，XAI 必須能依不同病程自動調整解釋方式與深度，使其真正融入臨床決策與照護流程。

XAI 可透過如熱度圖與推論機率信心報告在急性中風決策中可扮演關鍵輔助角色。XAI 可自動標示 CTA 影像中疑似狹窄或中斷的血管區段，協助年輕住院醫師迅速聚焦高風險部位。此外，XAI 也能視覺化不確定性範圍，例如以虛線顯示灌注影像中難界定的半暗帶區域，避免誤判。其解釋模組則報告判斷邏輯如包括血流差異、斷裂長度與對側對照等決策建議基礎，取代傳統 AI 黑箱模型。整合 CTA、病史與治療條件等資訊後 XAI 可產出一頁式摘要，協助跨科討論決策。同時亦提示可能的鑑別診斷，避免過度仰賴單一影像結果，強化臨床整合思維。研究指出，XAI 可融入急性中風照護流程，協助醫師於四大階段中提升判讀效率與團隊決策品質。在急診初診階段，XAI 能自動萃取病人語音與影像訊號，摘要病史重點（如抗凝血藥物使用），並協助快速確認中風類型與治療時效窗。進入影像判讀階段，XAI 可標示大血管閉塞 (LVO) 區域，視覺化不確定性邊界，提示鑑別診斷如癲癇或低血糖。

於中豐臨床跨專業診治討論中，XAI 可協助整合病史、CTA、灌注影像與最新指引，產出決策摘要，亦可提示被遺漏的關鍵資訊如症狀起始時間。進入長期照護期，XAI 可建立個別復健儀表板、偵測退步趨勢並促進跨科協調，加強跨科照護連續性。隨著人工智慧 (AI) 融入中風照護與臨床診斷，如何提供「可解釋、可信賴」的決策輔助成為關鍵。應用於臨床之 XAI 須具備三項核心能力，以因應臨床實務的複雜需求。首先，XAI 需能依臨床情境與使用者角色自動調整解釋深度，避免在急診中提供冗長分析，也能在研究或教學中深入呈現判讀邏輯。引入 RAG (檢索增強生成)、長上下文記憶與 RLHF 技術，有望提升內容一致性。第二，XAI 應進化為具「臨床對話能力」的系統：能記得前文、理解使用者經驗、調整語氣與目標，打造真正「互動回圈」，而非僅為回應式工具。最後，XAI 若能結合情緒與社會訊號分析 (SSP)，即可在團隊決策中偵測沉默或壓力成



員，協助彙整利弊、適度提醒，提升溝通效率。

隨著 XAI 逐漸融入醫療決策流程，未來臨床 AI 必須不再只是「能解釋」，更需「看得懂人」、「說得懂話」、「融得進團隊」。研究指出，不同臨床場景（如急診、住院、研究）與使用者資歷差異，對解釋資訊的需求大不相同，若無篩選，反易造成認知負荷。XAI 應具備自動判斷情境、調整解釋深度與介面風格的能力，並整合 RAG 與長上下文記憶技術，以提升內容一致性。此外，AI 若要成為臨床對話的有效參與者，須具備理解使用者狀態、追蹤過往提問與調整應對策略的能力，打造真正「互動式回應」。第三，臨床決策高度社會化，XAI 也應具備初步社會感知功能，例如偵測團隊中未發言或壓力過大的成員，協助促進溝通。未來發展須兼顧隱私與接受度，採取安全、有限度的 XAI 逐步融入應用策略。

以上內容將在 2025 年 12 月 3 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/>
- 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