



星球永續健康線上直播

可解釋 AI (XAI) 醫療應用

XAI 法律規範治理與臨床運用風險決策 (2)

2025 年 12 月 31 日

可解釋人工智慧 (Explainable AI, XAI) 為導入人工智慧科技創新連結臨床專業領域與法律治理的重要橋樑。它不僅讓傳統黑箱 AI 模型能夠呈現模型特徵以及提供使用者建議與判讀機轉及原因，更是確保醫療決策得以被理解、被檢視，並能追溯責任的關鍵基礎。本週我們將聚焦於人工智慧 XAI 的法規治理，以及 XAI 在照護場域中的實際應用案例。

星球健康新知

北極主權美國與歐洲盟國角力：「觸及紅線」

美國總統川普再度公開表示美國「必須擁有格陵蘭」，並任命路易斯安那州州長傑夫·蘭德里為格陵蘭事務特別特使之後，圍繞格陵蘭主權與北極安全的爭議迅速升高。川普多次強調，美國對格陵蘭的需求完全出於國家安全考量，而非礦產或能源利益，並宣稱在格陵蘭周邊海域可見俄羅斯與中國船隻，認為丹麥未能對該自治領地提供足夠的防衛與投入，因此美方有必要介入甚至取得控制權。他也拒絕排除使用強硬手段的可能性，使得相關言論不再僅被視為政治修辭，而被歐洲各國視為具有實質威脅性的戰略表態。這一立場在丹麥與格陵蘭內部引發強烈反彈，丹麥首相與格陵蘭總理一致重申，國界與主權根基於國際法，任何國家都無權吞併他國領土，並明確指出格陵蘭屬於格陵蘭人民，其未來只能由當地民主程序決定。丹麥外交部也對特使任命表達強烈不滿，認為相關言論已嚴重損害丹麥主權，並著手召見美國外交代表要求說明，同時宣布大幅提升在格陵蘭的軍事與防務投資。歐洲國家回應迅速且一致。多國領袖與歐盟高層公開表態支持丹麥，強調領土完整與國家主權是國際法不可動搖的基本原則，即便在北極安全與盟友合作的脈絡下，也不能被模糊或犧牲。歐盟執委會主席指出，北極安全確實是歐盟的重要優先事項，但這必須建立在尊重主權與法治的前提之上。法國、西班牙、瑞典、



挪威等國亦相繼發聲，表明在此議題上與丹麥站在同一陣線。相關討論也觸及北約內部的潛在矛盾，因美國與丹麥同為成員國，若出現對盟友領土的壓力或威脅，將動搖集體防衛體系的政治基礎。報導同時指出，儘管格陵蘭社會普遍支持在未來條件成熟時走向獨立，但民調顯示絕大多數居民明確反對成為美國的一部分，凸顯川普構想與當地民意之間的落差。這一連串事件不僅讓格陵蘭議題重新成為跨大西洋關係的敏感焦點，也反映出在氣候變遷、航道開放與大國競逐背景下，北極地緣政治正持續升溫，並考驗既有國際秩序的穩定性。

2026 人工智慧熱潮全球市場與資源布局：「以能制智」

2025 年為具有轉折意義關鍵年，人工智慧的快速擴張不僅改變科技產業本身，也深刻牽動能源體系、原物料市場與整體經濟結構。隨著大型語言模型與雲端運算需求暴增，資料中心的建設規模出現前所未有的擴張，相關投資金額被估計將累積至數兆美元等級，而電力與水資源的壓力已成為限制 AI 成長的實質瓶頸。為了確保長期且穩定的電力供應，科技巨頭開始從單純的能源採購者轉為直接掌握能源生產與基礎設施，透過併購大型再生能源與儲能開發商，將發電、儲能與資料中心整合於同一體系之下，藉此縮短建設時程、避開電網壅塞，並降低未來營運成本與碳排放風險。這種垂直整合的策略不僅反映出 AI 對能源的高度依賴，也顯示傳統公共電網已難以承受 AI 訓練與推論所帶來的負載成長。在資本市場層面，AI 熱潮被不少觀察者形容為可能正在形成泡沫。美國股市的估值指標在 2025 年攀升至歷史罕見的高位，過往經驗顯示，一旦估值長期脫離企業實際獲利能力，市場往往難以避免劇烈修正。儘管 AI 的確是史上擴散速度最快的技術之一，企業與個人使用率在短時間內大幅提升，但支撐這波投資浪潮所需的長期現金流與生產力提升，仍存在高度不確定性。大量資金湧入資料中心、晶片、能源與機器人領域，使得投資人對未來回報的期待被推升到極高水準，這種期待一旦落空，可能引發類似過去網路泡沫的調整。不過，也有觀點指出，即使資本市場出現修正，AI 對社會結構與勞動市場的衝擊仍將持續，特別是人形機器人與自動化技術的成熟，可能在未來數年內大幅取代人力工作，帶來稅制、社會福利與生活意義層面的深層挑戰。



在實體經濟與原物料市場中，AI 與能源轉型的同步推進，正將金屬推向新的戰略核心地位。銅、鋁、鋼鐵與鋰等工業金屬成為資料中心、電網擴建、電動車與儲能系統不可或缺的基础材料，其中銅因為廣泛用於電力傳輸、冷卻系統與晶片內部連接，被形容為能源轉型與 AI 基礎建設的「中樞神經」。在需求急遽上升的同時，全球供給卻受到礦場事故、極端氣候、地緣政治與貿易政策干擾，使市場面臨結構性短缺風險。分析指出，若新礦產開發與冶煉產能無法加速，未來幾年可能出現高達三成的供需缺口，進一步推升價格並壓縮下游產業利潤。這種供需失衡也促使各國試圖降低對單一國家的依賴，重新布局礦產與冶煉產能，以確保能源與科技發展所需的物質基礎。同樣的趨勢也反映在整體金屬市場表現上。2025 年不僅貴金屬創下新高，工業金屬價格亦大幅上漲，背後驅動力並非避險需求，而是 AI 建設與能源轉型對實體材料的長期消耗。市場分析普遍認為，全球經濟正從以化石燃料為核心的結構，轉向高度依賴金屬與電力的科技體系，而現有供應能力難以在短期內完全跟上這種轉變。對投資人與政策制定者而言，未來經濟的風險不只為金融市場的波動，也包含於能源、原物料與基礎設施的長期配置是否足以支撐 AI 所描繪的成長藍圖。

無人計程車發展契機：「遲來必爭」

自動駕駛計程車產業經過三十年技術積累，正式邁入商業化階段，目前呈現美中兩強競爭格局。在美國，Waymo 以約 2,500 輛自駕車在五座城市營運，月活躍用戶突破百萬，穩居領先地位；Tesla 則快速擴張至 1,655 輛車，但仍需配置安全監督員。中國方面，百度 Apollo Go 部署超過千輛自駕車並計劃於 2027 年擴至兩萬輛，Pony.ai 和 WeRide 等企業也積極布局，目前已有超過 50 座城市允許測試、至少 10 座城市展開商業營運。中國在這場競賽中具備顯著的結構性優勢。政府大力支持自動駕駛發展，地方政府積極建設智慧交通基礎設施。成本差距尤為突出：Waymo 單車成本高達 13 至 20 萬美元，中國 robotaxi 平均僅 4 萬美元，百度 RT6 更低至 3.5 萬美元。中國企業主導的光達市場佔全球九成，關鍵零組件價格已從過去的 10 萬美元降至僅千餘美元。然而，所有 robotaxi 服務目前均處於虧損狀態。營運成本約每英里 7 至 9 美元，遠高



於傳統叫車服務的 2 至 3 美元，預計需十年才能降至具競爭力的水準。安全與公眾信任是另一挑戰，調查顯示約 75% 美國人對自駕計程車缺乏信任，但實際搭乘體驗可顯著提升信心達 56%。在市場布局方面，Uber 和 Lyft 積極與自駕技術公司合作，中國企業則拓展至歐洲、中東和亞洲市場。2026 年起，百度 Apollo Go 將透過與 Uber、Lyft 的合作進入英國市場。分析師預測到 2035 年，中國 robotaxi 市場可達 500 億美元，全球市場（不含美國）可能超過 2,100 億美元。這

人工智慧經濟時代轉型動能與風險治理：「利器雙刃」

人工智慧正以前所未有的速度滲透至經濟與社會運作的各個層面，一方面帶來顯著的潛在效益，包括加速科學研究、提升生產力、改善決策品質與風險管理能力，以及促進醫療照護的進步；另一方面，也逐步暴露出對金融體系與整體社會穩定構成的深層風險。隨著大型語言模型與高度複雜的機器學習系統被廣泛整合至既有平台與企業流程，其直覺化的使用介面與規模化部署，促成極快的擴散速度，但也使得錯誤難以察覺、偏誤可能被系統性放大，並增加使用者對模型輸出的過度信任。在金融體系中，人工智慧的運用可能透過多種機制放大既有的系統性風險來源，包括流動性錯配、共同曝險、高度連結性、缺乏可替代性以及槓桿與順循環行為。模型高度同質化與市場集中，可能形成新的單點失效風險；交易與反應速度的提升，則可能加劇市場波動並使負向動態更難以中止。此外，人工智慧系統的高度不透明性，不僅削弱監理與內控能力，也可能被惡意行為者利用於詐欺、操縱市場或散播錯誤資訊，而其對歷史資料的高度依賴，使其在面對極端或前所未見事件時更容易失靈，進而引發連鎖反應。這些風險的累積，使得單純依賴既有監理架構已不足以因應，亟需結合競爭政策、消費者保護措施與宏觀審慎監理工具的調整，包括提高透明度、強化責任歸屬、要求「風險自負」與相應能力門檻，以及提升監理機關在技術、人力與跨境合作上的能力，否則在人工智慧高度全球化的背景下，任何一國的監理落後都可能對其他司法管轄區造成外溢性衝擊。人工智慧也重塑創業活動的樣貌，特別是在發展中國家中，微型、小型與中型企業以及新創公司，已開始在行銷、客服、物流、財務與產品設計等多個環節導入以人工智慧為基礎的工具。大



型語言模型逐漸成為一種基礎性技術，使資源有限的企業能以相對低成本快速採用人工智慧，但實際應用仍呈現高度不均衡，關鍵障礙並非技術本身，而是對其商業價值的理解不足、管理能力與技術人才的缺口，以及缺乏可預期且一致的治理環境。若缺乏清楚的規則與支持性生態系，企業往往難以判斷人工智慧應如何融入長期策略，甚至可能因風險與不確定性而延遲投資。因此，建立清楚且協調一致的治理框架、降低技術與資金取得門檻、推動開放原始碼與協作平台，以及透過分階段導入策略逐步累積能力，成為促進實質採用的核心條件。更重要的是，技能被普遍視為人工智慧採用的決定性因素，不僅包括操作工具的實務能力，也涵蓋對「在哪裡使用人工智慧最具價值」的策略判斷，而針對這一缺口所提供的實作導向訓練、輔導與示範案例，被認為能有效協助企業將人工智慧轉化為可持續的競爭力。在邁向以人工智慧為核心的經濟體系過程中，資安風險亦同步升高，並呈現質變。隨著經濟活動從「AI 輔助」轉向「AI 原生」，具備自主推理、行動與記憶能力的人工智慧代理人將大量取代或補充人類勞動，形成機器與代理人數量遠超人類的工作環境。這種轉變雖可大幅提升效率，卻也使「身分」本身成為新的攻擊核心，深偽技術與身分仿冒可能在即時情境中誘發自動化系統連鎖行動，導致難以挽回的損失。同時，自主代理人若配置不當，可能成為具備高度權限的內部威脅，讓攻擊者透過提示注入或工具濫用，在極短時間內完成資料外洩、系統破壞或金融操作。再加上資料投毒、模型信任危機與量子運算對既有加密體系的壓力，使得傳統以周邊防禦為主的資安策略已不足應對。建立以資料與身分為核心的整合式治理與防護架構，結合即時可視性、執行階段防護與責任明確的治理機制為 2026 人工智慧轉型與產業發展關鍵。

AI 重塑全球經濟挑戰：「順勢而為」

經濟學家對 AI 影響的預測差異極大，從保守的未來十年 GDP 增長 0.9% 到樂觀的 2045 年為全球增加 17 至 26 兆美元產出不等，但這些信念已實際影響當今經濟，引導年輕人的職業選擇、政府政策方向以及對半導體和數據中心的巨額投資。然而，傳統的自然實驗和隨機對照試驗面臨兩大根本挑戰：研究結果快速過時，例如早期研究顯



示客服人員效率提升 15%、軟體開發者產出增加 26%，但 ChatGPT 發布以來 AI 能力已成長三倍；此外，受控研究無法捕捉更廣泛的連鎖效應，實驗中經驗較少的員工從 AI 輔助受益最多，但現實中管理者可能重組工作或直接取代這些員工。作者形容這些研究如同「明亮但狹窄的聚光燈」，在更廣泛效應仍未知的情況下，大眾辯論往往滑入科幻式投機。為應對這些不確定性，Nature 雜誌為文作者提出三種互補的研究方法。第一種是經濟學家 Jean Tirole 提出的「社會科學虛構」，指植根於基本經濟原則的未來推測，運用模型探索技術與市場力量的互動。例如研究模擬自動駕駛汽車可能使交通更糟，因為乘客可在車內工作或娛樂使時間成本下降，導致更多人選擇開車造成擁堵，結果取決於政府是否實施擁堵收費等政策。另一例子探討市場力量如何限制 AI 的破壞性：當自動化提升某些任務生產力時，難以自動化的活動如創意指導或審核輸出的相對價值會增長，可能提高此類工作的需求和工資以緩衝衝擊，但也可能加深不平等。第二種方法是「前瞻性數據」，需結合不同類型資訊形成完整圖景。基準測試雖是評估 AI 能力的常見方法，但分數常與現實表現脫節，因為實際任務更混亂、複雜且依賴情境，例如醫療 AI 可能在教科書式問題上表現良好但會誤解患者溝通。即時監測指標包括追蹤人們將 AI 用於哪些任務、就業變化以及整合 AI 的公司是否獲得更高利潤。一個重要參數是 AI 運行成本持續下降，塞拉利昂研究顯示 2022 年初查詢 AI 的成本是載入網頁的 12 倍，到 2025 年使用 AI 反而便宜 98%，暗示 AI 可能在網路昂貴的低資源環境中擴大資訊獲取。研究者還可設計「綠野仙蹤」試驗，讓參與者以為與 AI 互動但實際由專家執行，觀察用戶如何回應類似 AI 的互動即使技術尚無法獨立運作。

第三種方法是「試點經濟」，不只觀察 AI 如何改變經濟活動，而是測試構建市場、公司和機構的新方式。政策制定者可從軟體開發等前沿部門汲取教訓，並在教育和醫療等領域啟動試點，功能類似「AI 特殊經濟區」可安全測試技術的指定空間，包括激勵私人組織採用 AI 工具以及提供臨時規則豁免的監管沙盒。實例包括 Cloudflare 實驗讓網路創作者向訪問其內容的 AI 系統收費，以及多所學校測試 AI 支持自主學習的課程結構。文末作者強調理解 AI 的社會後果必須成為共同項目，需要技術專家、公民創



新者和社會科學家密切合作。這項工作需要關於世界如何改變的想像力，以及通過行為模型、前瞻性實驗和新協作形式試點獲得的紮實證據，從被動應對轉向主動設計未來，探索多種可能並幫助我們對想要建立的社會類型做出深思熟慮的選擇。

澳洲社群媒體禁令 AI 自然試驗：「破格立法」

澳洲成為全球首個禁止 16 歲以下兒童使用大多數社群媒體平台的國家，這項開創性政策涵蓋 Facebook、X、Reddit、YouTube、Threads 和 Snapchat 等平台，科技公司必須採取合理措施阻止青少年創建或保留帳號，否則將面臨最高 4,950 萬澳元的罰款。澳洲政府主張社群媒體正在損害青少年心理健康，造成睡眠不足並使他們接觸有害內容，但部分研究人員指出相關證據仍有爭議，因為線上交流對少數群體和偏遠地區的青少年來說是重要的支持系統。儘管許多青少年對此感到憤怒，這項政策卻為社會科學家提供了獨特的「天然實驗」機會。多個研究團隊已經展開相關研究，墨爾本莫道克兒童研究所追蹤 177 名 13 至 16 歲青少年在禁令前後的心理健康和社群媒體使用變化；珀斯的多機構合作研究檢視禁令帶來的育兒挑戰和家庭衝突；西雪梨大學調查青少年科技使用和面對面互動的改變；莫納許大學研究禁令對學生政治知識和參與的影響；昆士蘭科技大學則追蹤青少年是否遷移到不受禁令限制的平台（如 Lemon8 和 Yope）。然而，禁令也帶來意外後果，例如 Facebook 的新驗證機制阻礙了公共衛生研究人員使用假帳號追蹤線上酒精廣告的能力。澳洲獨立網路安全專員將主導為期兩年的評估計畫，由史丹佛大學領導的專家小組監測禁令對青少年戶外活動時間、同儕連結感、網路安全和處理錯誤訊息能力等面向的影響，這項全球首例的政策實施成果將為其他國家提供重要參考。

LLM 影響科學產出數量與品質：「信號失靈」

康乃爾大學和加州大學柏克萊分校研究團隊發表於《科學》期刊的研究，分析了 2018 年 1 月至 2024 年 6 月期間三大預印本資料庫共 210 萬篇論文，系統性檢驗大型語言模型對科學生產的影響。研究涵蓋 arXiv 的 120 萬篇論文、bioRxiv 的 22.1 萬篇以及 SSRN 的 67.6 萬篇，研究團隊應用基於文本的 AI 檢測算法，比較人類撰寫文本與 GPT-3.5 改寫版本的詞彙分布，以識別 ChatGPT 發布後可能使用 LLM 輔助



的手稿。研究發現 LLM 採用與科學產出大幅增加相關，在排除核心 AI 子學科後的作者層級固定效應模型顯示，arXiv 採用者的手稿提交率提升 36.2%，bioRxiv 提升 52.9%，SSRN 更高達 59.8%。更重要的是生產力提升存在顯著的語言背景異質性：亞洲姓名且隸屬亞洲機構的學者獲得最大提升，在 arXiv 達 43.0%、bioRxiv 達 89.3%、SSRN 達 88.9%，而高加索姓名且隸屬英語國家機構的學者提升幅度較小但仍顯著。這表明 LLM 能顯著降低以第二語言寫作的成本，科學生產佔率可能正向非英語母語地區的學者轉移。然而研究揭示了令人擔憂的現象：傳統科學質量信號正在快速侵蝕。研究發現三個關鍵模式：LLM 輔助手稿的寫作複雜度分數顯著高於自然語言撰寫的論文；在非 LLM 輔助論文中，寫作複雜度與發表概率呈正相關；但最關鍵的是，LLM 輔助手稿的寫作複雜度與同行評審結果之間的關係發生反轉，寫作複雜度增加反而與較低的科學價值評估相關。這一反轉模式在 2024 年國際學習表徵會議的獨立數據集中得到驗證，證實複雜的 LLM 生成語言通常掩蓋了薄弱的科學貢獻，隨著產生精緻散文所需努力下降，它作為作者掌握主題的信號效用也隨之下降，可能使文獻充斥大量表面令人信服但科學上令人失望的研究。在既有文獻閱讀與下載方面，研究利用 arXiv 論文 2.46 億次瀏覽和下載數據，分析 Bing Chat 發布前後用戶閱讀行為的變化。分析顯示 Bing 用戶在 Bing Chat 引入後發現了更多樣化的文件集：書籍訪問比率提高 26.3%，訪問手稿的中位年齡減少 0.18 年，且更多發現較少被引用的參考文獻。事件研究方法進一步顯示 LLM 採用者引用書籍的可能性整體提高 11.9%，引用文件平均年齡更年輕，被引文獻累積的引用次數更少，提供一致證據表明 AI 輔助引導學者走向更廣泛的知識體系。該文結論強調 LLM 已開始重塑科學生產，其中英語流利度的價值將減退，但穩健的質量評估框架至關重要。隨著傳統規則轉移，編輯和審稿人可能越來越依賴作者聲譽和機構隸屬等作為稿件品質評估重要參考，抵消了 LLM 對科學生產的民主化效應。

台灣通過人工智慧基本法：「蓄勢待發」

行政院於 114 年 8 月 28 日通過《人工智慧基本法》草案，由數位發展部主責推動，立法院於 12 月 23 日三讀通過，主管機關為國家科學及技術委員會為中央主管機關。



人工智慧技術近年快速發展，世界各國普遍認為 AI 可為產業與社會活動帶來廣泛的經濟及社會效益，並為企業及國家發展提供關鍵競爭優勢。然而，AI 技術在帶來效益的同時，也可能對個人或社會產生風險或影響。為因應 AI 技術快速創新與隨之而來的挑戰，本法以鼓勵創新為主軸，同時兼顧保障人民權益，期望在促進產業發展與風險管理間取得平衡提升國家競爭力。本法提出 AI 研發與應用的七大基本原則，包括永續性、人類自主性、隱私保護及資料治理、安全性、透明及可解釋性、公平性，以及可問責性，以定錨我國 AI 發展方向。在推動重點方面，政府將透過補助、委託、出資、獎勵、輔導，或提供租稅、金融等財政優惠措施，積極推動 AI 研發、應用及基礎建設，並鼓勵公私部門進行國際合作，促進與各國間人才、技術與設施的交流及利用，同時建立資料開放與共享機制，提升 AI 資料的可利用性。在人才培育方面，政府將持續推動各級學校、產業、社會及公務機關的 AI 教育，以提升國民 AI 素養。為加速 AI 技術落地應用，本法將建立或完備 AI 研發及應用服務的創新實驗環境，並鬆綁學術研究限制。在風險管理層面，政府將推動與國際接軌的風險分類框架，建立高風險 AI 應用的責任、救濟、補償及保險機制，促進 AI 的穩健與安全發展。同時，針對 AI 的應用以風險管理為基礎，建立標準、驗證、溯源或問責機制，提升 AI 決策的可驗證性及人為可控性，透過評估驗證工具避免 AI 應用觸法，並針對高風險 AI 的應用明確責任歸屬及建立救濟、補償或保險機制，以保障人民權益。在權益保障方面，本法特別關注 AI 對勞工的影響，針對 AI 利用所造成的失業者提供就業輔導，保障勞工權益。在 AI 研發及應用過程中，將避免不必要的個人資料蒐集、處理或利用，以保障隱私。此外，政府將致力於提升 AI 使用資料的品質與數量，以利訓練及產出結果，並維繫國家多元文化價值及維護智慧財產權。台灣政府以跨部會協作推動多元化的 AI 發展政策，藉由《人工智慧基本法》推動政策，建構良善發展與應用環境，讓台灣躋升全球 AI 發展的關鍵推動地位，同時維護國家主權與文化價值，創造更智慧、安全且永續的數位社會。

人工智慧 XAI 法規治理



《黑色止血鉗》電影以心血管手術案例生動呈現臨床導入創新科技可能面臨之挑戰與病患安全議題。劇中運用創新心臟手術醫狙首例發生東城大學。該機構心臟瓣膜修補高度仰賴佐伯教授「職人型」外科醫師以純熟手術技巧，在心臟跳動的狀態下完成手術。他的大弟子渡海也承襲高超手術技術。另一方面高階醫師則引進了創新醫療科技，SNIPE（醫狙），為新型導管瓣膜治療的原型系統。SNIPE 使一般外科醫師也能在心臟跳動的情況下執行瓣膜修補，擴大困難心臟疾病治療可近性。然而，新技術的導入也帶來了預期之外的風險。劇中案例首位接受新技術治療的患者小山接受高階醫師引進 SNIPE 療法。術前團隊透過多次模擬與演練，希望在正式手術前熟悉新技術並提升治療成功率。然而因為團隊對新技術尚未完全熟悉，首例使用 SNIPE 的手術中發生了人為操作失誤。人工二尖瓣雖已成功送達預定位置，但在釋放過程中定位不夠精準，導致人造瓣膜脫位並卡入心肌。此一狀況使手術陷入進退兩難。面對突發狀況，手術團隊請求高階醫師即時進入手術室協助。然而佐伯教授隨即指出，若每當新技術發生問題都必須仰賴少數專家親自介入，那麼技術的普及性與初衷便無法實現，最終仍會回到高度依賴個人技藝的舊模式。他因此要求高階醫師僅能透過遠端方式進行指導，而不得親自操作。最終，這場危機由渡海醫師以傳統手術方式化解。他憑藉熟練的手工技術，成功取出脫位的人工瓣膜，並迅速完成瓣膜修補手術，挽回病患生命。佐伯教授也藉此提醒團隊，新興醫療科技即便設計精良，仍可能在臨床應用中產生不可預期的風險。

第二位潛在受試者為年輕音樂家，正準備參加即將到來的重要音樂比賽，同時又需要接受心臟瓣膜手術。相較於傳統手術，新技術若能縮短治療與復原時間，因此成為適合的治療對象。另一方面，外科醫師期刊的前主編楠木秀雄亦同期住院，原本預計接受由佐伯院長主導的傳統手術。兩名病患在同一時間進入醫院，分別代表「創新技術」與「傳統手術」兩種治療路徑。檢查後發現，年輕音樂家的心室結構異常且顯著肥厚，使得新型 SNIPE 技術在置換二尖瓣時面臨極高風險。該手術需經由心室進入，若導管稍有偏差，碰觸心肌便可能引發致命性心律不整。這使得手術路徑的選擇，猶如在高度敏感區域中操作，容錯空間極小。為降低風險，高階醫師結合心臟影像、數位模型與人工智



慧生成技術，反覆進行模擬與計算，試圖找出一條可行且安全的進入路徑。經過長時間的嘗試，團隊終於建立出一條理論上可行的手術方案。然而，在實際臨床決策上，主導權並未交由高階醫師，而是由渡海醫師負責執行。渡海醫師坦言，憑藉自身經驗與對影像的判讀，其實無需如此大量的模擬，即可判斷以約 83 度角進入是最合適的方式。同時，他也清楚告知病患，先前使用新技術的案例曾發生嚴重失誤，病患一度瀕臨死亡，讓病患在充分理解風險的前提下，自主做出選擇。病患最終仍決定接受手術。手術當日，渡海醫師憑藉穩定的操作與精準判斷，成功完成治療。然而，就在同一時間，原本未排定手術的期刊前主編卻突發冠狀動脈瘤破裂，必須立即手術搶救。偏偏此時佐伯院長外出開會，無法即時返回，病患生命陷入高度危險。在極度緊急的情況下，渡海醫師與高階醫師分工合作，同時運用新技術與臨床經驗，最終以 SNIPE 成功完成兩位病患的治療。佐伯院長也再次提醒團隊，即使在術前透過數位預測與人工智慧模擬找到了理想路徑，新技術在實際臨床執行時，仍存在無法完全預測的不確定性風險。兩個高度對立、同時發生的臨床場景，清楚呈現創新技術導入醫療的雙面性。醫療領域導入創新科技確實提供挽救生命機會與拓展治療可能，但也伴隨著複雜的風險、責任與抉擇。

這部電影呈現 4P 精準健康照護治理架構。《黑色止血鉗》中的案例清楚顯示，當新技術遭遇複雜個案時，最終仍需仰賴臨床醫師的專業判斷與經驗接手，凸顯單純技術導向 AI 的侷限。因此，若要從「技術型 AI」走向能夠被制度、法律與臨床共同承擔的法律 XAI，必須啟動 4P 治理架構。第一個是「預防」，首先是「預防」，法律關注的是在風險尚未發生前，是否已建立可對比、可說明的決策依據。第二個涉及「預測醫學」的範疇與尺度。電影中的案例呈現每一位病患生理結構與臨床條件都不同，單一模型無法適用於所有人，使 AI 治理必須考慮個人化。第三是「個人化」，而個人化的核心問題在於「深度」。治理上必須思考，是否需要對所有人進行全面監督與防治，或僅針對高風險個案進行選擇性介入。最後是「參與」。透過條件式說明不同選擇的風險差異，讓病人能在理解下做出決策。最終，病人得以在充分理解的前提下，選擇是採用 AI 輔助手術、個人化臨床判斷，或傳統手術方式。這正是 4P 醫學與 XAI 臨床導入治理所核心精神。



以大腸直腸癌作為例子，在臨床實務常見情境是：病人詢問，上一次糞便免疫化學檢測結果為 10 ng/ml，這一次為 60 ng/ml，原本篩檢間隔是兩年，接下來應該多久再檢查一次。在傳統的一體適用篩檢策略中，不論糞便潛血濃度高低，篩檢間隔一律設定為兩年一次。這樣的制度已被證實能降低整體大腸癌死亡率，約可下降三成，且在遠端大腸的效果尤為明顯。然而，糞便潛血濃度本身已反映出不同的風險層級。當結合精準篩檢與 XAI，預防策略便不再只遵循固定年限，而是能根據濃度變化，清楚解釋「為何需要調整篩檢間隔」，使預防決策從一體適用，轉向可解釋、可對比的個人化預防。

大腸癌發生率與死亡率趨勢顯示不同風險分層之間存在明顯差異。民眾透過風險分數可區分為高風險與低風險等不同風險層級，對應不同篩檢間隔。然而風險分層伴隨代價，高風險若被誤判為低風險，可能產生偽陰性；低風險若被過度篩檢，則容易出現偽陽性。這正是從族群式篩檢走向個人化篩檢時，必須考量的治理問題。實證研究顯示，糞便潛血濃度愈高，大腸癌的發生率與死亡率皆隨之上升，當濃度與前一次檢測相比出現明顯變化，篩檢間隔便有調整的合理依據。XAI 應用在於清楚說明為何篩檢間隔會從兩年調整為三年，或從三年縮短為兩年。這樣的解釋不僅能回應個別病人的疑問，也在法規與政策層面建立正當性，同時讓有限的醫療資源得以更有效地分配。

透過依糞便潛血濃度進行風險分層，篩檢間隔可依風險高低，從半年延長至六年不等。個人化篩檢策略使篩檢間隔不再一體適用，而能更精準地反映實際風險。在資源配置層面，這樣的調整可顯著降低整體檢測需求，約有近一半的人口不再需要頻繁接受糞便潛血檢測，檢測量可減少約 49%。後續大腸鏡檢查的需求，則依轉介門檻不同，整體可降低約 11% 至 46%。這類外部效應顯示，個人化篩檢不僅影響個別決策，也能有效減輕國家與醫療體系的資源負擔。然而在個人層次，仍會產生關鍵疑問：若篩檢間隔由兩年延長為三年，是否會提高風險；反之，若縮短間隔，是否會增加偽陽性的可能性。基於既有實證研究，糞便潛血濃度已被證實與大腸癌發生率及死亡率呈現明確關聯，因此，依濃度調整篩檢間隔具有科學依據，也能在合理範圍內控制風險。惟現行政策多仍停留在族群層次，若要真正落實個人化篩檢，仍需進一步透過可解釋的方式，回應個別



風險與決策差異。

隨著大型語言模型的發展，個人化篩檢情境逐漸浮現，使原本以國家政策為核心、著重族群平均風險的篩檢模式，轉變為需要回應個別病人差異的臨床問題。以此案例為例，一名 45 歲女性接受大腸癌預防性健檢，具備多項重要個人資訊，包括既往病史、用藥史、家族史，以及基因檢測結果。其家族中曾有大腸癌與子宮內膜癌病史，基因分析顯示符合 Lynch syndrome 特徵，屬於遺傳性大腸癌風險最高的族群之一。進一步的多模態檢測顯示，血液學與生化指標、糞便潛血濃度皆偏高，腸道菌群分析與內視鏡影像亦提供了補充風險資訊。同時，透過語音敘述蒐集病史脈絡，使家族發病年齡與症狀演變得以被納入整體評估。此一案例顯示，當篩檢決策從族群層次走向個人層次，單一指標已不足以支撐決策，必須整合電子病歷、檢驗數據、影像、基因與語意資訊。這正是多模態 XAI 在個人化預防醫學中所扮演的關鍵角色。

結合大型語言模型的個人化大腸癌精準防治架構，整合病歷資料、基因檢測、腸道菌群、內視鏡影像與生活型態資訊，透過 LangChain 串接多個語言模型進行語意理解、推理與回饋建議。系統可辨識個體風險因子，如腸道菌群失衡（*Fusobacterium nucleatum* 上升、丁酸菌下降）與 MMR 基因缺陷（MSH2），進而提出分層預防策略。初級預防著重於飲食與益生菌介入，以改善腸道微環境並降低發炎風險；次級預防則依家族史與基因風險，建議提早進行大腸鏡與相關癌症篩檢。此架構強調跨模態資料整合與即時決策支援，提升大腸癌早期偵測與個人化預防效益。並以深度學習為核心建構個人化的大腸癌風險評估模式，強調多元因子與病程時序進展的整合分析。透過真實世界資料（RWD），建立個體專屬的因果網絡，納入年齡、性別、生活型態、家族史、代謝指標與腸道相關因子，動態刻畫疾病由正常黏膜、腺瘤到癌前病變之演進過程。模型結合多狀態風險架構，分別估計進階腺瘤的絕對風險、惡性轉化年危險率與疾病進展年危險率，量化不同子風險層級之轉移機率。藉由深度學習捕捉非線性關係與時間依賴特性，本模式可隨個人資料更新即時調整風險預測，支援精準篩檢與追蹤策略制定，提升大腸癌早期偵測與預防效益。



運用 XAI 中的 SHAP 分析，可系統性評估腸道共生菌與臨床特徵對大腸癌發生與進展的影響。結果顯示，在訓練資料集與獨立測試資料集中，特徵重要性的排序趨勢高度一致，顯示模型具備良好穩定性與可重現性。分析指出，多種與發炎及腫瘤微環境相關的菌屬（如 *Fusobacterium*、*Peptostreptococcus*、*Porphyromonas*）在 SHAP 值上呈現正向貢獻，與癌變風險上升密切相關；相對地，部分短鏈脂肪酸相關或代謝保護型菌群則呈現負向 SHAP 值，顯示潛在保護效果。此外，BMI 等臨床變項亦在模型中占有重要地位，反映宿主代謝狀態與腸道菌群之交互作用。整體而言，SHAP 分析不僅提升模型透明度，也有助於辨識關鍵致癌與保護菌群，為精準風險分層與個人化預防策略提供具生物意義的依據。透過 XAI 進行互動式網絡分析，深入解析腸道共生菌與大腸癌風險之關聯結構。結果顯示，*Peptostreptococcus*、*Fusobacterium* 與 BMI 位於整體影響網絡的核心位置，扮演關鍵樞紐節點，主導癌症風險的傳遞與放大效應。進一步分群分析發現，不同高風險亞群呈現差異化的菌群互動模式，例如在高風險群中，致癌相關菌屬之間連結更為緊密，顯示協同促癌的潛在機制。次高風險群則呈現不同的網絡拓撲結構。此結果指出，大腸癌並非單一菌種驅動，而是由多菌群與宿主因子交互作用所形塑的風險模式，為分群式精準預防與臨床風險分層提供重要依據。

XAI 照護應用議題實例

Tammy Dobbs 患有腦性麻痺並需長期居家照護的美國阿肯色州居民，因 AI 評估系統介入照護時數分配而引起應用 AI 於健康領域廣泛討論。Dobbs 原每日可獲得 8 小時的居家照護，在專業評估下得以維持在社區中自主生活。但自 2016 年該州政府導入 AI 評估演算法後，照護需求改由演算法計算，結果導致她的照護時數驟減至每日僅 4 小時。根據資料，評估系統會直接對應固定時數，專業人員無法依個別情況調整，引發對於 AI 決策是否合理的質疑。Tammy Dobbs 因 AI 介入導致照護時數遭刪減，最終在法律援助下提出訴訟。她主張演算法缺乏透明性，未能正確反映其實際生活所需，嚴重侵害殘障者權益。阿肯色州最高法院最終判決此制度違憲，並要求「立即停止使用」該 AI



系統。法院裁定應恢復照護時數，改由專業評估主導，強調演算法不得取代人為判斷的重要性。Tammy Dobbs 案例顯示 AI 在健康照護評估應用可能遭遇問題。關鍵爭議包含：採平均標準而非個案評估，忽略個別疾病影響。對細微評分變動過度敏感，可能造成時數誤判。未考量長時間無人協助的實際風險，如夜間翻身、外出需求等。評估結果未設有適當的覆核或申訴管道，缺乏專業修正機制。AI 系統設計應更貼近實務需求，避免將個案生活權益數據化簡化處理。此案例也突顯 AI 在照護應用治理的潛在風險包括：解釋性與透明度不足，公眾以及專業人員難以理解 AI 決策依據；演算法偏誤、訓練數據缺陷未納入如腦性麻痺等特殊因素，此外當時導入之 AI 系統設計目的為「資源分配」，並非以個人實際需求為主要目標，應用此 AI 於個人照護需求評估有違健康權益保障，出現照護剝奪問題。Tommy Dobbs 訟案後阿肯色州導入 AI 於照護評估於 2018 年遭法院裁定違憲。AI 作為輔助工具雖具潛力，但專家提醒須確保其公平性與人性化設計，避免對弱勢群體造成傷害。

以上內容將在 2025 年 12 月 31 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