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

2025 年 12 月 24 日

人工智慧在醫療場域應用已逐步由輔助工具推展為可能影響臨床判斷與決策重要角色。面對醫療中高度不確定性與病情、個人特徵以及偏好等多重因素之高度複雜性，可解釋人工智慧（Explainable AI, XAI）對於當前醫療應用發展成為銜接技術創新、臨床可用性與法律責任的核心橋樑。本周我們將探討 XAI 法律規範治理以及 XAI 醫師臨床應用風險與決策。

星球健康新知

盟國承諾安全保證 俄烏和平談判現轉機:「和平懸峙」

2025 年 12 月中旬，美國、烏克蘭與歐洲主要國家在柏林密集會談，討論俄烏戰爭以停火為前提的戰後安全安排。美方表示，若能促成停火，願向烏克蘭提供近似北約第五條的高規格安全保障，並稱其為可提交國會審議的「白金標準」方案，但同時強調提案具有時間壓力。2025 年 12 月中，美國積極推進俄烏戰爭停火或和平協議，相關外交與安全討論快速升溫。美國、烏克蘭與歐洲主要國家在柏林展開密集會談，首次觸及涉及實質政策讓步、軍事風險承擔與長期安全架構的核心議題。美方明確表示若能促成停火，願意向烏克蘭提供高度強化、具制度性與法律穩定性的安全保證，其設計方向參照北約第五條的集體防禦精神，一旦烏克蘭遭受攻擊，相關國家將視為對自身安全的威脅而採取行動，美國官員將這套構想形容為「白金標準」的安全保障，並指出若協議成形，將提交國會、特別是參議院審議，以確保其可執行性與長期約束力，但同時也提醒，這項提案並非無限期有效，烏克蘭需要在有限時間內做出回應，而領土與主權問題依然是談判中最難突破的僵局。烏克蘭總統澤倫斯基首度公開釋出重大讓步訊號，表示願意放棄加入北約的長期目標，並將此視為推動和平談判的一項妥協。他坦言基於俄羅斯長期反對以及現實政治限制，烏克蘭在可預見的未來幾乎不可能真正成為北約成員，但前

提是西方必須提供具約束力、可執行的安全保障，以防止停火後俄羅斯再次發動攻擊。澤倫斯基特別指出，這些保障不應只來自美國，也應涵蓋歐洲主要國家，並包括加拿大、日本等其他夥伴國家，其本質必須與北約第五條相當，而不能只是政治聲明或模糊承諾。儘管在安全保證層面出現某種突破跡象，領土問題仍是談判中最難化解的核心矛盾，美國提出在烏克蘭東部前線、特別是頓巴斯部分地區設立特殊經濟區或經濟自由區，作為暫時性安排，藉此降低軍事衝突風險並為重建鋪路，但這項提議實際上要求烏克蘭自仍由其控制的部分地區撤軍，而俄羅斯卻未被要求對等後撤，澤倫斯基表示，這樣的安排缺乏公平性，等同默認俄羅斯以武力改變現狀，他強調，最合理的停火方式，應是雙方就地停火，在既有前線位置凍結戰線，其他主權與領土爭議再透過外交途徑處理。在歐洲方面，德國總理梅爾茨進一步將這套安全構想推向更具實質性的軍事意涵，他在接受德國媒體訪問時指出，根據美歐對烏克蘭提供的停火後保障設計，若任一方違反停火條款，擔保國在某些情況下將有必要「擊退」俄羅斯軍隊。他說明，未來可能在交戰雙方之間設立非軍事區，並在俄羅斯入侵或攻擊時採取實際行動，這顯示西方正在討論的不只是象徵性的維和存在，而是具備強制力與軍事回應能力的安全安排。梅爾茨也直言，美國承諾在停火情況下保護烏克蘭，如同保護北約領土一般，對華府而言是一個相當引人注目的新立場，但前提仍是俄羅斯必須同意停火，而目前莫斯科尚未點頭，也未同意允許西方部隊進駐烏克蘭。歐洲內部對戰後安全風險的警戒並未因談判升溫而降低。芬蘭總理奧爾波公開警告，即使俄烏之間達成停火或和平協議，俄羅斯仍將構成長期威脅，並很可能把目前部署在烏克蘭戰場的部隊轉而調往北約東翼與波羅的海周邊地區。北歐與波羅的海國家的評估普遍認為，俄羅斯可能在三到五年內具備與北約發生重大軍事對抗的能力，因此不能因戰事趨緩而低估安全風險。奧爾波呼籲歐盟加大對前線國家的財政與軍事支持，特別是在防空、無人機與地面作戰能力方面，並指出在當前情境下歐洲必須更清楚也更主動地承擔自身防衛責任

主要經濟體展現轉型韌性應對結構性壓力:「穩中趨緩」

全球經濟面臨多重結構性轉變，成長動能持續受限。除了成長放緩與貿易摩擦外，人口老化、人工智慧對勞動市場的衝擊、關鍵資源短缺、金融市場風險偏好升高，以及歐洲資本市場的結構性弱點，正共同壓抑中長期成長潛力。除了成長趨緩、貿易衝突與資本流動受限等在 2025 年已廣泛討論全球經濟變局現象外，五項結構趨勢將深刻改變 2026 年全球經濟發展。首先是人口結構的快速老化與區域分化，全球勞動年齡人口與退休人口的比例正急遽下降，先進國家面臨勞動力萎縮與財政壓力，而開發中國家則將迎來龐大的勞動人口與隨之而來的移民壓力。其次，人工智慧帶來的勞動市場衝擊，可能在提升生產力的同時，大量取代例行性工作，並加劇資本與勞動之間的分配不均。第三，關鍵自然資源如銅、鋰與水資源的結構性短缺，將限制能源轉型與高科技產業的擴張。第四，美國金融市場風險偏好持續上升，資產價格高漲、投機活動擴張，使金融體系潛藏不穩定性。最後，歐洲與英國長期存在的風險趨避文化與資本市場碎片化，已削弱其在創新與 AI 浪潮中的競爭力。這些趨勢共同構成一個背景，使當前歐洲低成長但穩定的狀態，不只是循環問題，也帶有結構性限制。英國的情況呈現出更明顯的勞動市場降溫訊號。路透社報導指出，截至 2025 年 10 月的三個月內，英國失業率升至 5.1%，為 2021 年初以來最高水準，私部門薪資成長降至 3.9%，創近五年新低，同期受薪就業人數在 11 月單月減少 38,000 人。官方與民間分析普遍認為，經濟動能趨弱、政策不確定性，以及先前提高雇主社會保險負擔，已促使企業縮減招募。這些數據強化了市場對英國央行即將降息的預期，儘管通膨仍高於 2% 目標，但就業與薪資壓力的降溫，讓央行更有空間調整政策。整體而言，英國的例子顯示，在高利率環境下，勞動市場往往成為最先反映經濟壓力的部門。在連續第三次降息後，聯準會將政策利率調降至 3.5%–3.75%，鮑爾坦言美國勞動市場正在降溫，失業率已升至 4.4%，就業成長明顯放緩，但他同時否認出現急遽惡化的風險。聯準會內部預測顯示，失業率可能在 2025 年底升至 4.5% 後趨於穩定。就業放緩部分來自移民與勞動參與率下降，也與企業招聘需求轉弱有關，但目前的貨幣政策已接近中性，有助於勞動市場在不劇烈衝擊下逐步調整。這種低招聘、低解雇的環境，與歐洲與英國所面臨的狀態形成呼應，也反映主要先進經

濟體同步進入後緊縮時期的調整階段。2025 年 10 月歐元區工業產出出現明顯回溫，月增率達到 0.8%，高於 9 月的 0.2%，也符合市場原先的預期。若以年增率來看，10 月工業產出成長 2.0%，同樣高於前一個月的 1.2%，並略優於經濟學家調查所預期的 1.9%。德國工業表現相對突出，單月成長 1.4%，彌補了義大利下滑 1.0 以及法國表現偏弱所造成的拖累。這些數據支撐了歐元區在 2025 年整體經濟表現比年初擔憂來得穩健的說法，歐洲央行總裁拉加德也已公開表示，最新的成長預測可能再度上修。不過，報導同時強調，這樣的改善並非景氣反轉或進入強勁擴張階段，歐元區整體經濟成長仍僅略高於 1%，出口部門持續疲弱，近年能源成本高漲與中國高科技製造快速擴張，使歐洲產業在國際競爭上長期承壓。即便產業可能已觸底，但並無明顯復甦動能，新一輪美國關稅政策如何重塑全球貿易結構，仍存在高度不確定性。在此背景下，歐元區展現的是下行風險受限而非成長動能明確的狀態。歐洲主要經濟體在經歷高通膨、高利率與美國貿易政策衝擊後，確實避免了明顯衰退，但經濟動能依舊疲弱。最新的歐元區綜合 PMI 在 12 月降至 51.9，為三個月低點，顯示企業活動擴張幅度不如預期，主因仍是德國製造業持續萎縮，儘管法國略有改善。家庭消費保持謹慎，各國政府債務水位偏高，也限制了財政刺激空間。值得注意的是，實體數據如工業產出表現優於企業調查，加上就業市場依然緊俏、能源成本顯著下降，使歐洲央行判斷目前貨幣政策已接近適當水準。市場普遍認為歐洲央行的降息循環可能已結束，甚至開始押注未來數年內出現升息的可能性。同時，德國擴大國防與基礎建設支出，被視為中期支撐經濟的重要因素，部分經濟學家估計，相關政策可能在 2026 與 2027 年各自為 GDP 貢獻約 0.6 個百分點。不過，整體而言，這仍是一種撐得住，但跑不快的經濟狀態。經濟分析認為歐洲目前呈現的是「下行風險受限，但成長動能不足」的狀態，即使有國防與基礎建設支出作為中期支撐，整體經濟仍屬撐得住、但跑不快。

全球經濟貿易地緣座標再定位：「實力定錨」

法國總統馬克宏上週訪問北京，清楚映照出歐洲在中美競逐格局下所承受的結構性壓力。此行原本著眼於爭取可見的商業成果，以回應法中貿易逆差擴大與歐洲對中國在

烏克蘭議題上角色的期待，但會談結果顯示，中方更重視戰略與政治訊號，而非短期經貿交換。法國力推的空中巴士採購案未能落實，被視為中國已不再將商業合作與政治立場切割的象徵。這一轉變突顯歐洲在戰略自主與談判槓桿上的不足。歐洲既缺乏足以平衡美國的硬實力，也難以在面對中國時提出具約束力的條件。這種轉變使歐洲所面臨的結構性弱勢更加浮現。歐洲既缺乏足以平衡華盛頓的硬實力，也難以在面對北京時提出具約束力的條件。這種戰略自主不足的狀態，與歐中之間日益擴大的經濟失衡相互交織。中國對全球其他地區的貿易順差已達約一兆美元，對歐盟的順差在十年間幾乎翻倍，來到約 3,000 億歐元。在美國提高關稅、中國國內需求成長乏力的情況下，大量中國出口品持續轉向歐洲市場，對歐洲產業與就業形成壓力。單靠關稅或配額的防禦性措施，難以回應失衡背後的深層因素。問題不僅在於中國長期以出口導向支撐成長，也在於歐洲自身生產力與競爭力的相對疲弱。因此，重新平衡被視為一項需要中、美、歐共同參與的宏觀調整過程。對歐洲而言，政策重點包括深化能源、醫療與數位市場整合，強化創新能力，並透過制度設計，引導約 30 兆歐元的本地儲蓄投入高成長潛力產業。同時，歐洲仍須保留關稅、反脅迫等工具，以因應不公平競爭。對中國而言，則需要透過財政政策刺激內需、擴大服務業比重，並改善外資進出條件的對等性。若合作管道無法有效推進，歐洲轉向更具保護色彩的政策選項，將難以避免。英國的經驗，則提供了另一個層面的對照。原被形容為「歷史性突破」的美英科技繁榮協議，已因整體貿易談判卡關而陷入停滯。該協議原於川普九月訪英時公布，涵蓋人工智慧、量子運算與先進能源等領域，但僅以不具法律效力的備忘錄形式存在，且明確與更廣泛的經濟協議進展掛鉤。美方對英國數位監管、食品安全等制度的疑慮，使科技合作成為談判籌碼的一部分，反映出即便在政治語言高度親近的情況下，英美經濟關係仍呈現高度交易性。即使美國科技企業對英投資計畫短期內不受影響，這種「投資與制度談判分離」的狀態，本身就凸顯英國在制度協商上的槓桿有限。相較之下，英國與南韓達成的自由貿易協議，呈現出截然不同的圖像。該協議不僅確保 98% 的關稅項目永久維持零關稅，也為英國服務業帶來可量化的出口成長預期，並更新原產地規則、數位貿易條款，為汽車、製藥與食品

產業提供制度確定性。英國政府藉此強調，其獨立貿易政策能帶來具體經濟利益，並透過深化與亞洲主要經濟體的連結，分散對美國與歐洲市場的依賴。然而，這項成功同時反襯出另一個現實：即便政治同盟關係緊密，英美之間仍難以跨越結構性分歧，反而在與其他夥伴的談判中，更容易達成完整且可落實的協議。在當前地緣政治與全球經濟重組的環境下，單靠政治象徵、歷史情誼或文化互動，已不足以支撐實質利益交換。無論是歐洲還是英國，真正的談判空間，終究取決於自身制度整合、經濟韌性與政策一致性的累積程度。

氣候變遷失控警訊: 印度洋周邊極端氣候衝擊：「治理失速」

近年印度洋周邊接連出現多個罕見熱帶氣旋，重創印尼、斯里蘭卡、泰國與印度等國，多座城市與村落遭洪水淹沒，造成上千人死亡、數百萬人受災，災害規模接近海嘯等級。研究指出，氣候變遷導致海溫上升與印度洋偶極子變化，使氣旋更易生成且降雨更極端，加上森林砍伐與採礦等環境破壞，進一步放大災害衝擊。2025 年 11 月下旬發生在印度洋沿岸的一場複合型氣候災難。短短一週內，熱帶氣旋 Senyar、Ditwah 與 Koto 接連生成並侵襲印尼、馬來西亞、越南、泰國、印度安達曼—尼科巴群島、菲律賓與斯里蘭卡等地，對一向少有熱帶氣旋侵襲的區域造成極端衝擊。豪雨、強風、土石流與暴洪同時出現，造成至少一千人死亡，大量村落被厚重泥流掩埋，道路與橋樑被沖毀，數百萬人被迫離開家園，陷入飢餓、疾病與孤立無援的狀態。作者形容降雨之猛烈，彷彿在陸地上形成洶湧海浪，河流化為具備海嘯破壞力的急流，連混凝土橋樑與巨木都被捲走，顯示這並非一般洪災，而是結構性失控的極端事件。在災害規模與人道需求龐大情況下，各國政府與國際社會反應遲緩。部分國家如斯里蘭卡已宣布國家緊急狀態並對外求援，但印尼政府在第一波災情發生十天後才召開首次內閣會議，且遲遲未正式宣布國家緊急狀態，導致國際援助無法快速啟動。許多受災城市與村落長時間與外界斷聯，糧食短缺、物價飆升，醫院甚至因災損而停擺。地方政府雖嘗試提供協助，卻受限於嚴重受損的交通與通訊基礎設施，使救援行動進展緩慢。作者同時指出，國際領導人多僅

止於表達慰問，真正的資源與行動明顯不足，歐盟與美國甚至未明確表態，凸顯這場災難在全球政治與輿論中的邊緣化。國際媒體的報導普遍將事件零碎地分割為各國洪災，忽略這些災害其實來自同一氣候系統的異常行為，因而削弱了外界對災害整體性與嚴重性的理解，即使後續改以東南亞洪水作為統稱，仍未真正觸及問題核心。印度洋地區對極端天氣特別脆弱，原因不僅在於全球氣候變遷，也包括長期的環境破壞，如森林砍伐、礦業開發與河岸不當建設，這些人為因素大幅削弱自然的防洪能力。令人憂心的是，許多國際氣候倡議者同樣未將此事件明確定義為一場氣候悲劇，使其在全球氣候正義的討論中幾乎失聲。在科學層面氣候模型研究顯示南亞與印度洋周邊國家本就被預測將成為極端氣候災害風險上升最快的地區之一。近年來氣候變遷已影響印度洋偶極子這一關鍵的海溫循環機制，進而改變全球天氣型態與熱帶氣旋的生成條件。過去因赤道附近科氏力較弱，熱帶氣旋極少影響印尼，但 2021 年 Seroja 氣旋首度登陸印尼已是一項警訊，而僅隔四年再次出現 Senyar，顯示這種「例外」正快速成為常態。官方資料指出，異常升高的海溫可能使局部地區單日降雨量超過 300 毫米，達到平常的數倍，進一步放大災害強度。蘇門答臘自 1990 年代以來已失去超過八成的低地森林，取而代之的是油棕園、紙漿林、煤礦、金礦與各類基礎建設，許多直接位於河岸地帶，使洪水成為常態。這些開發行為往往與居住在雅加達或新加坡等大城市的權勢人物相關，真正承擔後果的卻是對環境破壞毫無責任、也非溫室氣體主要排放者的地方居民。進前之氣候災難應視為氣候影響之整體事件，透過更多關注與政治壓力加速救援與重建，同時也迫使全球社會正視人類正在進入的新氣候現實，及早辨識其規律，避免未來付出更多生命作為代價。

極端降雨大都會居民死亡風險影響 孟買生存危機：「韌性失配」

極端降雨與洪水已成全球大城市常見現象，但其對死亡率的影響過去缺乏系統分析。最新研究以印度孟買為例，結合氣候與生命資料，建立氣候變化加劇死亡風險的預測模型。結果顯示，季風期間的極端降雨與顯著死亡風險上升有關，突顯都市在氣候衝擊下的公共衛生脆弱性。Bearpark 等人在孟買所做的實證研究顯示極端降雨與洪水在高度都市化、人口密集的沿海巨型城市中轉化為嚴重且持續的公共健康風險，並且這種風險

在氣候變遷背景下正快速放大。雖然全球大型城市中極端降雨與洪水事件早已被廣泛記錄，但它們與死亡率之間的量化關係，特別是對不同脆弱族群的差異影響，長期以來仍缺乏細緻分析。過去的多城市研究甚至完全未涵蓋印度，而印度正是人口快速都市化、經濟成長迅速、氣候風險高度集中的國家。孟買在季風期間常出現短時間、高強度的次日尺度降雨，這類降雨極易造成積水與排水失效，進而引發一連串健康危機，但其健康後果過去並未被充分評估。Bearpark 等人透過分析孟買的死亡資料，估計極端降雨與洪水所造成的超額死亡，也就是實際死亡數高於在正常情況下預期的部分。結果顯示，在季風季節中，超過 8% 的死亡可歸因於極端降雨，而其中約 85% 的死亡負擔集中在居住於非正式聚落（貧民窟）的人口身上，凸顯社會不平等在氣候衝擊中的放大效果。研究也進一步發現，風險並非只侷限於降雨當下，而是會延續數週，顯示死亡並非單一短暫事件所致，而是透過疾病傳播、醫療與公共服務中斷、居住環境惡化等間接機制，形成持續性的健康衝擊。這樣的死亡負擔，換算成經濟損失，每年約達 12 億美元，規模已可與印度全國層級的直接洪水損失相比。在脆弱族群分析上五歲以下兒童在遭遇單日降雨量達 150 毫米後的五週內，死亡率增加約 5.3%，而女性的風險上升幅度也明顯高於男性。這些差異與水媒傳染病暴增、生理壓力較高、照護資源取得不足密切相關。非正式聚落因排水系統不足、衛生條件惡劣與居住安全性低，更進一步放大這些危險，使城市中形成高度集中的死亡熱點。文章清楚指出，這些降雨相關死亡不只是氣候現象的直接結果，而是深深嵌入社會與經濟結構中的不平等所共同影響。當極端降雨與高潮位同時發生時，排水受限會使死亡風險顯著上升。僅一小時 30 毫米的降雨，就可能在五週內使死亡率上升 0.6%，而在高潮汐水位條件下，影響會更加嚴重。對未來的推估顯示，即便只是 5 至 15 公分的海平面上升，也可能使孟買的降雨相關死亡增加 7% 至 21%。這些結果與 IPCC 的結論相呼應，顯示極端降雨與海平面上升這類複合事件的發生機率正在上升，對沿海巨型城市構成長期且系統性的威脅。

AIx生物學 AlphaFold 重塑蛋白質科學: 「範式轉移」

人工智慧系統 AlphaFold 的出現快速改變蛋白質科學研究。過去仰賴 X 光晶體學與

冷凍電子顯微鏡的結構解析，往往需耗時數月甚至數年，如今 AI 可在數分鐘至數小時內提供高可靠度的三維結構預測。此技術不僅加速科學發現，也降低研究門檻，成為實驗設計的重要輔助工具。AlphaFold2 自 2021 年由 DeepMind 公開以來，至今已滿五年，這段時間內它對生命科學研究方式造成的改變，已不只是技術進步，而是深刻影響了研究者「如何提出問題、如何規劃實驗、以及如何解讀結果」。在 AlphaFold 出現之前，蛋白質三維結構的取得高度仰賴 X 光晶體學、冷凍電子顯微鏡或核磁共振等實驗技術，這些方法不僅耗時、成本高，且存在成功率不確定的風險，使得結構資訊往往成為研究進展的瓶頸。AlphaFold2 則大幅降低了這道門檻，讓研究者能在研究初期即取得高可信度的結構預測，進而改變整個研究流程的節奏與重心。以斑馬魚受精研究為例，2018 年研究者已發現卵表面的蛋白質 Bouncer 對於精卵辨識至關重要，但其作用機制長期無法釐清。直到 AlphaFold2 的結構預測顯示精子蛋白 Tmem81 可能形成穩定複合體，並在結構上創造出可供 Bouncer 結合的特定區域，研究者才得以提出具體可檢驗的分子機制假說，後續實驗結果也支持了這一結構導向的推論。這類案例顯示，AlphaFold 能實質縮短從觀察現象到機制假說之間的距離促成科學理解的突破。在全球使用層面，AlphaFold 的影響力主要來自其高度可取得性。隨著原始程式碼與 AlphaFold Database 的全面開放，研究者幾乎可以對任何已知蛋白質序列取得結構預測。目前資料庫已累積超過 2.4 億個蛋白質結構模型，使用者人數達到 330 萬人，分布於 190 多個國家，其中包含大量來自中低收入國家的研究者。這種規模代表結構資訊首次成為一種近乎「基礎公共資源」，而不再侷限於少數具備昂貴設備與專門技術的實驗室。引用數據的分析也反映了 AlphaFold2 的持續影響力。描述 AlphaFold2 的 2021 年 Nature 論文，其引用次數已接近四萬次，且在發表後多年仍維持高水準，並未出現快速衰退的現象。與此同時，引用 AlphaFold 的研究主題分布顯示，它不僅被廣泛應用於傳統的結構生物學，也成為 AI 蛋白設計、計算藥物探索與蛋白質動力學分析等計算導向研究的重要基礎工具。這顯示 AlphaFold 的角色更接近一個通用平台，使結構資訊能被直接整合進多種研究脈絡之中。在實驗結構生物學領域使用 AlphaFold 的研究者向 Protein

Data Bank 提交實驗結構的數量明顯高於未使用者，平均差距約為五成，顯示 AlphaFold 預測結果能協助研究者更有效地設計實驗、解讀繞射或影像資料，判斷哪些結構問題值得投入昂貴的實驗資源。若以短期的論文數量或臨床轉譯成果作為衡量標準，AlphaFold 的影響並不一定會立即顯現，整體分析顯示，與 AlphaFold 直接或間接相關的研究可涵蓋二十多萬篇論文與近八十萬名研究者，但在平均產出上，使用 AlphaFold 的實驗室僅呈現小幅增加。AlphaFold 讓研究者能夠把注意力從基礎結構取得，移到更高層次功能解釋、系統整合與假說驗證，提高其對臨床應用與產業轉化影響。

網路互動轉型 次世代智慧網路模式:「行動轉譯」

網際網路逐步邁向以 AI 代理操作為核心邏輯新階段。網路目前以人類透過瀏覽器主動搜尋、點擊與比對資訊，即使網站與服務不斷演進互動方式仍以人類手動操作為主。隨大型語言模型的成熟 AI 不再只負責回答問題或整理資訊，而是透過結合工具、權限與決策能力的智慧代理人，實際承擔原本由人類完成的工作，例如規劃行程、比價選擇、訂票、管理信件與處理各種日常事務。網路的使用方式，因而從「人操作網站」逐漸轉向「人委託代理人，由代理人操作網路」。代理人真正能在網路世界中自主行動，關鍵除了模型能力尚須網路底層結構調整。現有的網站與線上服務多半透過 API 提供功能，但這些介面與文件長期以來術設計使用者目標為工程師且缺乏一致性，對以自然語言推理代理人而言整合成本高且脆弱。當前產業開始推動新的通訊與協作標準替 AI 代理人建立一套共通語言。如 MCP 等協定目標在於讓代理人不必逐一理解複雜技術文件，能直接詢問系統功能與限制，安全地執行任務。A2A 則聚焦於代理人彼此之間的識別、能力宣告與協作，使多個代理人能分工合作、彼此信任，完成單一代理難以處理的複雜任務。這種轉變也迫使人們重新思考網站本身的角色。多數現有網站是為人類視覺與操作習慣打造，資訊被拆解成層層頁面與選單，方便人慢慢瀏覽，卻不利於代理人快速理解整體意圖。如 NLWeb 嘗試把網站轉換成能以自然語言互動的形式，讓使用者或代理人直接表達完整需求，而不必經過繁瑣的點擊流程。更進一步來看，這類技術也讓網站

同時成為代理人可直接存取的端點，逐步把「給人看的網路」與「給機器用的網路」接合。隨著代理人成為新的行動主體，平台競爭的重心也開始轉移。過去的入口之爭圍繞著瀏覽器、搜尋引擎與 App，如今則延伸到誰能代表使用者在不同服務之間行動。代理人若能直接完成查詢、比較、購買與管理，平台對使用者行為的掌控方式勢必改變，電商、內容平台與服務供應商之間的權力關係也會隨之重組。廣告模式同樣面臨調整：原本以人類注意力為核心的設計，未來可能要轉向影響代理人的決策邏輯，說服的對象不再是使用者本人，而是行動數位分身。當 AI 被賦予實際行動能力，錯誤的代價也隨之放大，代理人可能在理解不完整的情況下做出錯誤決策，甚至在外部誘導下洩漏資料或執行未授權操作。這使得權限控管、最小授權原則、人類確認機制與安全邊界設計，成為不可或缺的配套。部分任務必須維持在人類監督之下，確保代理人的行動始終可被理解與修正。目前網路快速納入代理人化情境，包含大量資訊搜尋、比對與操作在背景中由機器完成，人類不再需要逐步操作每個介面，可委託 AI 決策。網路的重心，也從「讓人更容易瀏覽」轉向「讓機器更容易行動」，人類則逐步退居為目標設定者與最終裁決者。

AI Token 用量成長與產業價值迷思:「量質脫鉤」

AI 市場常以 Token 用量成長作為需求與產業價值指標，但專家提醒，Token 耗用與模型能力、實際需求並非線性關係。隨著晶片與模型技術進步，專業問題的單位 Token 成本快速下降，各家語言模型定價策略差異也極大。當前人工智慧產業在需求判讀與獲利預期之間所存在的結構性落差。以 token 作為人工智慧需求的指標，其實相當模糊。儘管 Google、阿里巴巴、OpenAI 等企業反覆強調 token 用量呈現爆炸性成長，但這些數字並不必然對應到真實、可付費的外部需求。其中相當一部分成長，來自既有產品大量導入 AI 功能後所產生的內部消耗，例如搜尋摘要、推薦系統、影像與內容優化等，這類使用多半是功能升級，而非新增營收來源。另一部分則來自模型本身的「自動膨脹」現象，隨著模型能力提升，回答變得更冗長、推理步驟更細碎，單一任務自然消耗更多 token，使得總量看似快速攀升。從成本面觀察達到相同能力水準的單位推論價格在短時

間內大幅下滑，顯示硬體進步與模型效率改善確實讓生成單一 token 的成本持續下降。然而，這並不代表整體 AI 服務正在變得便宜。但模型規模不斷擴大，完成同一任務所需的 token 量持續增加，使得總成本居高不下，邊際成本也難以明顯壓低。這種「單位成本下降、總體消耗上升」的結構，正是 token 成長容易被誤讀的關鍵所在。模型供應商普遍將獲利期待集中在輸出 token，透過拉高輸出價格來維持高階模型的能力溢價。然而，在中階與效率型模型快速進步、競爭者不斷湧現的情況下，實務應用對「頂級模型」的依賴正逐步下降，價格壓力隨之升高。廠商既難以大幅調升價格，又必須持續投入資本以維持模型競爭力，導致獲利前景愈發吃緊。OpenAI 等企業已預期在相當長一段時間內仍將承受營運虧損，且未來數年的虧損規模可能遠高於當期收入。在這樣的脈絡下，產業前景的「破滅風險錯估」便浮現出來。token 的快速成長，並不等同於 AI 產業的獲利能力同步上升，更不能直接視為可持續的價值創造指標。若投資人將 token 視為新時代的 AI 需求評估指標，忽略其背後來自內部使用、效率變化與模型膨脹的因素，極可能重演網路泡沫時代對成長指標的誤判。人工智慧的應用前景仍然廣闊，但關鍵在於這些使用是否能在競爭激烈且成本壓力沉重的環境下，轉化為穩定且可持續的商業回報。

自我改進強化學習人工智慧:「人機共鑄」

《Nature》報導一項前沿研究，透過 meta-learning 讓 AI 不僅執行強化學習演算法，更能在高層次上自行發現與優化演算法本身。結果顯示，AI 設計的學習規則在多項任務中超越部分人類專家成果，並展現跨任務泛化能力。該創新研究聚焦新穎方向：讓人工智慧不只是使用學習演算法，能夠發現與設計演算法本身如同生物演化一般創造更高效率以及能力之 AI 模式。meta-learning 核心概念在兩個層級上同時進行學習，基層（base layer），以強化學習代理人在環境中行動學習，例如玩 Atari 遊戲，在設計層（meta layer）另一個神經網路則負責產生、調整與評估這些強化學習演算法，根據代理人在環境中的表現回饋，持續修正演算法的結構與更新規則。研究結果顯示當訓練所涵

蓋的環境數量增加、以及可用的運算資源變多時，這套系統所「學到」的強化學習演算法，其表現會隨之提升，並在多個標準基準任務上超越部分由人類專家精心設計的演算法。這一點本身已相當引人注目，因為強化學習演算法通常涉及許多細緻的數學與經驗設計，而這裡卻是由神經網路在搜尋空間中自行摸索出有效的更新規則。由此設計中 AI 發現的演算法，在訓練過的任務上表現良好，還能在一些從未見過的新任務中維持不錯的表現。作者認為跨任務的泛化能力是這項工作與先前自動化演算法搜尋研究相比關鍵進展。研究團隊強調，此成果凸顯人機協作潛力，但仍受限於人類預設框架與現實回饋設計問題。

人像資料倫理示範 FHIBE 資料集:「以人為本」

《Nature》介紹 Fair Human-Centric Image Benchmark（FHIBE），一個以明確同意與自我回報標註為核心的人像影像資料集。FHIBE 強調公平性分析用於檢視 AI 在性別、年齡與互動情境上的偏誤表現。FHIBE 影像資料集理念是在人工智慧研究中，嘗試以合乎倫理的方式處理與「人」相關的影像資料。作者指出，過去電腦視覺領域的重大進展，往往仰賴從網路大量蒐集影像所建立的資料集與競賽基準，這類作法雖然成功推動了技術突破，卻也逐漸暴露出偏見、刻板印象、不當脈絡，以及未經當事人同意使用影像等倫理問題。FHIBE 的建立正是為了這些長期存在的爭議。這個資料集完全由志願參與者提供影像，並在蒐集過程中取得明確同意，用途僅限於研究評估。資料集包含一萬多張影像，涵蓋近兩千名個體，來源遍及八十多個地理區域，目標並非擴大量體，而是提供一個可用來檢視模型公平性的參考標準。在資料標註方式上，FHIBE 採取與傳統影像資料集顯著不同的做法。許多與個人身分與社會意涵密切相關的資訊，例如年齡、祖源、偏好的代名詞、姿勢與互動情境，都是由影像中的當事人自行回報，而非由第三方觀察者推斷。作者認為，這樣的設計有助於降低外部標註者將自身偏見投射到資料中的風險。除此之外，資料集仍保留了豐富的視覺層級標註，包括臉部與身體邊界、像素層級的分割結果、關鍵點、臉部幾何結構，以及外觀特徵與拍攝環境等資訊，使其在技術評估上仍具高度價值。當模型同時需要處理多個敏感屬性時，規模較小、任務較專一的

模型，表現差距會變得特別明顯。對於視覺語言模型，作者透過不同提示詞進行測試，發現與性別相關的提問容易誘發帶有偏見的回應，顯示即使是能力強大的基礎模型，也可能在語言與影像的交互推論中再現社會刻板印象。由於資料量有限，它無法支撐現今動輒需要數百萬甚至數十億影像的模型訓練流程。此外，建立這樣一個遵循倫理原則的資料集，所需投入的時間與資金成本極高，顯示未來在訓練資料層級全面落實倫理要求，仍是一項艱鉅挑戰。雖然自我回報式標註對處理性別等社會敏感屬性相當重要，但外部觀察者如何解讀影像，在心理學與社會科學中同樣具有研究價值。作者認為，未來若能同時保留自我標註與外部標註的資訊，或許能在促進公平的同時，也幫助研究者更深入理解人類與人工智慧如何進行視覺判斷。研究指出，建立合乎倫理的資料集成本高昂，顯示未來在 AI 發展中兼顧技術與倫理仍具高度挑戰。

美國 AI 創世任務 風險治理與科技競逐:「雄心壯志」

美國政府宣布推動「Genesis Mission」，整合國家實驗室資料與超級電腦資源，讓學界與企業訓練科學型 AI 與研究代理人。計畫目標在十年內提升科研生產力，但資金、資料權限與治理機制仍不明朗。學界警告，在安全監管不足下快速擴展 AI 能力，恐放大研究誤導與資源錯配等風險。白宮於 2025/11/24 以行政命令推動的 Genesis Mission 由美國能源部（DoE）建立平台，讓學界與 AI 公司能在 17 個國家實驗室的科學資料集與超級電腦資源上訓練大型科學 AI 模型，並進一步發展可自動化科研流程、提出假說的 AI agents；計畫口號是希望在十年內把美國研究與創新「生產力與影響力加倍」，並以全球科技主導權競賽作為敘事背景。DoE 需在 60 天內提出 20 個潛在的科學/技術挑戰題目（如核融合、量子資訊、關鍵材料），盤點聯邦算力、挑出第一批可用資料資產、規劃如何安全納入外部資料，並被期待在 9 個月內對其中一個挑戰做出平台能力示範。企業端則計畫為外部夥伴，涵蓋 Microsoft、IBM、OpenAI、Google、Anthropic、以及量子公司等，在平台上的精確角色、是共同研發、使用者設施協議、或其他合作模式仍不清楚。此計畫著重於下降「資料＋算力」門檻：國家實驗室有大量高價值、通常

不易取得的科學資料與儀器產出，若能在合規前提下被更廣泛地用於訓練，可能催生更「科學能力更強、較可信」的模型，加上像 Argonne 的 Aurora 這類 GPU 導向超算對 AI 訓練特別合適，對研究節奏可能有直接推力。第二為促進 AI 運用於科研工作流程發展：以 agents 結合專業軟體與程式工具自動化資料處理、模擬、假說生成與實驗設計，讓更多研究者能接近原本只有少數頂尖團隊才用得到的基礎設施。此計畫亦具有生治理與可行性挑戰。資金面上計畫沒有公開明確價碼，額外預算仍得過國會；同時報導提到川普提案對 DoE 科學預算有削減幅度，AI 經費若不足，或必須從其他研究項目挪移，使推行落差加劇。資料安全與權利界線也是計畫執行挑戰，行政命令要求符合分類、隱私與智財規範，但一旦形成龐大的公私協作、讓公司以前所未有的方式接近聯邦科學資料，外界自然會關切誰能用、能用到什麼深度、產生的模型與成果歸屬如何界定。把分散、異質、品質不一的資料變成可用的平台引擎任務艱鉅，過去成功的大型科學資料/設施（像 Protein Data Bank、CERN 的大型強子對撞機）之所以能放大效益，與其開放性與資料結構化程度高度相關，而國家實驗室的大型資料並不一定都已達到可直接整合可重用狀態。參與公平性也是重要考量，若沒有清楚設計讓更多美國科學家能運用機制，利益可能主要流向原本就已在 DoE 體系或既有平台工作的研究者與公司。最後是 AI 安全與監管環境，川普在 2025 年 1 月撤銷了拜登政府的 AI 安全行政命令，且政府立場偏向產業、甚至主張對過度管制 AI 的州扣留聯邦 AI 資金。在缺乏全面聯邦立法的情況下，當計畫把目標拉到更廣能力的通用模型與 agents，研究社群擔心的風險（缺少足夠 guardrails、能力擴張帶來不可預期後果）會被放大，且在科學場域裡一旦導入錯誤或偏誤，可能造成研究方向被誤導或資源錯配。

AI 治理需求與國際組織提案:「WAICO」

《Nature》近期社論指出在 AI 能力快速擴張但國際監管薄弱的情況下，中國提出的世界人工智慧合作組織（WAICO）。AI 能力快速增長，可能推動科學與經濟，但模型不真正理解世界、會以難以預測的方式失誤，風險包含加劇不平等、助長犯罪、擴散錯假訊息，甚至有研究者擔心未來超智慧 AI 造成毀滅性後果，然而這些風險在競相開發

的節奏裡沒有被給到相稱的注意力。如果各國只靠宣言與白皮書、卻缺乏具約束力或可執行的機制，那麼再多共識語言也不會自動轉化成安全行為。文章援引美國、歐盟、中國三種路徑對照。對美國描述是沒有全國性 AI 法規，主要靠州層級拼布式法律與企業自律；並引用 Future of Life Institute 的 AI Safety Index（2025/12/3 發布）指出美國大公司分數沒有高於 C+，同時又提到當時的美國總統川普啟動 Genesis Mission、讓開發 AI 的公司與研究者更容易取得政府資料集，形成「風險治理不足、能力推進更快」的張力。對歐盟則是點出 AI Act 分階段上路、要求最強大的先進系統加強威脅分析，但也提醒高額罰款是否有效仍不明朗，且媒體報導企業在施壓希望弱化規範。這段的隱含訊息是：即使有立法，執行力度與產業政治經濟仍可能稀釋效果。中國 AI 策略推行一方面以政府力量積極把 AI 嵌入社會與產業如地方政府聊天機器人、工廠機器人等，另一方面監管者也推動可追溯、可問責要求。自 2022 年以來推出多項法律與技術標準，要求生成式模型在部署前提交安全評估，並對 AI 生成內容加上數位浮水印以防詐騙與假訊息，而且 2025 上半年推出的國家級 AI 需求數量，與前三年加總相當，顯示監管加速。文中亦提到中國公司在 AI Safety Index 的表現下降以及中國的動機包含維持政治穩定又想促進經濟成長，控制輸出與對齊社會價值在技術與治理上皆是棘手議題。在全球治理工具上，現有的 OECD AI Principles、Council of Europe 的 AI Framework Convention 多偏向非強制或缺乏執行，WAICO 若有國際適當介入也許可以借鏡 IAEA 的模式，運用各國同意限制、並接受檢查以便互相驗證是否遵守達到國際 AI 治理目標。

XAI 法律規範治理

《黑色止血鉗》其中一位主角是來自東城大學的佐伯教授，以技術高超聞名，能在心臟持續跳動的情況下直接進行縫合修補手術，手法穩定且判斷精準。整個日本僅有佐伯教授一人能完成此類手術。與之形成對比的，是來自維新大學的西崎教授。他所關注的並非個人技藝的極致，而是透過現代科技建立可被複製與推廣的醫療技術體系，使更多醫師能夠達到相近的手術水準。因此，他投入機器人手臂、人工智慧與 3D 列印等技

術研發，試圖以科技取代高度依賴個人能力的治療模式。西崎教授以國立大學教授身分前往東城大學這所地方型地區醫院，親自觀摩佐伯教授的手術操作，驗證其技術實力。觀摩過程中，西崎教授迅速判斷該名病患正是其所研發手術技術的理想適用對象，然而這項技術目前仍僅能由他本人執行。示範手術是在大型會議廳同步直播，會議室聚集了數百名醫師與觀摩者，然而示範手術進行到一半時，另一名原本安排於隔日接受手術的病患，突然發生急性主動脈剝離。該名病患本身已有心臟疾病史，病況在短時間內急速惡化，屬於若未即刻手術便可能致命的危急狀態。監測畫面顯示，病患心率明顯加快，血氧飽和度下降至 87%，已進入高度危險範圍。在當下的設定中，能夠處理此類病灶的醫師，只有佐伯教授一人。然而，他此刻正身處另一間手術室，進行示範開刀，無法立即親自執刀。佐伯教授隨即指揮其團隊資深醫師在另一間手術室執行主動脈置換手術。透過即時監視畫面判斷，病患確實出現主動脈剝離。然而，受限於遠端監視的視角，畫面僅能顯示主動脈狀況，心臟本身潛藏的損傷並未被即時察覺。當主動脈置換完成、體外循環恢復、心臟壓力上升後，原先未被辨識的心臟破損部位，在血流重新灌注下立刻引發大出血，病況瞬間惡化，陷入高度生命危機。此時，佐伯教授正進行示範手術，現場團隊一時之間無人敢中斷示範流程回報危機，病患生命岌岌可危。就在關鍵時刻，另一位具備頂尖手術能力的醫師渡海臨危受命介入救援。他並未依賴任何科技輔助，而是憑藉精準的臨床判斷與純熟的手術技巧，迅速確認出血源頭，完成關鍵縫補，成功挽救病患性命。劇中的對照呈現出兩條截然不同卻同樣迫切的醫療需求：一端是仰賴極少數職人級醫師的高風險手術能力，另一端則是希望透過 AI 與科技，讓關鍵判斷與技術得以被理解、被支援，甚至被複製。

AI 治理中的 Transparency Gap，也就是人工智慧應用中的「透明度落差」，並非來自技術不足，而是源於技術解釋與法律需求之間的本質差異。當前對 AI 的討論，特別是在醫療領域，多半聚焦於演算法與模型表現，強調預測準確度，以及模型如何透過神經網路進行推論。XAI 正是在這樣的脈絡下出現，試圖說明模型運作方式與影響預測的因素。然而，這類解釋多半停留在技術層次，無法直接回應法律與治理所關切的問題。

法律所要求的透明性，並非單純理解模型如何計算，而是涉及責任歸屬、程序正當性、可質疑性與監督可能性。當決策結果產生爭議時，是否能釐清誰應負責、決策是否符合規範、權利是否受到保障，才是治理的核心。若僅以模型特徵貢獻或因子權重作為解釋，往往無法支撐這些法律判斷。因此，透明度落差的關鍵，不在於是否已經提供了解釋，而在於這些解釋是否能被納入法律制度中使用。若將 XAI 誤認為只要補足技術說明即可符合法規，反而可能導致治理流於形式，無法真正落實責任追究與制度監督。這正是當前 AI 應用中，必須正視的核心治理挑戰。

以目前醫療 XAI 的發展來看，技術層面的透明性，與臨床上可理解、可使用的解釋需求之間，仍存在明顯差異。現階段最蓬勃發展的，多半是技術層次的模型透明性，重點放在呈現特徵貢獻、決策依據與敏感度分析，主要用途在於模型驗證、偏誤檢查與技術審查，其評估核心在於是否忠實反映模型行為。然而，這類技術解釋本身，並不等同於臨床說明。從治理角度來看，醫療 XAI 可清楚區分為兩個層次：一是技術層的模型透明性，二是面向使用者的解釋性透明。前者提供的是模型內部的中介性資訊，強調忠實性與穩定性，在法律上屬於事前必須具備的基礎條件；後者則必須將模型輸出轉化為臨床語言，結合病況、指引與情境，生成可被醫師與病人理解、可實際使用的最終解釋，這才是法律意義上的關鍵層次。也正因為這兩個層次在目的與功能上並不相同，才會引出責任配置的問題。技術層透明性主要對應的是模型開發與導入前的事前責任，而臨床可理解的解釋，則關係到決策發生後的事後責任、責任釐清與審查。就目前的 XAI 發展而言，這兩種責任層次仍難以被完整銜接，技術解釋與法律所需的可用解釋之間，依然存在尚未被填補的落差。要滿足這些應用問題層面關鍵除 XAI 技術之外必須建立將 XAI 技術與法律規範治理系統整合的整體架構。這個架構的核心概念，是以法律所要求的透明化解釋義務與治理目標，反向檢視現有 XAI 技術是否具備相應能力。首先，必須系統性地整理不同 XAI 方法的設計目的與技術特性，釐清各類解釋在忠實反映模型行為、可理解性與可審查性上的能力差異，並透過文獻回顧，辨識技術解釋與法律期待之間的落差。這一步的目的，不是證明技術是否先進，而是確認其是否能被不同利害

關係人理解，並支持事後審查與質疑。其次，法律解釋的需求與治理目標必須被清楚界定，包括受影響者是否能理解並質疑決策、主管機關是否具備監督與稽核的基礎，以及責任是否能被明確歸屬。這些要求，並非技術層面自然會滿足，而必須作為 XAI 設計與評估的起點。在此基礎上才能進一步進行 XAI 合規性評估，將一般法規治理通則與特定產業或情境的法律要求納入考量。歐盟目前正嘗試透過這種方式，重整其 AI 法規架構，雖然在產業發展速度上相對保守，卻強調制度信任、責任追究與人類實質控制。最終，法律導向的 XAI 共識，必須建立在跨法律與工程領域的合作之上。未來的 XAI 發展，不能僅以演算法可解釋性作為唯一指標，而必須同時回應法律對權利保障、責任歸屬與治理可行性的實質需求。只有在這樣的前提下，XAI 才可能真正成為可被信任、可被監督、也可被問責的醫療人工智慧基礎。

XAI 的發展由「模型如何被解釋」逐步走向「解釋如何被轉化為法律上可使用的說明」。這正是接下來幾個主題要聚焦的核心方向。從架構上來看，人工智慧原本的黑箱決策，透過 XAI 進入模型解釋層，產生可解釋資訊；但這些資訊本身仍不足以構成法律意義上的「解釋」。真正關鍵的是，這些資訊是否能被不同的利害關係人理解與使用，包括一般民眾、主管與監理機關、企業與組織，進而支撐監督、審查與責任判斷。在技術層面，XAI 必須滿足事前評估所需的指標，例如模型的穩定性、是否忠實反映模型行為，以及資訊呈現的精簡度與一致性；但在法律層面，解釋的評估標準則截然不同，重點放在是否清楚明確、是否有助於行動理解、呈現形式是否適切，以及在實務上是否具有可用性。也因此，XAI 的法律意義不在於技術上能否萃取更多模型資訊，而在於能否針對特定對象，產生情境化、可理解、可檢視的解釋。只有當解釋能支撐盡責監督、責任追究與系統性評估時，XAI 才真正完成了從「技術透明」走向「治理可用」的轉化。

關鍵問題就在於：如何透過「需求導向」，縮小 XAI 技術與法律規範之間的透明度落差。這並不是單靠強化演算法本身就能解決的問題，而是必須從法律與應用場域的實際需求出發，重新校準 XAI 的設計方向。當人工智慧逐漸以代理人的形式介入決策，例如自動搜尋、判斷與推薦，它不只是處理資料，而是在回應並推斷使用者的意圖與價值

取向。在這樣的情境下，所謂的可解釋性，已不再只是模型如何運作，而是這些運作結果是否符合法律對透明性、可質疑性與責任歸屬的要求。因此，XAI 的應用理念必須同時整合演算法能力與法律目標。具體而言，應先釐清應用領域中的法律解釋需求，再反向界定 XAI 所需具備的可解釋屬性，並將這些需求回饋至技術評估與設計過程中。透過這樣的循環，才能逐步縮小技術透明與制度期待之間的落差。以歐盟法規為例，其在資料保護與 AI 治理架構中，明確將權利保障、責任追究與監督機制作為核心目標，並要求解釋義務能實際支撐這些制度運作。這種以法律需求作為起點、再回饋至技術評估的模式，正是技術與法律得以契合的關鍵。

要讓人工智慧在法律與倫理層次上具備可接受性，XAI 必須同時滿足清楚且可被檢驗的解釋性特徵。在事前層次，解釋必須具備忠實性與穩定性，也就是能真實反映模型的判斷依據，且在資料微幅變動時維持一致表現；同時，解釋內容必須精簡且具一致性，使不同利害關係人對相同病況的理解不致出現相互矛盾的解讀。最後，解釋本身必須具備可理解性，能讓目標對象快速掌握其意涵，而非僅停留在技術描述。進入法律層次後，評價重點則進一步轉向信任性與責任性。解釋是否符合醫療常識、是否在出現爭議時具備合理說服力，直接影響其可信度。同時解釋必須可被追溯與稽核，能清楚連結模型版本、輸入資料、輸出結果與實際採納的決策理由。效率亦成為不可忽視的要件，特別是在醫療情境中，解釋必須能在合理時間內產出，而非延誤臨床處置。此外真正具有法律與臨床價值的 XAI，還必須具備可行動性與互動性。解釋不僅要說明風險來源，更應指向下一步可行的處置方向，協助醫師進行治療決策。同時，解釋內容應能依角色調整，讓醫師、病人與管理者各自取得符合其需求的資訊層次，實現個人化與可溝通的解釋。

XAI 醫師臨床應用風險與決策

研究聚焦兩個核心問題，一是醫師認為在使用醫療 AI 系統時，哪些因素最可能導致臨床錯誤；二是醫師對 AI 潛在錯誤的理解，如何影響其對法律風險的認知。研究以 10 位具臨床經驗的住院與資深醫師為對象，透過線上訪談，請受訪者針對假想醫療 AI 情境，討論其使用方式、錯誤預期及法律因應策略。研究請參與者設想自己是大型醫院

的主治醫師，必須與醫療團隊合作，並實際使用 AI 系統來協助決定病人的用藥與處方。透過系統介面，參與者需要說明任務需求、溝通對象與預期目標，同時反思自己在使用 AI 時所依賴的資訊類型與原因。訪談重點聚焦於四個面向：對 AI 介面建立信心的來源、不放心的風險與補強資訊、AI 系統可能犯的錯誤，以及醫師在使用 AI 時可能產生的人為錯誤，藉此理解 AI 與臨床決策的互動關係。整合了病人的基礎資料、主訴、過去病史與生命徵象，讓醫師能快速掌握整體健康狀況。系統同時提供 AI 治療建議，列出不同藥物選項、劑量、風險與是否符合臨床指引，並保留醫師接受、修改或拒絕的決策權。此外，平台連結實證準則、學術文獻與支援資源，協助醫師在繁忙臨床中做出更有依據、可追溯且安全的治療決策，強調 AI 作為輔助而非取代醫師判斷。

醫師對 AI 輔助工具的使用感知中，系統設計本身即存在重大風險。臨床資料無法充分反映醫病互動、個別差異與心理社會背景等難以量化的關鍵資訊，使得 AI 難以產出完整決策依據。此外，資料缺乏族群與地區多樣性代表性，也可能導致決策偏誤。AI 無法即時更新最新醫學指引，以及輸入錯誤或資訊不完整，都是潛藏的系統性錯誤來源，值得高度關注。AI 在臨床初期導入階段，醫師須逐步建立信任並仰賴額外參考資料進行決策。然醫療現場常面臨資料雜亂與不完整，亦難以即時預期系統行為，造成不確定性。過度仰賴 AI 建議與記錄系統異常反應，亦可能影響後續決策準確性。進入長期使用階段後，醫師若未持續更新專業知識，恐將忽略病史或背景因素，導致錯誤判斷，並且在壓力情境下依賴 AI 建議，風險更難控管。

隨著 AI 工具普及，醫師在臨床決策中若過度依賴 AI，恐被視為未盡專業判斷義務，成為法律追責依據。研究指出，醫師應主動採取防禦性紀錄策略，與專科醫師諮詢並妥善記錄病史。AI 系統引導下的記錄與建議接受與否，也可能成為日後訴訟爭點。醫師普遍期盼由專業審查機構進行 AI 決策稽核，確保決策透明且具防禦力。醫師在臨床使用 AI 工具時，關鍵在於保持批判性思考與主動判斷。研究顯示，絕大多數醫師傾向先檢視病人資訊、再依據既有知識判斷，最後才視 AI 建議作為補充依據。唯有堅持臨床判斷優先原則，並將 AI 視為參考而非替代，方能視為低風險使用行為。AI 建議應作為輔助

性提示，非可質疑的唯一依據，才能真正落實醫療安全與信任。

以上內容將在 **2025 年 12 月 24 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