



星球永續健康線上直播

AI 醫療治理(2)

AI 安全閘門治理沙盒(AI Airlock)

2026 年 7 月 8 日

近年英國提出「AI 安全閘門治理沙盒 (AI Airlock)」概念，透過模擬測試、真實世界驗證及法規學習等機制，協助人工智慧產品在安全且受控的環境中導入臨床應用。此治理模式兼顧技術創新與病人安全，為智慧醫療的永續發展提供重要方向。本週將進一步探討 AI Airlock 的核心理念、運作機制，以及其在智慧醫療領域中的實際應用案例。

健康科學新知

約旦折衝美國-伊朗商議停火細節：「和而未穩」

美伊雖透過 6 月 17 日的「14 點諒解備忘錄」暫時降低衝突，但目前仍未形成完整和平協議。備忘錄原意在於停止美國、以色列與伊朗之間的軍事行動、恢復荷莫茲海峽通航，並於 60 天內就核計畫、制裁、凍結資金及區域安全展開談判。然而停火後仍有貨輪遇襲，美伊亦再度發生軍事摩擦，顯示局勢依然脆弱。6 月底杜哈會談未能進入美方期待的直接高層談判，伊朗否認近期與美國直接接觸，只同意派員赴卡達討論凍結資產與協議執行；卡達也澄清，美方特使主要是與調停方會面，而非直接會見伊朗官員。目前談判主要針對三項爭議。首先是荷莫茲海峽通行與管理權。美國要求伊朗不得干擾國際航運，確保商船自由通行；伊朗則主張其與阿曼對海峽具有管理權，甚至認為協議承認其在談判期間安排航線的權限。由於荷莫茲海峽承載全球約五分之一的石油與天然氣運輸，任何襲擊、扣船或航線限制，都可能迅速推升油價、航運成本與全球通膨。伊朗則聚焦於凍結資金與制裁鬆綁。伊朗希望取回存放於卡達等地的受限制資金，並恢復石油出口收入。美國則要求資金釋放須與伊朗履約進度掛鉤，且用途應受到監督，避免被用於重建飛彈、無人機或其他軍事設施。卡達強調自身僅扮演金融中介，是否撥款仍取決於美伊談判。批評者擔心，美國若過早解除制裁或允許伊朗以美元取得石油收入，



可能削弱原有施壓效果，並協助伊朗恢復軍事能力。核計畫與區域安全則為另一爭執點。備忘錄雖重申伊朗不得發展核武，並提及可能在國際原子能總署監督下降低濃縮鈾濃度，但未明確要求伊朗完全停止鈾濃縮，也未處理飛彈、無人機及代理武裝等問題。這使沙烏地阿拉伯、阿聯、巴林與科威特等海灣國家感到不安，擔心美國為儘快結束戰爭而接受一項缺乏區域安全保障的安排。尤其美國先前已與海灣國家建立大規模經濟、軍事與人工智慧投資合作，區域盟友期待華府在對伊朗談判中優先維護海灣和平與發展，但目前相關保障仍不明確。卡達與阿曼雖努力維持斡旋管道，避免海上衝突再度升高，美國希望藉由停火與談判穩定油價、降低通膨壓力，伊朗則持續以荷莫茲海峽、凍結資金與核計畫作為談判籌碼。停火後港口、市場與漁業雖逐步恢復，部分船隻仍遭扣押，貨輪也需等待通行許可，正常航運尚未完全恢復。若美伊無法在 60 天內釐清海峽管理權、資金用途、制裁解除條件、核計畫限制與區域安全保證，現有安排可能只是下一輪危機前的暫停，而非中東局勢穩定的終點。

俄國-烏克蘭擴大長程攻擊：「以攻制攻」

俄羅斯上週末對烏克蘭首都基輔發動戰爭爆發以來最嚴重的夜間空襲之一，動用火箭、巡弋飛彈、彈道飛彈、高超音速飛彈及近 500 架無人機，總計約 570 個空中目標，造成至少 27 人死亡、91 人受傷。儘管烏軍宣稱攔截多數來襲武器，多波次、複合式攻擊仍使部分防空系統不堪負荷，住宅、旅館、電信設施及其他民用基礎設施受到破壞。俄方聲稱攻擊對象為軍事與軍工設施，但相關說法尚未獲得獨立證實。此次空襲反映俄烏遠程打擊持續升級。烏克蘭近來加強攻擊俄羅斯本土及俄占克里米亞的煉油廠、能源設施、港口、雷達與軍事物流據點，意圖削弱俄軍燃料供應、補給體系及戰爭財政。俄方宣稱攔截數百架烏克蘭無人機，但克里米亞仍出現燃料短缺、停電及民生壓力，俄方任命的當地政府甚至宣布經濟緊急狀態。俄羅斯總統普丁亦罕見承認部分地區存在加油站排隊、油品供應困難及克里米亞存油不足等問題，但淡化其嚴重程度。烏克蘭的戰略是透過國產長程無人機將戰爭壓力帶回俄羅斯，迫使莫斯科分散防空、維修、能源調度與後勤資源，並藉燃料短缺、物價上漲及民生不便，削弱俄羅斯社會對戰爭的承受能力。



俄羅斯則以規模更大、強度更高的飛彈與無人機攻擊回應，集中打擊基輔等城市，試圖增加烏克蘭的平民傷亡、防空負擔與政治壓力，形成相互報復與持續升級的循環。歐洲國家因此面臨戰略兩難。一方面，歐洲支持烏克蘭打擊俄軍後勤與能源供應，並主張加強防空援助及對俄制裁；另一方面，烏克蘭攻擊深入俄羅斯境內，也提高戰爭外溢、俄方報復及北約被捲入衝突的風險。歐洲必須在提升烏克蘭反擊能力與避免衝突失控之間取得平衡，同時因應能源市場波動、武器與情報支援所帶來的政治壓力，以及愛國者攔截彈等防空資源可能不足的問題。遠程攻擊升級亦使停火談判更加困難。俄羅斯不願放棄其長程打擊優勢，並拒絕限制縱深攻擊；烏克蘭則持續強調反擊俄羅斯本土軍事及能源設施的必要性。加上俄方要求烏克蘭讓出東南部領土，而烏克蘭拒絕割讓領土，雙方立場仍存在根本差距。整體而言，俄烏戰爭已由前線作戰進一步延伸至城市空襲、無人機戰、能源破壞、防空消耗、制裁與心理戰，戰爭更具長期消耗性，歐洲軍援、防空供應與升級風險控管，將成為影響後續戰局的重要因素

Anthropic-三星 AI 供應鏈成型：「算力自主」

AI 領先企業 Anthropic 傳出正與三星電子評估合作開發自有 AI 加速器。該計畫擬採用三星 2 奈米製程與先進封裝技術，顯示 Anthropic 欲減少對 Nvidia 等通用硬體的依賴，打造更貼合其 Claude 模型架構需求的專用晶片。三星、SK 海力士與美光先前已透過融資成為 Anthropic 的策略夥伴，顯示產業正轉向算力與封裝整合的策略共生模式。對三星而言，若能承接此訂單，將是其晶圓代工業務挑戰台積電、驗證先進製程實力的關鍵轉折。專家分析，AI 競爭已延伸至硬體底層的全棧對抗。Anthropic 透過自研晶片旨在優化成本、降低延遲並確保供應鏈自主。

亞洲 AI 晶片經濟榮景影響：「盛衰並行」

全球 AI 晶片需求激增，台積電、三星及 SK 海力士成為供應鏈核心，帶動台韓兩國出口與股市齊揚。然而，亮眼的數據背後卻隱藏著「K 型經濟」隱憂，科技紅利正加劇社會財富分配不均。南韓與台灣因全球生成式 AI、資料中心及高頻寬記憶體需求快速成長，帶動三星、SK Hynix 與台積電成為全球晶片供應鏈核心，推升出口、股市、



市值與企業獲利。然而，AI 晶片繁榮主要集中於少數大型科技企業、股東與高薪專業人才，未能平均擴散至傳統製造業、服務業、中小企業及一般受薪階層，形成科技產業快速成長、其他部門相對停滯的「K 型經濟」。南韓雖受惠於記憶體晶片出口，但地方商圈、零售業及中小企業仍面臨需求疲弱與倒閉壓力，三星不同部門間的獎金爭議，更反映 AI 紅利在企業內部的分配不均。台灣同樣高度依賴台積電與先進晶片出口，但半導體就業人口有限，多數非科技業勞工難以直接受益，並持續承受低薪、房價及生活成本上升壓力。石化、鋼鐵、電池與汽車零組件等產業，也受到全球需求減弱及中國競爭影響。與過去由大量中小企業共同支撐的經濟模式相比，今日 AI 晶片成長更加依賴少數企業、尖端技術與龐大資本，可能使南韓與台灣形成科技部門與傳統產業並存的雙軌社會。未來除持續投資半導體與 AI 基礎建設外，亦須加強區域發展、中小企業扶持、職業訓練、薪資改善、房市治理及企業分紅制度，才能避免 AI 紅利集中於少數企業與資產持有者，並使科技成長轉化為更具包容性的社會利益。避免社會矛盾因技術變革而深化。

TriNetX 大數據平台研究洪流：「慎思明辨」

醫療平台 TriNetX 因提供大量匿名病歷數位資料與便捷的分析工具，使相關論文從 2020 年起迅速成長，2025 年已近 2,700 篇。這類按鈕式分析快速成長引發嚴重偏誤與學術誠信危機。缺乏流行病學訓練者易產出誇張結論，如宣稱減重藥物能大幅防癌，卻忽視統計偏誤。部分論文疑似誤用 AI，描述平台根本不具備的分析功能，顯見研究者未理解工具限制。這類低品質研究恐誤導醫師對藥物或補充品效果的判斷，損害照護品質。對此，美國醫學院協會擬調整評選機制，從強調論文數量轉向學術實質品質。專家呼籲，應推動相關訓練課程，確保研究者在追求產出效率時，能對資料偏誤與方法限制保持高度警覺。

學術平台反制 AI 濫用維持學術品質：「信實為本」

arXiv 宣布，若投稿手稿被發現含有生成式 AI 產生的「幻覺式參考文獻」，或殘留明顯的 AI 對話語句，作者將面臨一年禁止投稿的處分。arXiv 電腦科學部門主席指



出，這項政策旨在應對日益嚴重的 AI slop 議題。處分期滿後作者仍需證明其作品已被具信譽的同儕審查場域接受，方可恢復投稿權限。目前各平台策略不一：PsyArXiv 與 SocArXiv 已祭出永久禁令，而 bioRxiv 則傾向開發偵測技術而非直接封禁。專家提醒，若 AI 錯誤內容被重複抓取訓練，恐破壞學術流通。研究者即便使用 AI 輔助，仍須對引註與事實負起最終審核責任，避免損害個人學術信譽與平台的可信度。

歐陸積極復興科研領導地位：「聚才立勢」

在美中科研競爭加劇下，歐洲正力拚轉型為全球科學超級強權。歐盟推出九億歐元的 Choose Europe 倡議，旨在將歐洲定位為穩定且開放的研究避風港，吸引頂尖人才移居。目前歐洲論文產出已超越美國，僅次於中國。為強化實力，歐盟擬將新一輪「Horizon Europe」預算提升至 1750 億歐元以上，並納入日、韓、加等夥伴，建立全球科研網絡。儘管面臨研發投入較美、中低及轉化創新能力不足的挑戰，歐洲仍積極強化制度整合。隨著俄烏戰爭後國防科研支出激增，歐洲正試圖在戰略自主與學術開放間取得平衡，開創美中之外的第三種科研聯盟模式。

AI 安全閘門治理沙盒(AI Airlock)

《黑色止血鉗》第二季第四集的劇情與 AI Airlock 法律治理與醫療監管議題非常契合。劇中的心臟外科醫師天城雪彥醫術高超，能夠完成許多高難度的冠狀動脈手術，因此在醫療界享有極高聲譽。然而，天城醫師雖然擁有拯救病人的能力，卻有著極具爭議的行事風格。他會以賭局決定是否接受病人的手術，病患必須先與他進行一場賭博，贏得賭局後，才能獲得接受手術的機會。他甚至認為，唯有得到「神的祝福」的人，才值得接受他的治療。除此之外，病患還必須支付高額的手術費用，甚至被要求付出一半財產作為醫療報酬。這種將醫療決策建立在運氣與財富之上的做法，雖然可能出於醫師個人的理念，但顯然已經違反社會對醫療公平性與可近性的期待，也無法被現代醫療體系與普世價值所接受。因此，劇中的醫療律師戶島和子率領醫療守護會前往醫院抗議，認為天城醫師的行為已涉及對病人的不當誘導甚至詐欺，並質疑醫療是否能夠由個人意志凌駕於制度與規範之上。抗議的場面一邊是東城大學醫院的高層與醫師團隊，另一邊



則是由醫療律師戶島和子所率領的醫療守護會與病人團體，雙方形成強烈對峙的局面。在社會輿論與訴訟壓力之下，東城大學醫院最終決定暫時取消天城醫師原訂的手術安排。醫院擔心，如果爭議持續擴大，不僅會影響醫院聲譽，也可能衝擊未來的發展與重大計畫，因此選擇先平息風波。對於天城醫師而言，這無疑是一個重大打擊，因為他始終認為，只要能救病人，自己的技術就應該被認可，而不應受到外界爭議影響。戶島律師離開醫院時，意外遇見了在東城大學醫院擔任護理師的女兒美和，母女兩人因而相認。然而，就在相認不久，戶島律師突然因身體不適而昏倒。經過檢查後發現，戶島律師罹患嚴重冠狀動脈狹窄，其中右冠狀動脈甚至接近完全阻塞。原本醫師計畫採取傳統冠狀動脈繞道手術進行治療，但在手術評估過程中發現，她的冠狀動脈存在嚴重鈣化，使得繞道手術以及其他常規治療方式都難以進行，因此病情一度陷入僵局。戶島律師最終拒絕接受天城醫師的治療，轉而前往維新大學醫院尋求其他治療方案。維新大學醫院當時正積極發展一套名為「艾卡諾 (Elcano)」 AI 手術輔助系統，希望透過人工智慧協助處理高難度心血管手術。戶島律師自願成為第一位受試者，接受這套 AI 系統所規劃的新治療方案。該系統結合病人的影像資料、臨床資訊以及最新醫學證據進行分析，並搭配一種促進血管新生的新藥，嘗試先誘導新的血管生成，再進行後續的繞道手術。根據 AI 系統的評估，這項治療策略的成功率可達 89%。維新大學的菅井教授與野田醫師公開提出所開發的 AI 手術輔助系統 Elcano，未來有可能讓任何醫師都能完成過去只有少數天才外科醫師才能完成的困難手術。然而這樣的想法也讓病患家屬感到不安，模擬環境中的成功並不代表真實臨床一定安全，因此戶島律師的女兒也質疑，母親是否成為 AI 技術進入臨床前的第一隻白老鼠。另一方面，世良醫師則不斷練習，希望能靠自己的努力克服技術上的不足。然而天城醫師卻提醒他，醫療真正需要的不只是技術，醫師必須理解疾病、理解人體，面對疾病多變特徵也必須具備創造新的治療方法與突破困境的能力，是為每位病人設計最適合的治療方案。手術過程中，野田醫師依照 AI 系統的指示一步一步完成操作，整個手術進展十分順利，也似乎證明了即使缺乏頂尖外科醫師的經驗，只要依循 AI 的引導，仍然有機會完成高難度手術。然而就在手術剛結束時，由於新藥



副作用，戶島律師突然發生心室顫動，病情瞬間陷入危急。此時，AI 系統並未能即時提供有效的處置方案。因為它所設計處理的是標準化的手術流程，而不是面對突發、複雜且超出訓練情境的臨床危機。當病人的狀況偏離既定流程時，AI 的能力也隨之受到限制。就在關鍵時刻，世良醫師想起天城醫師過去訓練時反覆演練的一種特殊繞道手術，決定接手搶救戶島律師。世良醫師根據眼前病人的血管狀況與自己的臨床判斷選擇調整手術策略。這個決定最終成功挽救了病人的生命，也凸顯出 AI 與醫師之間真正輔助關係是合作。AI 可以提供資訊整合、風險預測與決策建議，但醫療最終仍需要醫師根據病人的即時狀況做出臨場判斷。醫療不只是標準化流程，更是結合想像力、技藝以及對病人負責的專業決策。

台灣目前正在推動的人工智慧基本法，重要的核心是人工智慧治理的七大原則。第一個是永續發展與福祉，其次則包括人類自主、隱私保護與資料治理、資安與安全、透明與可解釋、公平與不歧視，以及最後的問責機制。這七大原則其實都是在回答同一個問題：當 AI 真正進入社會、進入醫療、開始影響人類決策之後，要如何確保 AI 是可信、可管理，而且是對人類有益的。其中最困難的是當 AI 發生錯誤時誰要負責？AI 的決策是否可以被解釋？資料隱私是否受到保護？不同族群是否會受到不公平待遇？以及人類是否仍然保有最終決策權。而要回答這些問題，光靠技術開發完成之後再進行監管，往往已經太晚。因此，在 AI 真正進入臨床與社會之前，需要一個能夠讓技術、法規與臨床共同學習的機制，這就是所謂的治理沙盒（Regulatory Sandbox）。

英國提出的 AI Airlock，正是全球第一個針對醫療 AI 醫材（AIaMD）所設計的監管沙盒計畫，由 MHRA、NHS 以及 DHSC 等單位共同推動。核心目的是在可控環境中提前發現問題、驗證風險，並協助產品開發者、監管機關與臨床使用者共同學習如何讓 AI 符合治理要求。AI Airlock 扮演從技術創新走向法規落地橋樑。透過產品測試、法規缺口分析以及臨床情境驗證，協助 AI 在真正進入醫療現場之前完成醫療應用必要的安全檢驗與治理準備。建立民眾信任、保障病人安全，同時兼顧醫療創新與公共利益。

AI Airlock 在智慧醫療應用設計中，核心理念是以真實產品與真實情境作為出發



點，直接面對人工智慧醫療產品於臨床應用過程中可能遭遇的監管挑戰。其採取真實產品導向的策略，利用實際的 AI 醫療器材或原型產品，探索產品在開發、驗證與部署過程中可能涉及的法規與治理議題，而非僅停留於實驗室環境中的技術驗證。AI Airlock 強調多方利害關係人的共同參與，包括監管機關、臨床專家、技術開發者、法規專家及學術研究人員，透過跨領域合作，共同評估人工智慧技術進入醫療體系後可能產生的效益與風險。第三，AI Airlock 採用情境化測試機制，依據產品特性設計測試計畫與驗證流程，並透過真實世界證據（Real-world evidence）的蒐集，評估產品於臨床場域中的安全性、有效性與可行性。第四，AI Airlock 重視法規學習與回饋機制，將測試過程中所發現的問題與經驗，轉化為未來政策修訂、技術指引與制度設計的重要依據，使監管制度能夠隨著技術發展持續演進與調整。因此，AI Airlock 的核心目標，在於促進技術創新、維護病人安全以及建立公眾信任之間取得適當平衡，進而支持智慧醫療技術的安全導入與永續發展。

目前 AI Airlock 第一階段計畫涵蓋四項智慧醫療應用案例，分別反映醫療人工智慧於不同應用場域所面臨的治理挑戰。第一項案例為 Philips Healthcare 的 AutoImpression，主要應用於放射科報告摘要與自動生成，其所對應的監管議題包括生成式 AI 內容的正確性、完整性以及驗證機制之建立。第二項案例為 AutoMedica 的 SmartGuideline，應用於臨床指引問答與工作流程支援，其主要挑戰在於大型語言模型可能產生幻覺（hallucination）與不確定性，因此透過檢索增強生成（Retrieval-Augmented Generation, RAG）技術降低錯誤資訊風險，提升決策可靠性。第三項案例為 OncoFlow，主要應用於腫瘤學資料分析與多專科團隊（MDT）決策支援。除了模型效能外，其核心監管議題在於人工智慧決策的可解釋性，以及系統建議與臨床證據之間的對應關係。第四項案例為 Newton's Tree 所開發的 FAMOS 系統，主要針對放射 AI 的安全監測，持續監控模型漂移（model drift）、資料品質下降以及自動化偏誤等問題，以確保模型在實際部署後仍能維持穩定與可靠的表現。

以 Philips AutoImpression 為例，其主要應用於放射科報告的自動生成與摘要撰



寫。隨著醫療影像檢查量持續增加，放射科醫師面臨龐大的判讀與報告撰寫負擔，因此如何利用生成式 AI 協助完成初步診斷與結論段落，成為目前重要的發展方向。然而，AI 所生成的報告是否能夠達到與放射科醫師相近的品質與一致性，是否能夠在降低人力負擔的同時維持臨床準確性與可靠性，則成為其導入臨床前必須驗證的重要議題。因此，國際間已陸續採用非劣性研究設計（non-inferiority study），評估 AI 生成報告是否能達到不低於專科醫師報告品質的標準。另一項重要案例為 OncoFlow，其主要應用於腫瘤學多專科團隊（Multidisciplinary Team, MDT）決策支援。在腫瘤治療中，無論是乳癌、大腸癌或其他惡性腫瘤，治療決策通常需要整合病理科、放射科、外科及腫瘤內科等多個專業領域的意見，共同制定最適合病人的治療策略。此類決策支援系統往往需要結合大型語言模型、檢索增強生成（RAG）技術以及外部知識資料庫，同時整合結構化與非結構化資料，以完成治療建議與臨床決策支援。因此，如何確保系統的準確性、可解釋性以及臨床可信度，便成為導入過程中的重要監管議題。這兩項案例充分反映出 AI Airlock 在治理沙盒中的核心角色，也就是透過早期驗證機制，釐清人工智慧技術是否符合現行法規與治理原則的要求，並協助辨識潛在的風險與規範缺口。其中，醫療人工智慧與一般資訊系統最大的不同，在於其最終將直接影響病人的診斷與治療，因此必須建立在充分的臨床證據基礎之上。無論是效能驗證、安全性評估，或是真實世界情境下的臨床表現，都需要透過系統性的驗證與評估程序加以確認。AI Airlock 的價值，正是在技術正式進入臨床與監管程序之前，提供一個可以提前學習、提前驗證以及提前修正的機制，使開發者、臨床使用者與監管機關能夠共同釐清技術風險與法規要求，並補足現行制度可能存在的證據缺口與治理缺口。

AI Airlock 建立了三種不同層次的測試環境，將人工智慧可能面臨的監管問題放置於不同成熟度與風險情境中進行驗證。第一個層次為模擬 Airlock（Simulation Airlock）。此階段主要透過工作坊、專家討論與情境模擬，探索人工智慧系統可能發生的失效模式、監管疑慮以及潛在風險。例如，當模型出現異常決策、資料遭到竄改，或模型超出原始設計範圍時，可能對病人安全與醫療流程造成何種影響，均可於此階段預



先進行模擬與評估。

沙盒測試結合合成資料 (Synthetic Data)、數位雙生 (Digital Twin) 及模擬情境設計，觀察模型的穩定性、準確性與韌性，作為早期產品與概念驗證的重要基礎。第二個層次為研究/虛擬 Airlock (Research / Virtual Airlock)。此階段利用真實資料或高擬真模擬資料，在虛擬環境中觀察模型行為，評估其準確性、穩定性與可重複性。此階段的目的是，在於確認人工智慧系統是否具備進一步進入臨床驗證的條件，同時辨識模型可能存在的偏誤、漂移或效能下降等問題。第三個層次則為真實世界 Airlock (Real-world Airlock)。此階段將人工智慧系統部署於真實醫療場域之中，但在適當監督與風險控制機制下，不直接影響病人的醫療照護決策。其主要目標在於評估人工智慧系統與臨床工作流程的整合能力、安全性以及實際運作表現，並透過真實世界證據 (Real-world Evidence) 的蒐集，持續優化模型與監管策略。值得注意的是，這三種場域並非彼此獨立，而是可以依據產品生命週期與風險程度進行單獨或組合運用，以回應不同階段的監管需求與治理挑戰。

以 OncoFlow 為例，其治理設計與台灣人工智慧基本法所提出的治理原則具有高度對應性。首先，在人本原則方面，OncoFlow 所提供的是多專科團隊 (MDT) 決策支援，而非取代醫師決策。人工智慧負責提供資訊整合、風險評估與治療建議，最終的臨床決策權仍保留於醫師與醫療團隊手中。其次，在安全原則方面，人工智慧系統必須先經過 Airlock 沙盒驗證與風險評估，確認其安全性與可靠性後，方能進入正式臨床應用。在公平原則方面，則需持續評估不同資料來源所可能產生的偏誤與代表性問題，以避免模型對特定族群產生不公平的決策結果。在可解釋性方面，目前人工智慧治理已逐漸朝向可解釋人工智慧 (Explainable AI, XAI) 的方向發展，例如利用 SHAP、LIME 或 RAG 等技術，說明模型為何做出特定風險預測或治療建議，提升模型決策的透明度與可信度。在透明原則方面，OncoFlow 強調完整揭露治療建議所依據的臨床證據、指引來源以及推論過程，使醫師能夠理解人工智慧決策背後的理由，而非僅接受模型輸出的結果。在問責原則方面，目前國際間普遍強調 Human Override 或 Human-in-the-loop 的概念，



也就是人工智慧系統必須保留人類介入、修正與否決的權限，避免將最終責任完全交由人工智慧承擔。最後，在持續改善原則方面，人工智慧系統需要透過真實世界資料(Real-world Data) 與持續回饋機制，不斷修正模型偏差、提升模型效能，形成持續學習與持續改善的循環。因此，AI Airlock 的核心精神並非限制人工智慧的發展，而是在促進創新、維護病人安全以及建立社會信任之間取得平衡，進而支持人工智慧技術安全且負責任地進入臨床醫療場域。

AI 安全開門治理沙盒實例

人口結構變遷使疾病防治與治療需求快速增加，影像判讀需求同步提升，然醫療專業人力提供難以如同需求面迅速提高，隨著醫師判讀負荷增加，報告產出及複檢排程均受到影響。在此情境下導入 AutoImpression 可輔助初步診斷草擬，協助擷取影像重點、產生初步報告，提升作業效率與判讀一致性，最終診斷仍由醫師確認。乳房影像判讀導入 AI 輔助初步診斷為例，AI 辨識右乳不規則腫塊及局部群聚性微小鈣化，風險分級屬 BI-RADS 4，並建議安排乳房超音波及切片檢查。醫師完成 AI 生成報告複核後可補充病史資訊並簽署正式報告。該初步診斷生成系統為輔助工具，最終診斷與醫療決策仍由醫師主導。影像初步診斷輔助 AI 工具 AutoImpression 發展採雙軌流程設計，首先以 MIMIC-IV 去識別化資料建立結構化放射報告，再透過大型語言模型 (LLM) 依提示詞產生初步診斷內容。系統後續結合 AI 自動化評估、放射科醫師人工審查及量化指標分析，持續檢視報告品質、一致性與臨床合理性，建立完整的品質管理流程。測試顯示合成放射報告在 ICD-10 分布與真實資料相近但仍有差異，BioBERT 相似度落在 0.6 至 0.8。LLM 裁判與放射科醫師評估具有一致性，在準確性、清晰度與有據性表現較佳，但完整性仍需加強。於 AI Airlock 模擬情境評估顯示利用 AI 驗證可能產生循環風險，未來實際運用須建立更明確法規與人工監理機制。

Ai Airlock 另一評估實例為腫瘤治療智慧決策輔助系統 OncoFlow，旨在解決臨床腫瘤照護中資料分散、整理耗時與判讀不一致等問題。透過將病理、影像、病歷與治療指引等資訊結構化，系統可協助多專科團隊更快掌握病人狀況，聚焦關鍵治療決策。此



工具有助提升討論效率、降低延誤風險，讓腫瘤治療更精準且即時。於 Ai Airlock 模擬情境評估結果顯示結合 LLM、RAG 與臨床規則，可提升腫瘤治療決策輔助的準確性與可解釋性。Rule-based 系統在臨床資料擷取上可達約 90%準確度。治療決策匹配建議再加入預先定義指引作為 RAG 後，ESMO/NICE 指南檢索準確度達 92%，SACT protocol 治療規範指引匹配達 100%，且幻覺率低(3.4%)。可解釋分析 SHAP 顯示遠端轉移與 HER2 狀態為關鍵因子。以沙盒環境評估 Philips AutoImpression 與 OncoFlow 兩項智慧醫療輔助工具目前皆屬早期階段。AutoImpression 聚焦放射科報告生成，採生成式 AI/LLM 技術，主要挑戰在於合成資料生成與驗證，測試顯示其相似度約 0.6 至 0.8，部分醫師仍難分辨真實與合成報告。OncoFlow 則結合 LLM 與 RAG，應用於腫瘤 MDT 決策支援，資料擷取達 90%、指南檢索 92%、SACT 匹配 100%。兩項工具皆尚未發表同儕審查成果，未來仍需最佳實務、專家驗證與獨立評估。

以上內容將在 2026 年 7 月 8 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：rl1847030@ntu.edu.tw