

星球永續健康線上直播

星球健康週新知 &

專題: 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

2025-12-24

CHE團隊：

陳秀熙教授、許辰陽醫師、陳立昇教授、嚴明芳教授、林庭瑀博士、
劉秋燕、羅崧瑋、林家妤、陳虹彥

資訊連結:
<https://www.realscience.top/7>



星球永續健康線上直播



<https://www.realscience.top/4>

Youtube影片連結: <https://reurl.cc/gWjyOp>

漢聲廣播星球永續健康:

https://audio.voh.com.tw/TW/Playback/ugC_Playback.aspx?PID=323&D=20240615

新聞稿連結: <https://reurl.cc/no93dn>

本週大綱

- 星球健康新知 (2025 / W50)
- XAI 法律規範治理
- XAI 醫師臨床應用風險與決策

星球健康新知

2025 / W50



盟國承諾安全保證 俄烏和平談判現轉機：和平懸峙



美國與歐洲盟國承諾向烏克蘭提供強有力安全保證 但烏東領土問題仍未解決



美國特使威特科夫積極折衝四方
提出烏東設立特別經濟區構想



澤倫斯基表態放棄入北約換取安全保證
德國總理表示若俄方破壞停火
盟國將肩負擊退俄軍責任

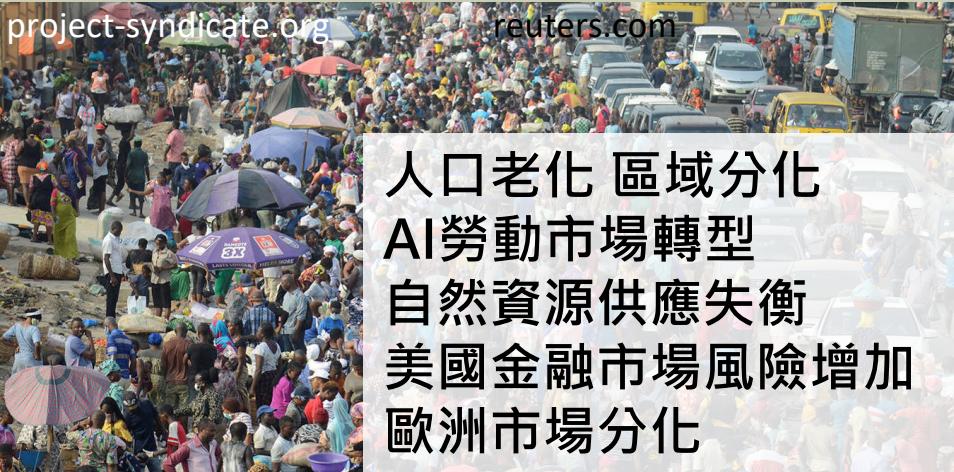
芬蘭警告俄烏和平協議將造成俄國於北約邊境部屬武力北移 造成新威脅



芬蘭總理歐爾波



主要經濟體展現轉型韌性應對結構性壓力：穩中趨緩



人口老化 區域分化
AI勞動市場轉型
自然資源供應失衡
美國金融市場風險增加
歐洲市場分化

五項結構性趨勢將改變2026全球經濟



英國就業市場放緩 英格蘭銀行
正考慮降息避免經濟狀況惡化

美國就業動能轉弱 聯準會已發佈降息 投資人預測2026年第一季美國將維持中性策略



美國聯準會主席鮑爾

歐元區工業成長回升
歐洲經濟展現韌性但前進動能仍弱



reuters.com



reuters.com



全球經濟貿易地緣座標再定位：實力定錨



馬克宏訪問北京反映歐洲面臨的困境



馬克宏提出歐-中關係再平衡
對全球經濟安全生產-供應重整必要性

美英科技合作協議因制度分歧陷僵局



英韓貿易完成協議 將強化雙邊經貿關係



bbc.com

氣候變遷失控警訊：印度洋周邊極端氣候衝擊 治理失速



- 三個罕見熱帶氣旋重創印度洋周邊多個國家的城市與村落
- 豪雨與洪水造成上千人死亡、數百萬人受災，規模接近海嘯等級
- 氣候變遷海溫上升與印度洋偶極子變化使氣旋更容易生成降雨更極端
- 環境破壞如森林砍伐、採礦等加劇災害衝擊

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

- 極端降雨與洪水在全球大城市常見，但與死亡率關聯未充分研究。
- 研究者分析孟買氣候與生命資料建立氣候變化對死亡風險加劇效益預測模式

脆弱族群首當其衝，複合氣候危機事件進一步放大死亡風險

- 季風期極端降雨佔總死亡 >8%，主要 (85%) 影響貧民地區居住民眾
- 強降雨 5 週內死亡率：兒童 (<5 歲) +5.3%，女性 +3.1% (男性 兩倍)。
- 持續影響：雨後數週因疾病傳播、醫療中斷、生活條件惡化，死亡風險增加持續
- 降雨 + 高潮汐水位阻礙排水：30 mm/小時降雨增加 5 週死亡 0.6%，漲潮期更劇。
- 氣候變遷將進一步加劇此風險，預測2030 年海平面升 5-15 cm，可增加 7-21% 雨相關死亡。

全球暖化極端氣候健康政策應對啟示

- 整合防災：現代化排水基礎設施、廢物管理、安全飲水、氣候韌性設計
- 早期預警：考慮地形 (orography) 影響，使用機器學習預測
- 擴大感測器IoT網路 + 群眾提供的淹水數據，提升即時應變力
- 衝擊監測：全球評估需納入雨相關死亡，而非僅限溫度

Subimal Ghosh, Nature, 2025



AI x 生物學 AlphaFold 重塑蛋白質科學：範式轉移

Ewen Callaway, Nature, 2025

- 長期以來，蛋白質三維結構的解析依賴 X光晶體學 與 冷凍電子顯微鏡
- 這些方法成本高且常受樣品品質與實驗條件限制 → 使得結構解析往往耗時數月至數年

AI系統 _ AlphaFold (由Google DeepMind開發)

- 只需幾分鐘 ~ 幾小時就能給高可靠度 預測的蛋白質三維結構
- AlphaFold2：在蛋白質結構預測競賽中表現幾乎等同實驗結果

1 加速科學發現

能有效指引實驗設計 → 顯著縮短生物學研究所需時間

2 降低研究門檻

程式碼與資料的全面開放 → 全球研究者皆能進行高階蛋白質研究

3 擴散跨領域影響

用於科學發現的人工智慧

蛋白質結構與動態分析

用於蛋白質結構與工程的人工智慧

AI / 計算式藥物研發

工業酵素生物催化

酵素 / 微生物生物合成

AI 蛋白質設計與工程

用於生物分子預測的深度學習

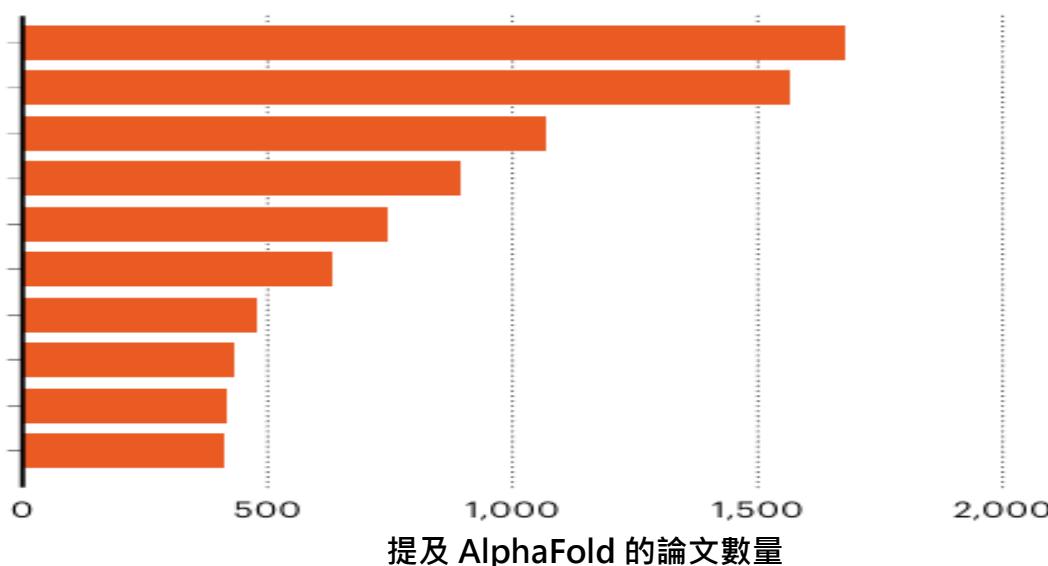
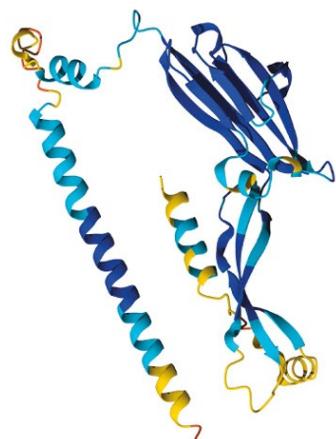
細菌蛋白質與細胞組織

癌症機制與精準腫瘤學

4 改變實驗模式

為傳統分析的重要輔助

➤ 讓實驗更快、更有效



提及 AlphaFold 的論文數量



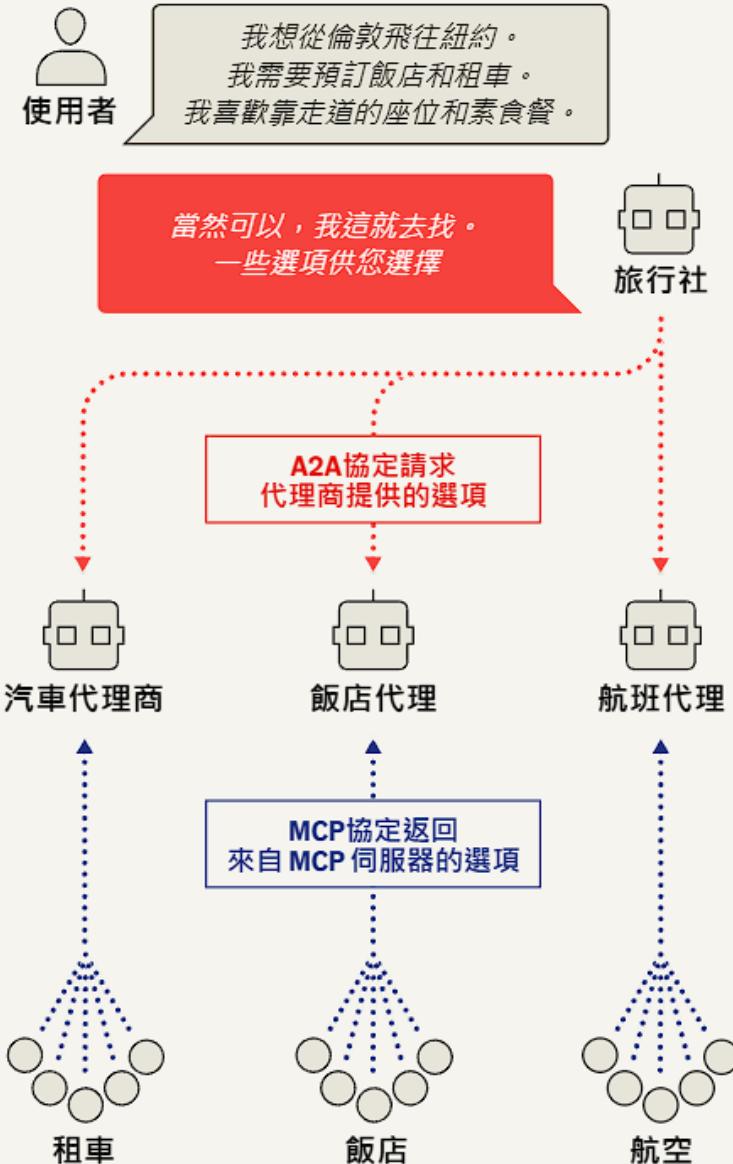
網路互動轉型 次世代智慧網路模式：行動轉譯

未來的網路互動將逐步轉向由AI智慧代理人
理解使用者意圖後主動規劃與協作

economist.com

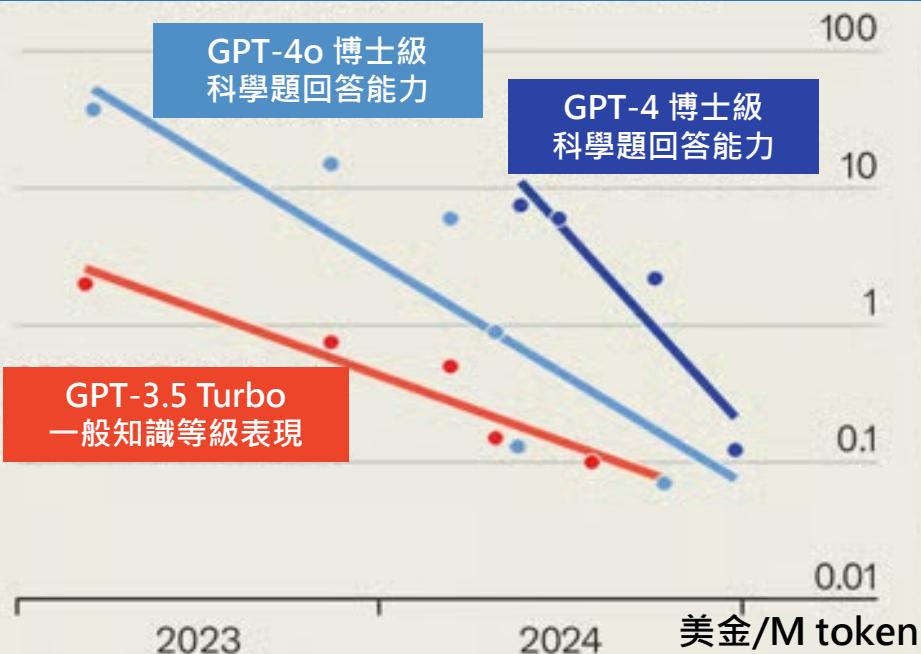


透過共通的通訊與存取協定
代理人能提供跨平台、跨服務的自動化數位服務
成為新一代網路架構的重要組成



AI Token用量成長與產業價值迷思：量質脫鉤

晶片與AI技術快速進步回答專業問題
單位Token成本迅速降低
Token用量 ≠ 模型能力與需求



語模供應商定價策略
差異極大
個語模Token耗用 ≠ 公司價值

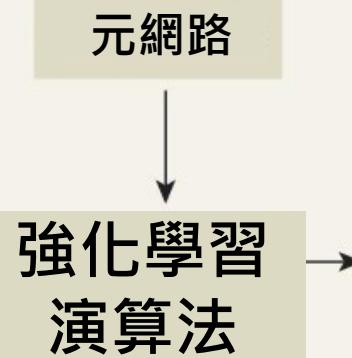
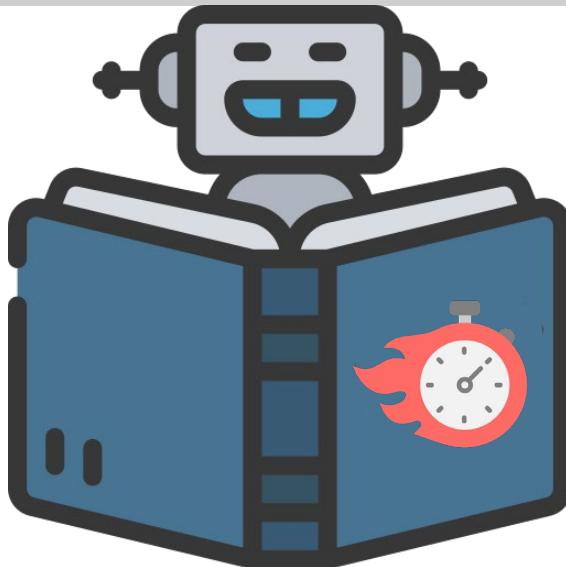


- AI市場普遍以Token用量反映產業需求與預期價值
- Token耗用成長≠AI需求、獲利能力等產業重要指標
- 以錯誤指標衡量產前景可能會造成網路泡沫時代錯估與投資失利

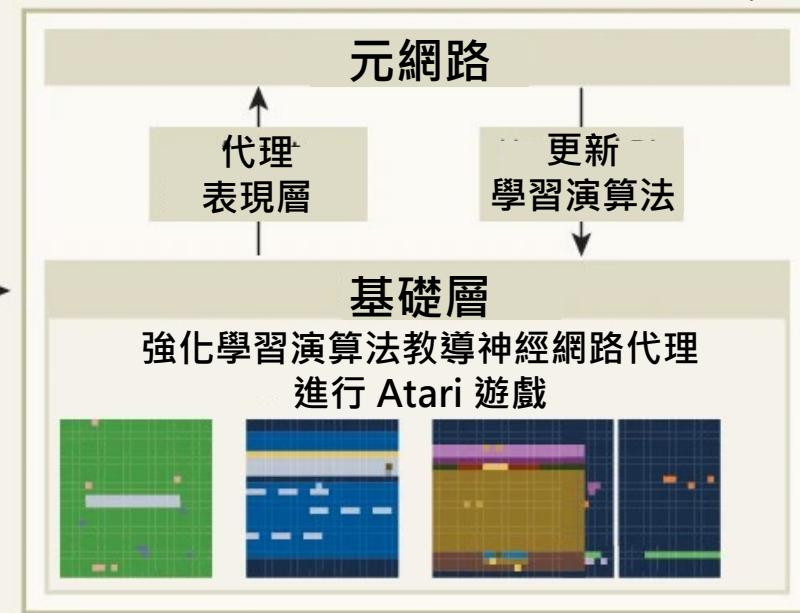


自我改進強化學習人工智慧: 人機共鑄

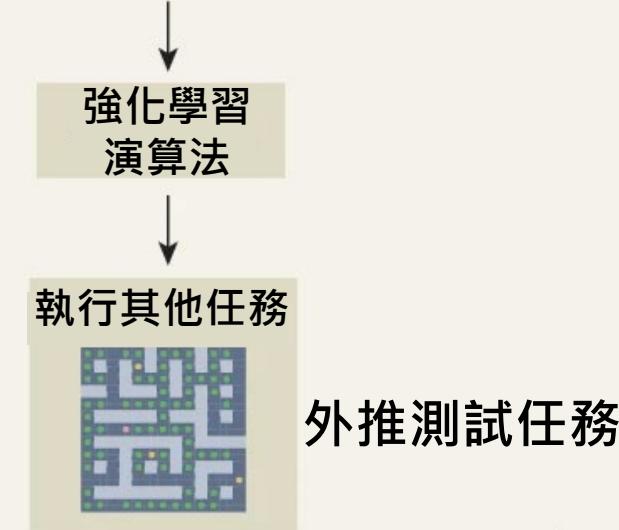
AI以強化學習
發現更有效率演算法



Nature 648, 283-285 (2025)



- 透過回饋機制設計並持續自主優化AI演算法
- AI 生成強化學習演算法在外推性測試任務中展現良好泛化能力
- AI 自主設計具潛力，但同時伴隨部署應用風險與挑戰

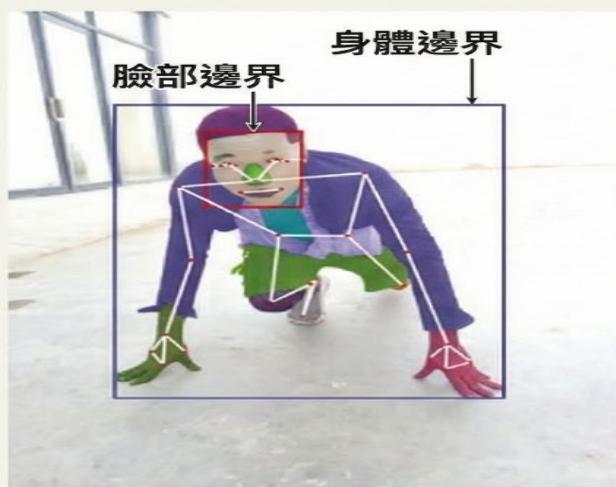




人像資料倫理示範 FHIBE資料集: 以人為本

Walter J. Scheirer, Nature, 2025

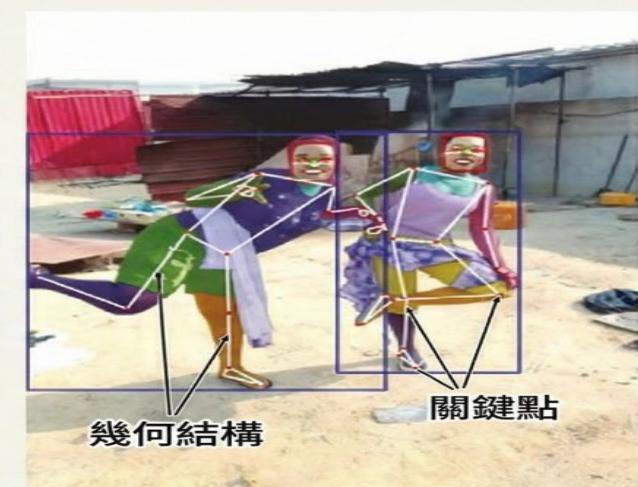
以FHIBE展示倫理蒐集與多元標註強化模型公平性評估



人口統計特徵
年齡
代名詞
祖籍
國籍



外貌特徵
膚色
眼色
髮型
面部特徵



影像內容
姿勢
互動
地點
環境特徵

以明確同意取代網路抓取
自我回報式身分標註

提供可檢視的公平性分析基礎
促進負責任的人像 AI 研究實務

- FHIBE規模有限，不適合作為大型模型的訓練資料
- 建置成本高、蒐集困難，短期內難以達到大規模擴展
- 需探索可兼顧自我認同與外部觀察的方式，使資料更完整反映實際使用場景

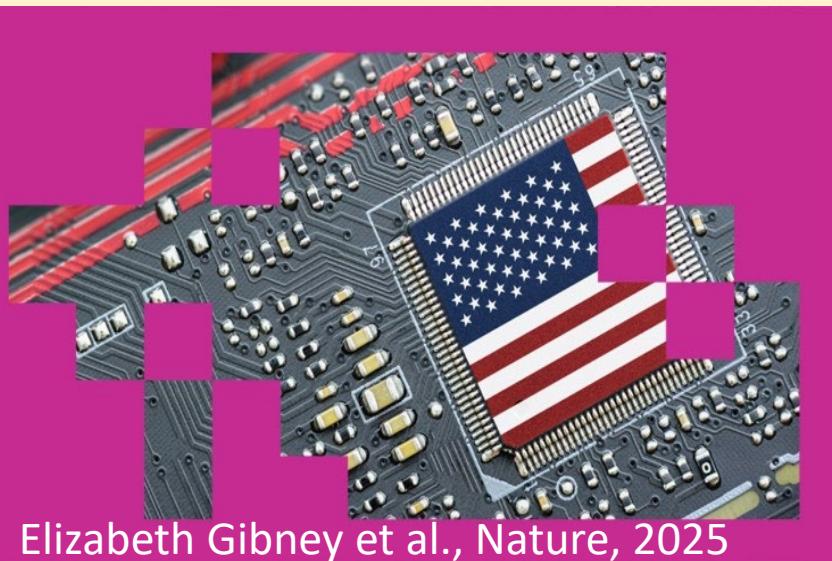


美國AI創世任務 風險治理與科技競逐：雄心壯志

「AI 創世任務」旨在鞏固美國 AI 全球領導地位，由川普總統提出。

提案主張：擴大美國科技能力，限制中國技術擴張，並建立 AI 自由市場陣營。

此與前任拜登政府強調倫理、風險管理、多元參與 AI 政策形成鮮明對比。



Elizabeth Gibney et al., Nature, 2025

風險

- 過度依賴大型企業，恐忽略民主治理與社會多樣性的 AI 需求。
- 學者憂心此舉將導致 AI 研究與發展規範政策受政治極化交互影響

AI 政策專家強調

制定 AI 策略須兼顧創新推進與安全保障，並納入多方利害關係人。倘若該任務實施，美國未來 AI 法規方向將大幅右傾，亦將對全球 AI 產業發展趨勢與治理架構造成影響。



AI治理需求與國際組織提案: WAICO

Nature 648, 2025

- 近年各國專家皆倡議AI治理原則，但尚未有AI監管國際組織。
- 中國提議成立世界人工智慧合作組織(World Artificial Intelligence Cooperation Organization, WAICO)

中國AI社會發展影響

- 政策層面推動晶片生產製造、AI演算法發展、工業化機器人生產製造
- 金融、貿易、社會、健康等多領域推展應用平台

AI治理推動要素

- 建立具體法律要件與安全評估測試項目與技術標準
- 階段式風險管推行規範
- 可追溯、可問責AI要件
- 安全評估準則共識

共同發展互補策略為健全AI治理共識關鍵

美國：

- 創新強，但監管零散、依賴企業自律

歐盟：

- 規範完整，落實速度緩慢、影響力還未成形

如果這些國家因為政治立場拒絕與中國互動：自己的權益會被忽略

XAI 法律規範治理

黑色止血鉗：心血管疾病致命危機

國立大學教授



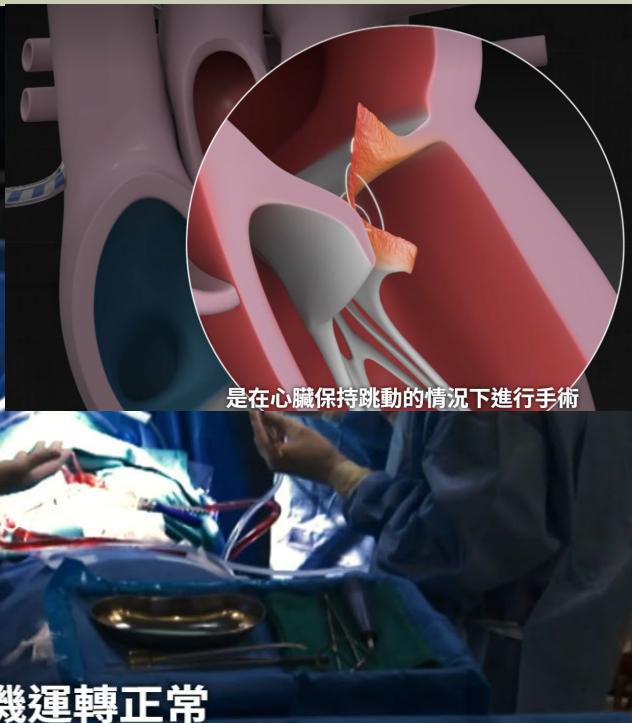
國立一流的大學醫院

怎麼會到我們這種地方小醫院呢？

地區醫院外科聖手



是在心臟保持跳動的情況下進行手術



人工心肺機運轉正常



- 東城大學佐伯教授技術高超，可在心臟跳動下直接縫合修補病灶，對心臟病變與療法判斷精準
- 佐伯教授競爭者西崎教授主張以科技取代高度依賴個人技術之療法

突發主動脈剝離 病患命懸一線



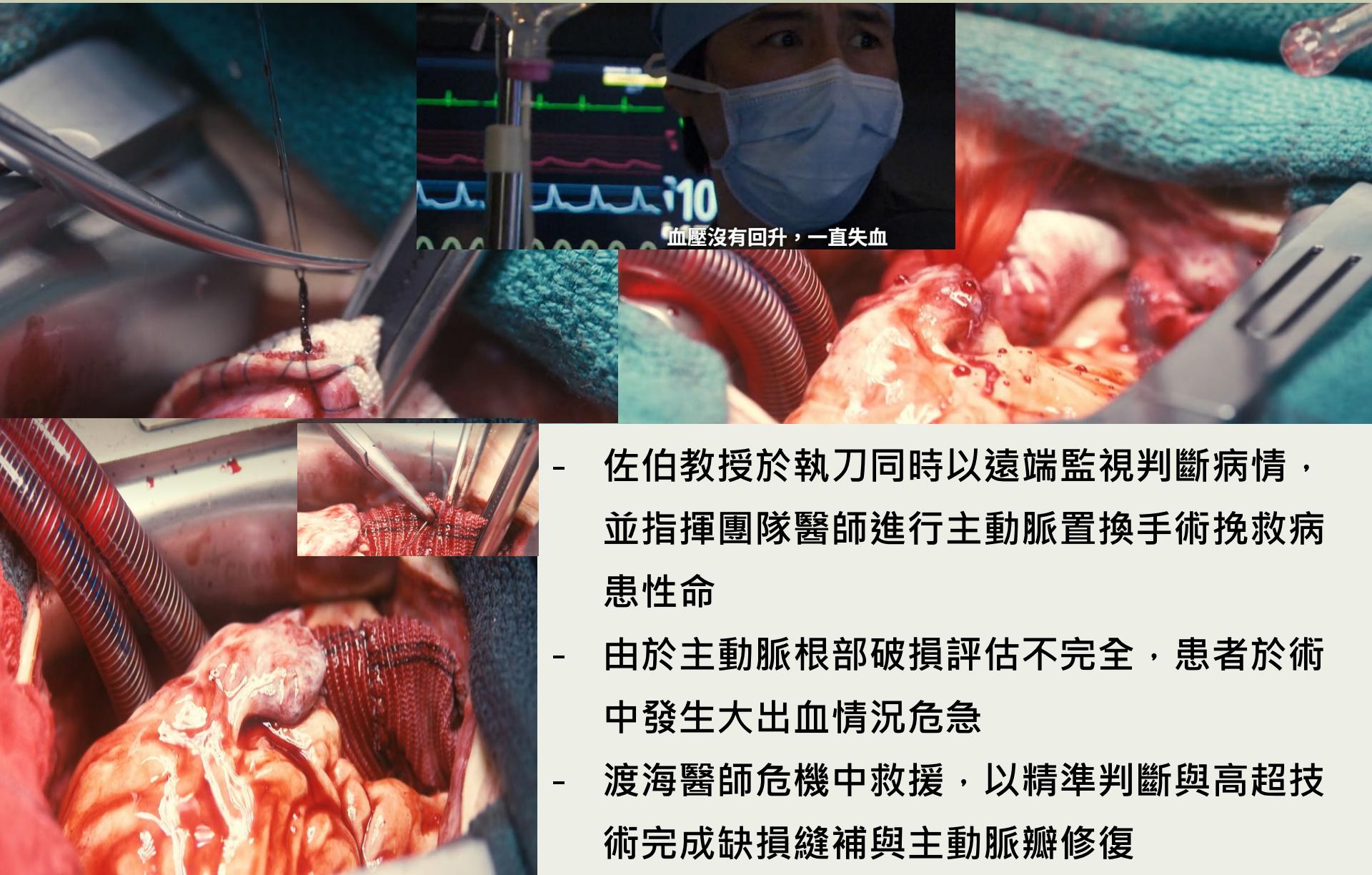
-去拿人工血管

-果然是主動脈剝離

升主動脈出現破裂

- 應西崎教授之邀，佐伯教授舉辦示範手術展現高超技術
- 期間原安排於次日開刀另一病患突然發生主動脈剝離，性命垂危需要立即進行手術治療

病灶智慧偵測 巧手修復病灶



- 佐伯教授於執刀同時以遠端監視判斷病情，並指揮團隊醫師進行主動脈置換手術挽救病患性命
- 由於主動脈根部破損評估不完全，患者於術中發生大出血情況危急
- 渡海醫師危機中救援，以精準判斷與高超技術完成缺損縫補與主動脈瓣修復



AI 應用治理議題: Transparency Gap

Sovrano et al., 2025



XAI 透明度落差: 技術 vs 應用法規

XAI可解釋性技術須滿足法律衡鑑所需求以支援權利保障、責任歸屬與監督

透明性
目的錯配

技術解釋與
法律理解落差

Transparency Gap
制度後果

技術說明與

法律應用需求錯配

- 技術透明偏描述模型行為
- 法律透明重制度與責任
- 兩者不契合形成落差

技術解釋未符合法律透明

- 技術性說明難以理解
- 缺乏責任判斷與程序保障
- 解釋內容呈現方式無法符合特定法規情境

落差致治理失衡

- 誤認XAI技術說明即可符合法規
- 缺乏合適解釋呈現判準
- 規範流於形式難落實治理



醫療XAI: 技術透明vs 臨床可理解解釋需求

Sovrano et al., 2025

Extractive / Inner XAI

技術層的模型透明

- 呈現特徵貢獻、決策依據與敏感度
- 用於模型驗證、偏誤檢查與技術審查
- 評估重點為忠實反映模型行為
- 不等同臨床說明

Surface / Explanatory XAI

臨床理解可理解解釋需求

- 模型輸出轉成臨床語言與溝通敘述
- 結合指引、病況與情境形成可用解釋
- 支援輔助醫師說明、病人理解
- 可於事後責任釐清與檢視

	Extractive / Inner XAI	Surface / Explanatory XAI
核心對象	模型行為	XAI使用者 (Explainee)
主要功能	萃取模型資訊	生成可用解釋
輸出性質	中介性資訊	最終解釋
評估重點	忠實性、穩定性等技術屬性	清晰透明、可用性等使用層面屬性
法律角色	前置必要條件	法律意義關鍵層次



XAI技術-法律規範治理系統整合架構

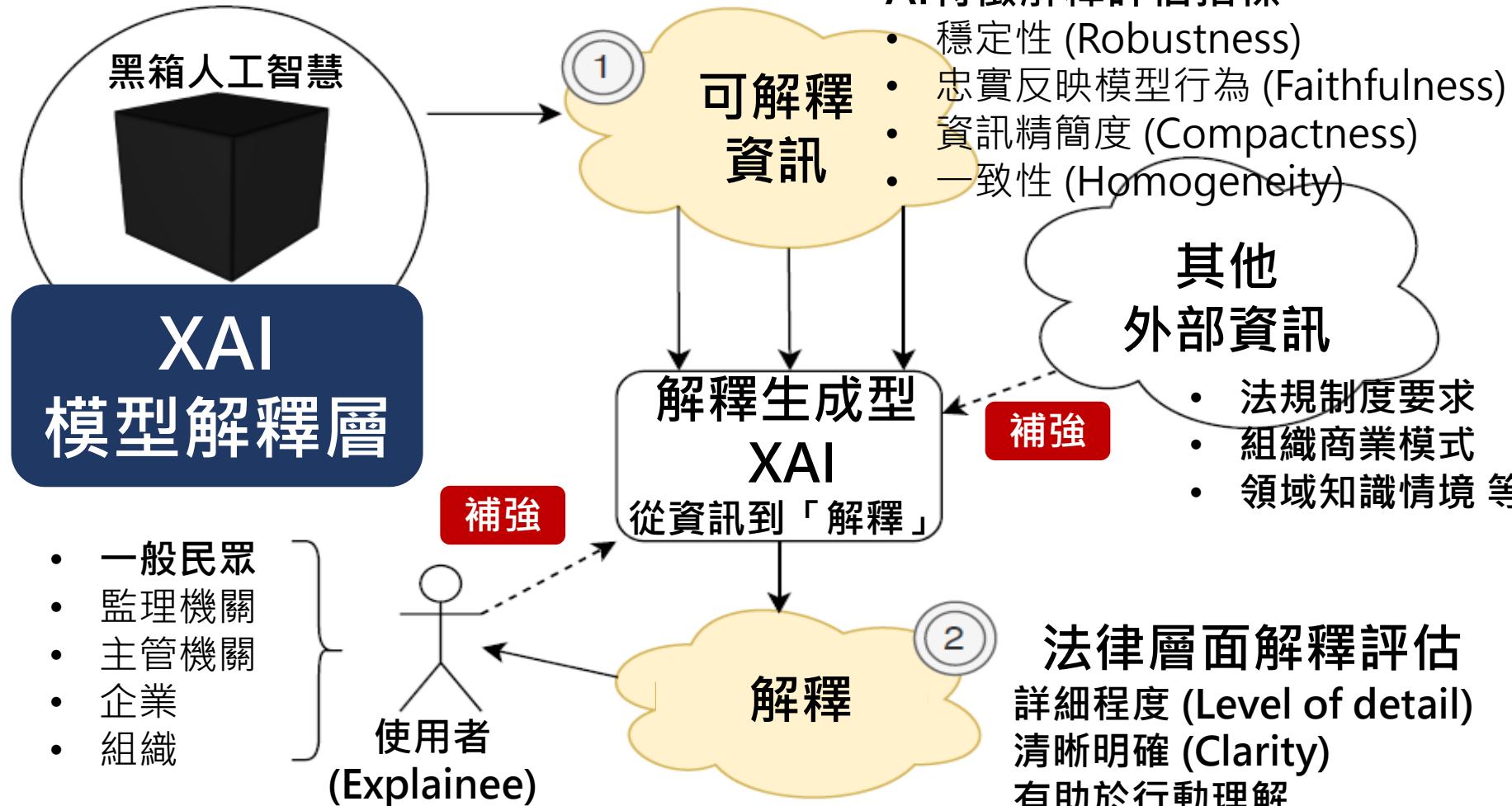
Sovrano et al., 2025

當前XAI技術能否滿足法律所要透明化解釋義務與治理目標？



XAI 解釋性透明轉化技術

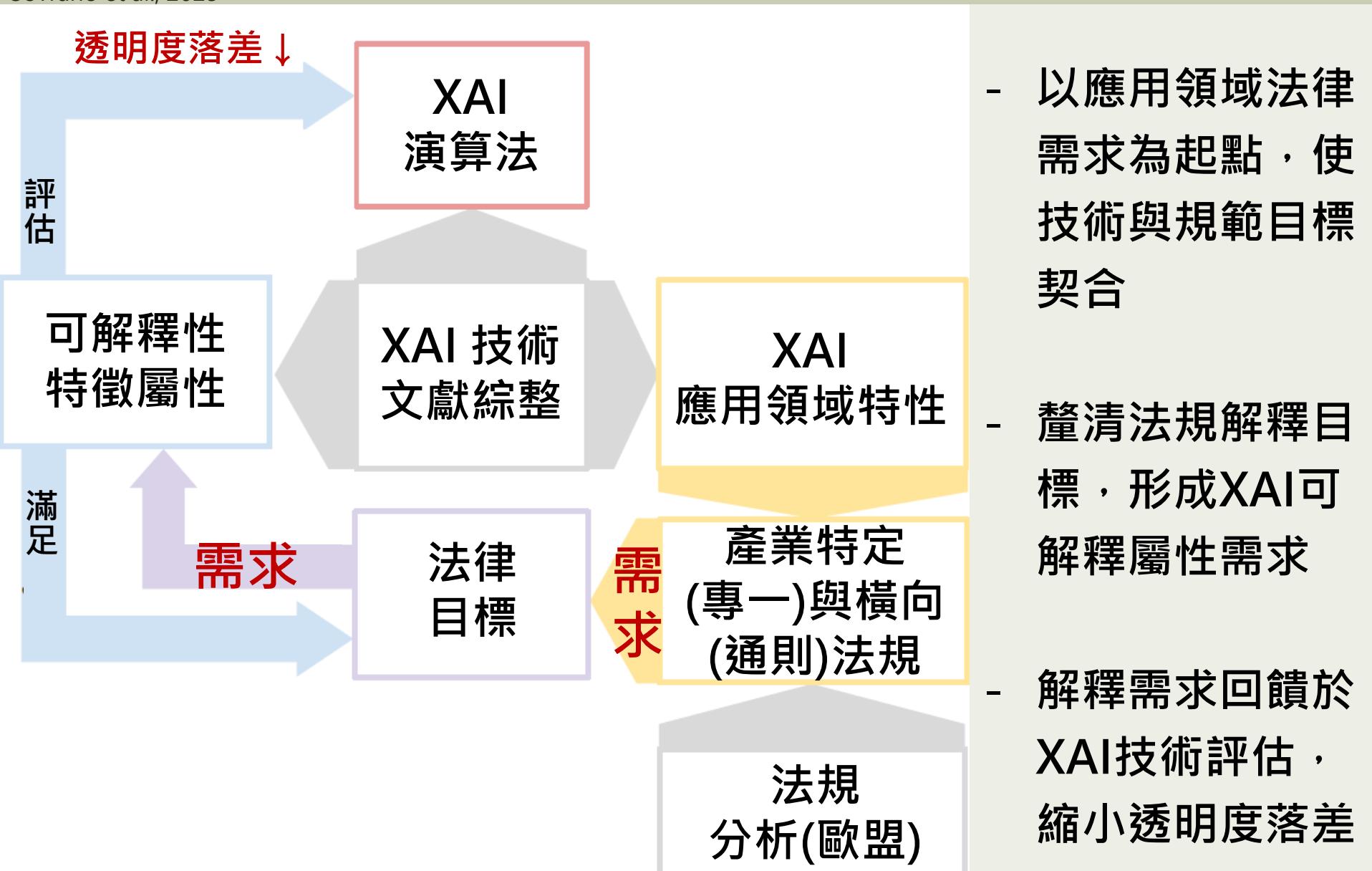
Sovrano et al., 2025



XAI的法律意義不在於技術萃取，而在能否產生對特定人(explaineer)可用情境化解釋

XAI 技術-法律契合架構: 以需求縮小落差

Sovrano et al., 2025





法律層次解釋性特徵 (1)

Sovrano et al., 2025

忠實性 (Faithfulness)

解釋須反映模型真實依據

如敗血症模型應著重於乳酸、呼吸速率，若以性別、住院天數解釋則屬失真

穩定性 (Robustness/Stability)

資料微幅變動解釋應近似

例如血鉀4.0到4.1呈現腎因性與心因性不同鑑別因素，表示模型穩定性不足

精簡度 (Compactness)

以少量關鍵因子清楚說明，例如僅列血氧、呼吸速率、浸潤範圍
優於增列其他影響極微之生化指標如血脂等

一致性 (Consistency)

相似病況應給一致邏輯，例如皆為高CRP發燒低血壓
但原因判斷分別出現感染與年齡不一致解讀，易引致使用者疑慮

可理解性 (Comprehensibility/Interpretability)

目標對象能快速理解；例如說明血氧低呼吸快較易理解，勝過權重向量呈現



法律層次解釋性特徵 (2)

Sovrano et al., 2025

信任性與可信度 (Trust/Plausibility)

解釋需符合醫療常識，例如CXR Grad-CAM熱區集中於肺浸潤區較可信，若在邊框或非合理解剖位置即使預測機率準確模型可信度亦降低

責任性 (Responsibility)

解釋須可追溯與稽核，如同步記錄模型版本、輸入、解釋與使用者採納理由

效率 (Efficiency)

解釋需在時限與合理成本內產出，例如急診分流系統回應時間應於數秒內提供簡要建議與主因，優於等待數分鐘後才提供詳細病情列表

可行動性與有用性 (Actionability/Usefulness)

解釋須指向下一步處置決策，例如指出乳酸下降至某範圍且平均動脈壓維持在門檻以上，預測風險將降低可幫助臨床輸液或升壓劑治療決策

互動性與個人化 (Interactive/Personalization)

解釋可依角色調整，例如醫師介面提供看趨勢、敏感度分析、因素排序等細節，病人介面則以白話說明風險來源與可能處置建議

XAI 醫師臨床應用

風險與決策

XAI 臨床應用醫師使用者關注議題

Mansi and Riedl et al., 2025

1. 醫師在使用AI系統時認為哪些因素會導致錯誤？
2. 醫師對潛在錯誤認知與使用AI系統時對法律風險的認知之間有何關係？

受訪對象：10 名具臨床經驗的醫師(住院醫師與資深臨床醫師)

訪談方式：在線上進行，平均約 58 分鐘

訪談內容：請受訪者針對假想醫療AI系統情境描述，討論其使用、誤差預期與法律相關策略。

AI臨床應用決策訪談情境

Mansi and Riedl et al., 2025

- 請參與者設身處地想像自己是一家大型醫院的主治醫師，需與其他醫護人員合作，為病患制定藥物劑量。
- 參與者必須使用AI系統決定患者處方。
- 參與者由系統介面，描述可能需要完成的任務、需要溝通的人員以及可能的目標。
- 要求參與者在解釋他們將使用的資訊類型、使用方式以及做出這些選擇的原因時，參考他們過去和現在的經驗。

訪談綱要

1. 您認為使用AI介面來照顧病人，有哪些方面會讓您感到有信心？您還需要哪些不同的或額外的資訊才能感到有信心？
2. 您在使用類似的AI介面時，有哪些方面會感到不放心？還有哪些資訊可以幫助您增強信心？
3. 你認為AI建議系統可能會犯哪些類型的錯誤？
4. 您認為您和其他醫生在使用AI時可能會犯哪些類型的錯誤？

數位醫療AI平台情境應用訪談

Mansi and Riedl et al., 2025

Jane Q.



MRN: 636978

DOB: March 3, 1953

Sex: Female

Insured – Medicare

Patient: (555) 555-5555

POA: Joe Q. (555) 555-5555

Husband (555) 555-5555

Allergens

Drugs ▾

Food ▾

Other ▾

Medications

Last updated 11/09/23. Open patient profile to edit.

Diabetic Tussin DM
10 mL, 1x per day, since 10/23/22

Nitroglycerin
0.4 mg, 3x per day, since 11/01/23

Amlodipine
5 mg, 1x per day, since 11/08/23

Treatment Options

Hover each bullet for an extended explanation for the factor.

DRUG #1

START DOSE: 10 mg
END DOSE: 100 mg
FREQUENCY: 3x/day, MR
INCREMENT: 25 mg/wk

- ▲ Avoids allergens
- ▲ No drug interactions¹
- ▲ Aligns with hospital guidelines
- ▲ Covered by insurance
- ▲ Minimizes side effects
- ▲ Aligns with societal guidelines²

Accept Modify Reject

AI recommendation

Recommended by patient's providers listed below:



¹Based on automated Lexicomp search.

²According to clinical practice guidelines from the American Association of Clinical Endocrinology <https://pro.aace.com/clinical-guidance/2022-clinical-practice-guideline-development-diabetes-mellitus-comprehensive>.

Patient Complaints 主訴

Based off most recent visit on 11/12/23. Click for medical record.

- Sharp pains
- Muscle weakness
- Difficulty performing everyday tasks

Documented Problem List

Select problem to see corresponding medical notes. Last updated 11/12/23. Click to edit list.

- Diabetes
- Peripheral Diabetic Neuropathy
- Coronary Artery Disease
- High Blood Pressure
- Asthma

Vitals

Last measured 11/12/23 @ 3:42 pm

140/95

Blood Pressure

See history

75 bpm

Heart Rate

See history

Recent Orders 過去病史

Lists order, date, and the specialty of the physician who made the order. Lists those tied in medical record.

- Autonomic Nerve Testing, 11/10/23
 Pain Specialist Access Note
- Electromyography, 11/09/23
 Pain Specialist Access Note
- Filament Testing, 11/05/23
 Pain Specialist Access Note
- Venous Sampling, 11/04/23
 Endocrinologist Access Note

More Labs

See Rest of Care Team

生命徵象監測

140/95

Blood Pressure

See history

75 bpm

Heart Rate

See history

26

BMI

See history

Patient Support 病患服務

Find Pharmacy

Used by other physicians for the following drugs

DRUG OPTION #1 DRUG OPTION #2 OTHER

CVS Bel Air

2176 Marsh Lane

Phone: (555) 555-5555

Patient Preferred Distance from patient address: 2.3 mi

Save my choice for this patient

Walgreens Bel Air

3923 Maridoe Road

Phone: (555) 555-5555

Distance from patient address: 3.1 mi

Target Josey Ranch

2392 Mallard Cove Drive

Phone: (555) 555-5555

Distance from patient address: 4.5 mi

See more pharmacies

Other patient support actions

Print Prescription

Export Medication Schedule for Patient

Export Care Summary

Contact POA

Contact Hospital Social Worker

Contact Hospital Translator

實證準則指引

I'm looking for...

UpToDate

PubMed

Hospital Guidelines

Suggested based on past searches

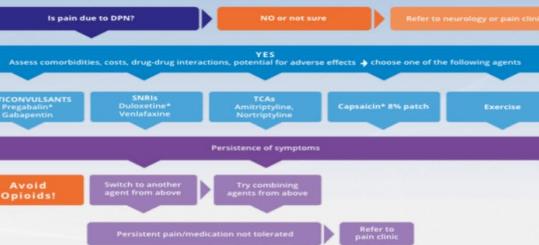
My Hospital's Guidelines

Up To Date: Management of diabetic neuropathy

PubMed Chincholkar (2020). Gabapentinoids: Pharmacokinetics, pharmacodynamics and Considerations for Clinical Practice

AACE Algorithm for DPN Treatment

ALGORITHM FOR TREATMENT OF DIABETIC PERIPHERAL NEUROPATHY



*FDA approved for treatment of DPN
DPN = diabetic peripheral neuropathy; DSNP = distal symmetrical polyneuropathy; FDA = Food and Drug Administration; SNRI = serotonin-norepinephrine reuptake inhibitor;
TCA = tricyclic antidepressant

Adapted from Pop-Busui, Blauth, et al. Diabetes Care 2017;40:136-154

Support Resources AI 決策輔助

AI Model Details

Message Patient Team
Go to hospital secure messaging system

Request Model Audit

On-Call Physician
Find contact info for on-call physicians

Nonurgent Technical Support
response time 1-2 business days
 help@utsrtn.ia

Emergency Technical Support
response time 10-20 min
 (555) 555-5555

醫師對使用AI輔助工具之風險感知 (1)

Mansi and Riedl et al., 2025

AI 系統設計相關風險

- 資料無法充分反映臨床實務中關鍵但難以量化的資訊
 - 例如過往臨床經驗、醫病互動、病人之社會或心理因素、個別病人差異，以及身體檢查結果
- 資料代表性不足，未能涵蓋病人族群多樣性或地區性差異造成建議決策錯誤
- AI建議無法即時配合最新醫學指引
- 數位病例輸入錯誤或訊息不完整

系統設計所呈現的風險

- | - 一、由系統實作所導致的錯誤
 - | - AI 開發者之專業能力不足或邏輯設計不當
 - | - 對人工智能系統適當應用情境之假設不正確
- | - 二、由資料所導致的錯誤
 - | - 資料無法反映或捕捉關鍵臨床互動
 - | - 資料代表性不足
 - | - 與醫學會或社會臨床指引存在差異
 - | - 放大醫療紀錄中既有的錯誤

醫師對使用AI輔助工具之風險感知 (2)

Mansi and Riedl et al., 2025

醫師實際使用AI系統時所產生風險

系統導入初期調適階段

- 調適階段需逐步建立對系統信任，並依賴外部臨床參考資料、系統開發者的專業背景，以及既有醫療實務規範進行建議採納判斷。
- 此過程中可能難以預期系統如何處理雜亂或不完整的資料，亦不確定自身應如何隨技術演進調整臨床決策與行動方式
- 如何依賴他人使用人工智慧建議及回報錯誤，以及如何記錄系統表現不佳或病人不良結果以協助後續修正系統亦存在不確定性
- 醫師同時擔憂在學習使用系統的過程中試誤行為可能對病人造成負面影響

系統長期使用過程中持續存在風險

- 未能充分行使其專業醫療知識與臨床判斷能力，思考範圍過度依賴建議、忽略病歷中的關鍵資訊、未將自身專業判斷納入決策過程
- 臨床繁忙情境下，直接預設採納人工智慧建議
- 此類風險被醫師視為具有高度臨床與法律風險，並可能影響病人安全與醫療責任歸屬



醫師對使用AI之法律歸責擔憂

Mansi and Riedl et al., 2025

- 過度依賴AI法律風險

- 醫師普遍擔憂未進行充分臨床思考即採納 AI 建議可能被視為醫療疏失並成為醫療責任追究依據
- 在高工作量與時間壓力下，此風險更為明顯

- 防禦性醫療作法與法律責任

- 遵循院內或專業指引被視為降低法律風險的重要策略
- 諮詢其他專科醫師並於病歷中記錄，有助於責任控管
- AI 系統導入後既有作法可能需重新調整

- 決策紀錄與責任歸屬的不確定性

- 接受或拒絕 AI 建議之紀錄，可能成為法律審查重點
- 醫師擔憂相關紀錄在訴訟中之解讀方式

- 對 AI 稽核與專業審查的期待

- 醫師期望由具高度專業能力之專家進行 AI 決策稽核
- 專業審查為降低醫療疏失風險的重要配套



AI 臨床應用決策：降低風險 增強信心

Mansi and Riedl et al., 2025

- 批判性思考作為降低 AI 風險的核心途徑
 - 醫師普遍將人工智慧視為思考夥伴而非直接提供答案或取代臨床判斷的工具，依賴自身臨床決策與批判性思考是降低風險的最重要方式
 - 實際決策流程
 - 多數醫師的資訊瀏覽順序為：
 1. 先檢視病人摘要與臨床資訊
 2. 以醫學演算法或既有知識驗證自身想法
 3. 最後參考 AI 建議作為確認或補充
- 對 AI 風險的整體看法

只要持續運用臨床判斷使用 AI 被視為低風險行為

AI 提供的建議為輔助思考與檢核，非不可質疑決策依據

星球永續健康 線上直播

林庭瑀
博士



陳秀熙
教授



國立台灣大學



林家妤



許辰陽
醫師



陳虹彥



曾暉哲



劉秋燕



梅少文 主持人



侯信恩主持人 楊心怡製作人



不只是科技



嚴明芳
教授



陳立昇
教授



羅崧璋

台北醫學大學

