

星球永續健康線上直播

星球健康週新知 &

專題: AI醫療治理(2)

AI 安全閘門治理沙盒(AI Airlock)

2026-07-08

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、林家妤、陳虹彬、邱士紘、尤翊庭、王斌俞



資訊連結:

<https://www.realscience.top/7>

星球永續健康線上直播



<https://www.realscience.top/7>

Youtube影片連結:

https://youtube.com/channel/UCCHTox4rUysI30QW4e_xliA?si=IDlj9qln3bZWMtNG

漢聲廣播星球永續健康: <https://reurl.cc/WbGALy>

新聞稿連結: <https://www.realscience.top/7>

本週大綱

- 健康科學新知 (2026 / W27)
- AI安全閘門治理沙盒(AI Airlock)
- AI安全閘門治理沙盒實例

健康科學新知

2026 / W27

約旦折衝美國-伊朗商議停火細節:「和而未穩」

美伊兩國上週波斯灣爆發連日互襲後衝突升溫 在卡達等鄰近國家斡旋下緩解局勢

Straitstimes

美國國務卿
魯比歐

科威特外交部長 阿薩巴

卡達首相兼外交大臣
阿勒薩尼

美國-伊朗派遣特使赴卡達首都杜哈在約旦斡旋下進行間接談判美方關注海峽治理 伊朗聚焦資產解凍

BBC.com

美國特使 庫什納

美國特使威特科夫



卡達發言人表示
正積極斡旋
美伊停火細節

BBC.com



記者乘快艇訪視海峽區域

卡達首都 杜哈

gulfnews



海峽恢復部分航運活動

CNN.com



美國總統 川普 沙烏地阿拉伯
王儲 薩勒曼

川普2025訪沙烏地阿拉伯達成鉅額投資與軍售協議 並簽署戰略防禦及AI備忘錄

俄國-烏克蘭擴大長程攻擊:「以攻制攻」

烏克蘭持續對克里米亞發動無人機攻擊，俄方控制區接連傳出火災與基礎設施受損

nbcnews.com



烏克蘭以長程無人機攻擊俄國能源與補給設施，戰事延伸至克里米亞及俄本土設施

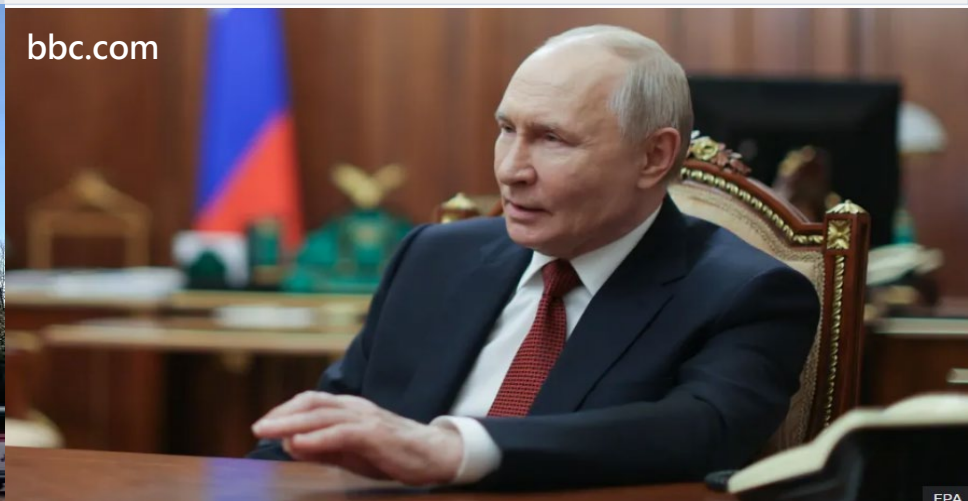
塞瓦斯托波爾變電站



烏軍攻擊克里米亞電力設施，停電與燃料短缺迫使當地進入緊急狀態

普丁拒絕烏克蘭提出限制長程攻擊提議，強調俄羅斯不會放棄長程火力優勢

bbc.com



Yahoo.com



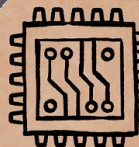
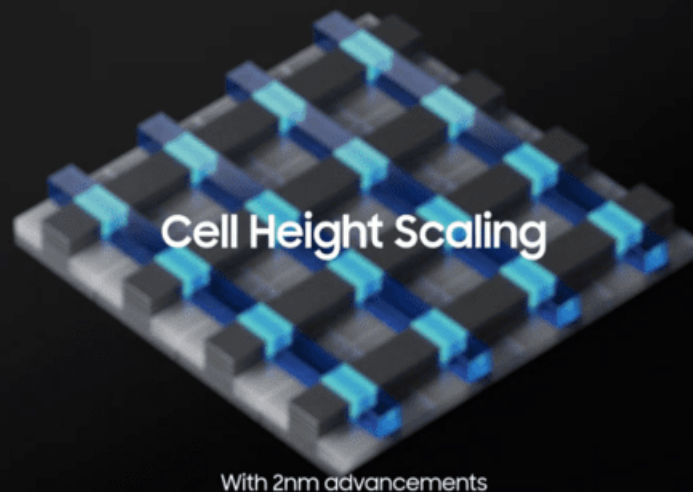
上週末俄羅斯發動大規模飛彈與無人機攻擊基輔，造成多人死傷 基輔濃煙遍佈

Anthropic-三星AI供應鏈成型:「算力自主」

三星和SK海力士上月公佈向人工智慧主要廠商之一Anthropic注資上億美元尋求合作

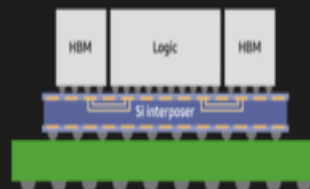


三星電子2奈米晶片設計



2.5D 큐브-S

2.5D Cube-S는 대형 인터포저에서도 탁월한 칩 제어 기능을 통해 높은 대역폭과 향상된 성능을 제공합니다. 로직 칩과 HBM 다이를 실리콘 인터포저에 수평으로 배치하여 강력한 컴퓨팅 성능을 위한 고대역폭, 저지연 데이터 경로를 구현합니다. 이를 통해 뛰어난 전력 안정성과 높은 메모리 밀도를 제공하는 동시에 고성능 애플리케이션의 열 효율을 크게 향상시킵니다.



- 더 많은 HBM 모듈을 탑재한 2.5D 패키지 개발을 위한 칩은웨이퍼(CoW) 기술을 지원합니다.
- 고급 로직과 최대 8개의 HBM 모듈을 통합한 3.3배 실리콘 인터포저를 탑재한 2.5D 패키지는 모든 인증을 완료하여 생산 가능합니다.
- 8개 이상의 HBM 모듈과 3,000nF/mm² 이상의 ISC™(통합 스택 커패시터)를 지원하는 확장된 실리콘 인터포저를 탑재한 더 큰 2.5D 패키지도 제공됩니다.
- 실리콘 인터포저 웨이퍼는 삼성 파운드리에서 제조합니다.

Anthropic 上週宣布與三星合作
設計生產研發專用加速晶片 建立AI供應鏈

thelec.net

亞洲AI晶片經濟榮景影響：「盛衰並行」

南韓投鉅資發展晶片與AI
力促區域均衡發展並確保技術領先



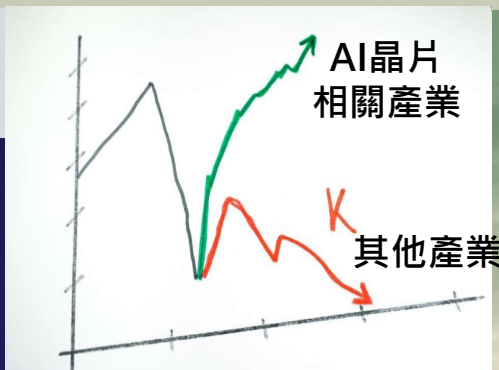
韓國總統李在明公佈
8800億美元晶片與AI投資計畫

BBC.com



因晶片產業利潤分配爭議
三星於5月發生大規模罷工

南韓AI晶片帶動強勁成長 卻加劇貧富差距與
房市投機 引發社會分化



business-standard.com



AI 晶片帶動台韓經濟 卻加劇整體貧富差距
產業發展形成K型分化 傳統產業陷入困境



aljazeera.com

AI帶動台灣科技產業與經濟快速成長卻加劇
K型分化 基層面臨通膨房價壓力

TriNetX大數據平台研究洪流：「慎思明辨」

 TriNetX 提供全球逾三億名患者電子病歷數據，五年內論文產量**激增近八十倍**

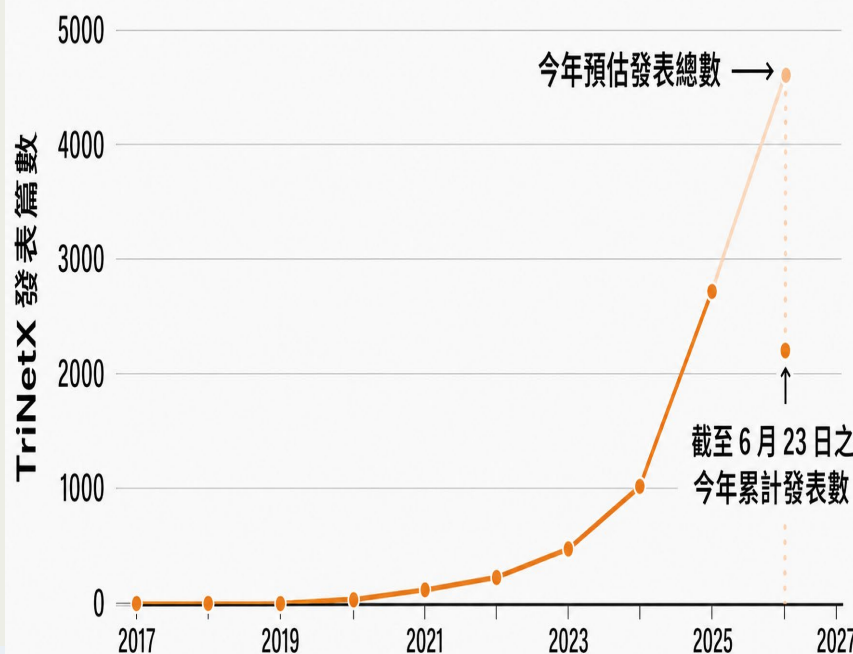
 學術界觀察到**大量數據偏差研究**，結果顯現過度樂觀治療推論結果

 醫療領域人員利用平台提供便利分析工具快速發表，成為工作申請加分籌碼



核心威脅與偏誤

1. 研究設計與統計漏洞：忽視科學因果與時序關係，致使研究產生誤導性結論
2. 數據操弄：易於操作分析系統誘發「**P 值修正**」行為，僅篩選顯著結果進行發表
3. AI 偽造：部分論文出現平台不具備之分析功能描述，疑受生成式模型錯誤引導



Frederik Joelsing, *Science*, 2026

案例分析與體制改革

- 抗癌神話破滅：GLP-1 藥物宣稱具備抗癌奇效研究，經驗證研究具**嚴重方法學缺失**
- 病患用藥風險：醫師若採納具偏誤之研究結論，將對患者用藥安全產生**負面衝擊**
- 選才標準調整：美國醫學會嘗試調整申請指標，引導學生由追求**數量轉向「質量」**

學術平台反制AI濫用維持學術品質:「信實為本」

- 少數研究者以生成式AI快速大製造文章投稿於開源平台 影響整體可信度
- 論文若含 AI虛假引用或未經作者查核痕跡，作者將被 arXiv 禁投一年
- 解禁後須先通過可信同儕審查，才能再次上傳 arXiv

Dalmeet Singh Chawla, Nature, 2026

AI slop 問題

① AI 文章氾濫

- 大量生成式AI製造低品質內容入侵學術平台影響整體科學品質

② 虛假引用問題

- AI 捏造參考文獻，顯示作者可能未確實查核

③ 學術信任危機

- 論文工廠與 AI 亂文，加劇研究品質與可信度問題

④ 禁令執行爭議

- 支持者認為可把關品質，反對者擔心只是治標

未來方向

- 多個預印本平台已開始採取禁令、關閉平台或發展偵測技術
- AI技術快速發展，未來辨識 AI 生成文章將更加困難
- 科學界需建立AI應用於研究執行與查核機制規範 確保研究品質與可信度



歐陸積極復興科研領導地位：「聚才立勢」

Elizabeth Gibney, *Nature*, 2026

背景

- 美國科學研究政策動盪，研究經費凍結、研究人員被裁撤，歐洲希望成為研究避風港
- 歐盟推動 Choose Europe，吸引科學人才，以美中競爭中建立第三股科學力量為目標

研究影響力

- 歐洲論文產量高，重要期刊表現接近美國與中國

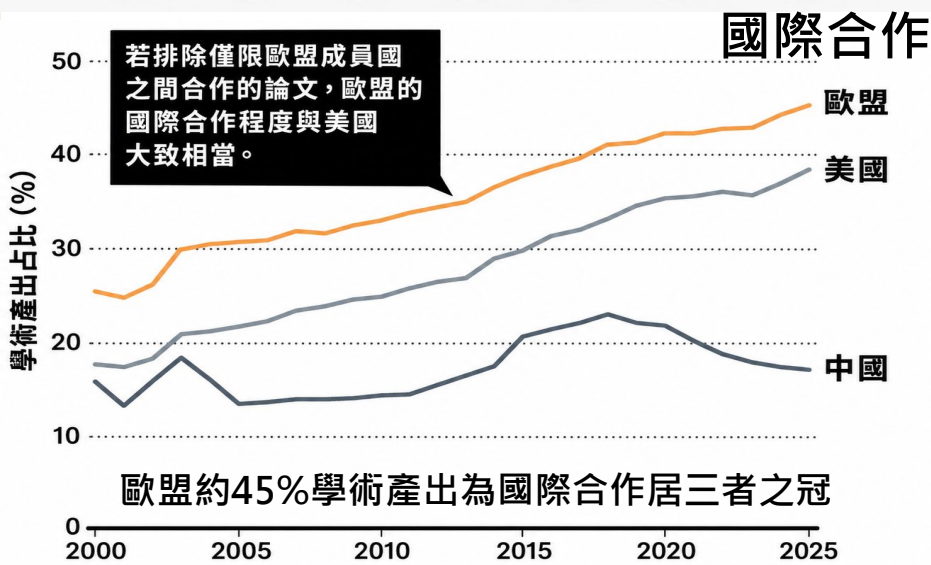
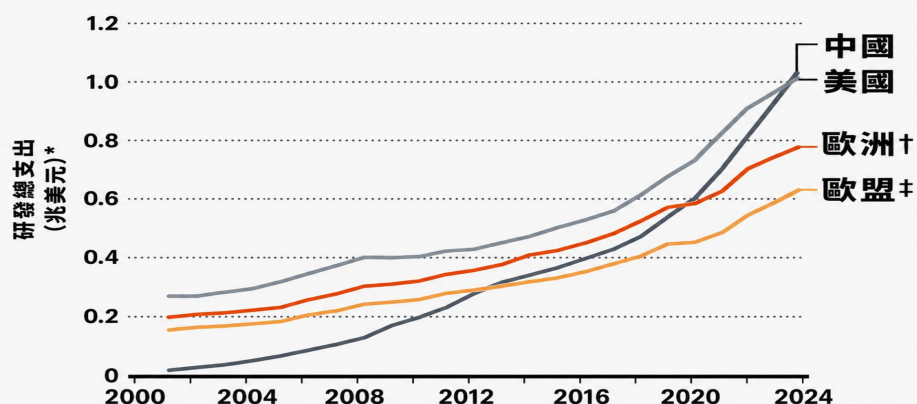
世界級機構

- CERN 等國際研究機構，具高度科學影響力
- 高比例跨國合作具多樣性與研究脈絡豐富性為歐洲研究體系重要競爭力

- 研發總投資仍低於美國與中國
- 研究成果較難轉化為產業與經濟效益
- 各國制度分散，人才流動與新創擴張受限
- Horizon Europe 預算可能大幅增加擴大與跨洲合作

研發支出

歐洲來自民間與公部門的研究與發展（R&D）總支出，如今已落後於美國與中國。



AI安全閘門治理 沙盒(AI Airlock)

NETFLIX

惡魔賭局爭議指控

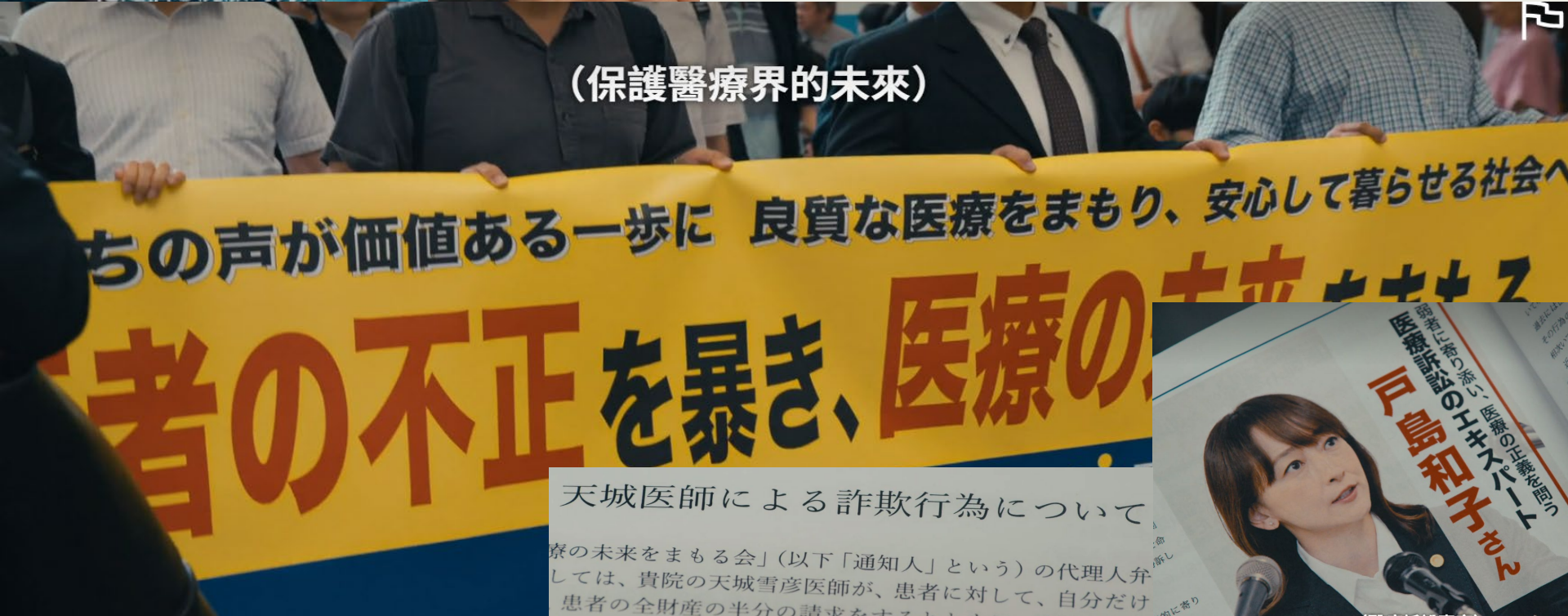


你必須先得到神的祝福
才能接受我的手術

並要求病患付出一半財產作為醫療費
他是個惡魔般的男人

- 心臟外科醫師天城雪彥醫術超群但行事風格受到醫療社群廣泛議論
- 天城醫師以籌碼賭局決定是否提供手術以及高額手術費用引起爭議
- 醫療律師戶島和子率醫療守護會提出控訴

(保護醫療界的未來)

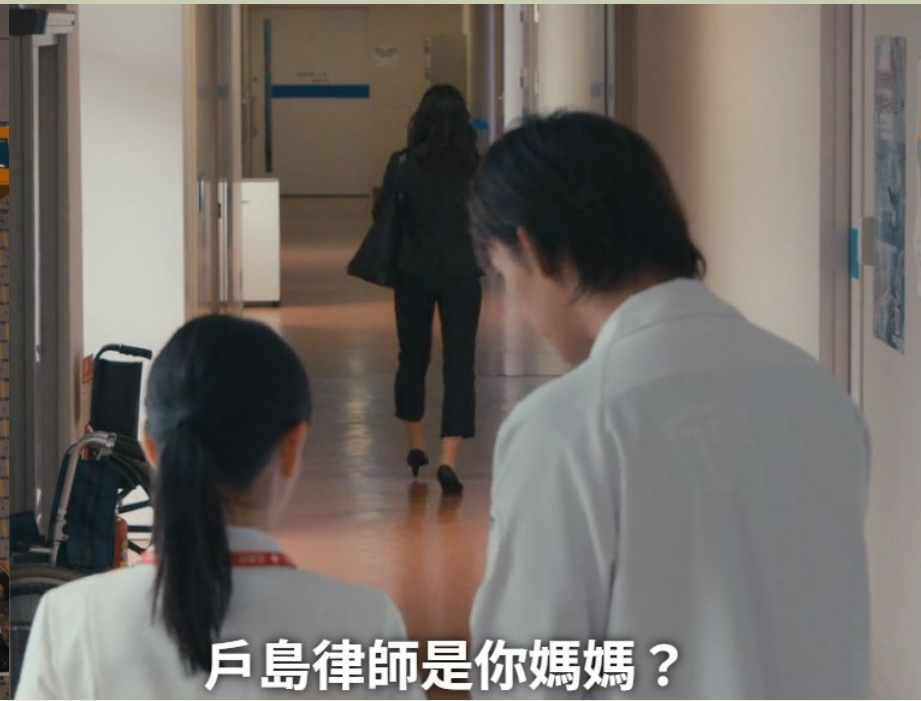


天城医師による詐欺行為について

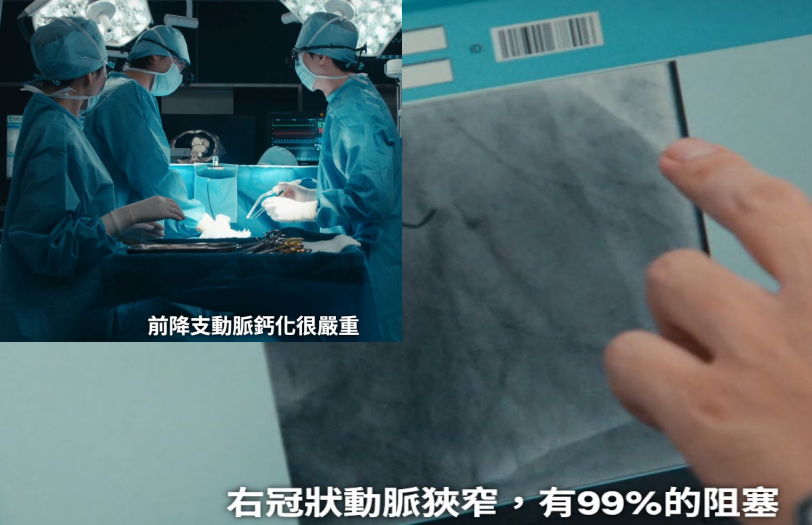
療の未来をまもる会」(以下「通知人」という)の代理人弁
しては、貴院の天城雪彦医師が、患者に対して、自分だけ
患者の全財産の半分の請求をするという行為は、明らかに



母女相認驟病告急



戶島律師是你媽媽？



前降支動脈鈣化很嚴重

右冠狀動脈狹窄，有99%的阻塞

- 東城大學醫院在說明會為平息爭議被迫取消天城醫師手術，訴訟風暴影響醫院決策
- 戶島律師在醫院中巧遇女兒美和護理師，且由於冠狀動脈狹窄暈厥 由於冠狀動脈嚴重鈣化使繞道手術與其他治療難以進行

智能手術新藥試驗



我願意親身當作這項新科技的試驗品

餵入病歷之後

這是維新大學研發的

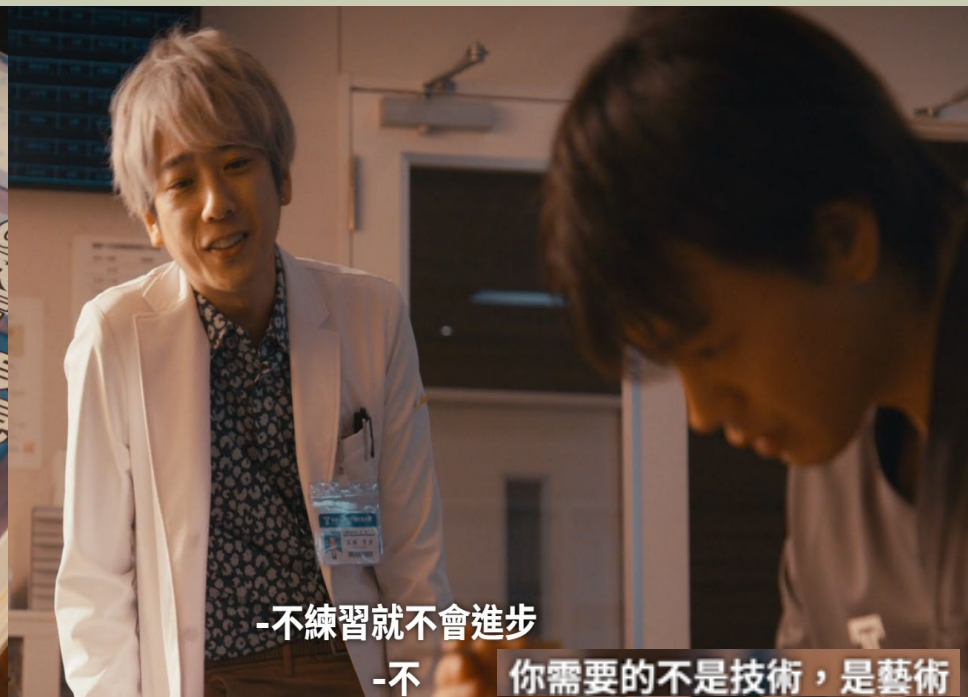
而且能透過進階影像分析
引導我們進行手術

- 拒絕接受天城醫師治療後戶島律師轉至維新大學醫院
- 戶島律師自願成為第一位受試者採用野田醫師所開發艾卡諾AI手術輔助系統，利用新藥促進血管新生再進行繞道手術 AI預測成功率可達89%

就能對新生血管進行繞道手術了

麻州醫學大學進行的研究顯示

AI 技術 vs 手術藝術

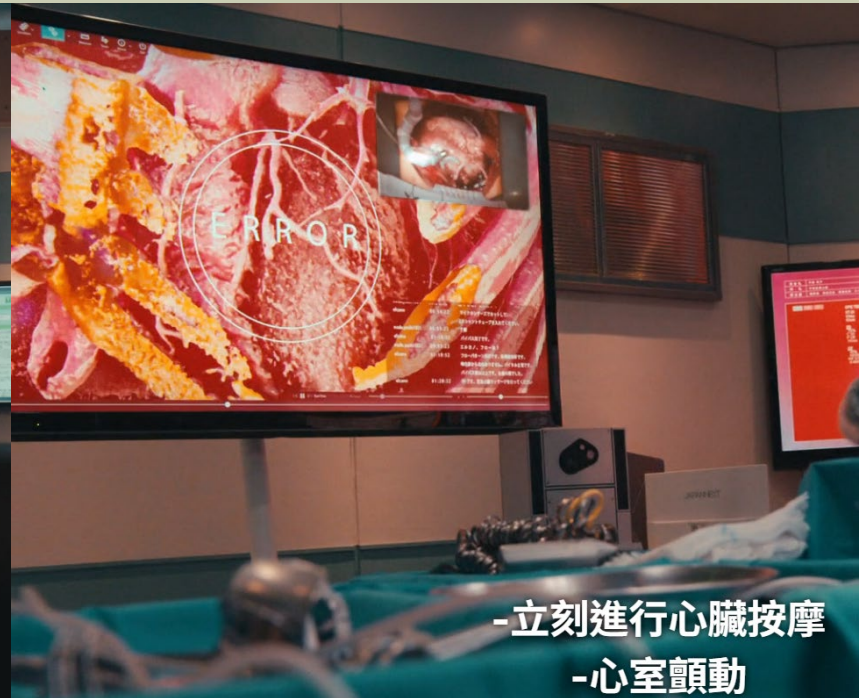


- 維新大教授菅井教授與野田醫師提出 AI 技術能讓任何醫師完成困難手術 挑戰天才外科醫師存在的必要性，AI 工具可信度讓病患家屬備感憂慮
- 世良醫師則練習不懈希望以努力克服失敗 天城醫師提醒創新性是療法關鍵

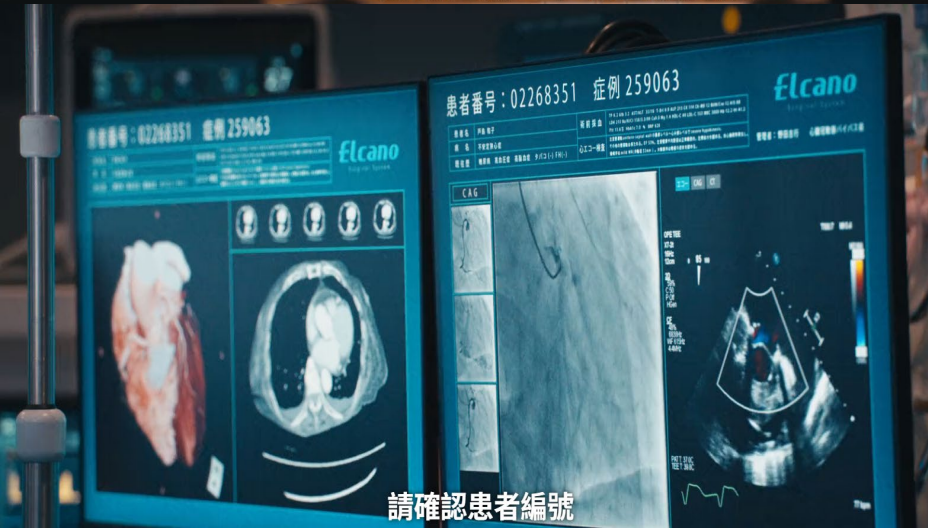
公開手術心室顫動



沒有經驗，手術也能完成



-立刻進行心臟按摩
-心室顫動



請確認患者編號

- 公開手術野田醫師依照艾爾卡諾 AI 指示執刀，手術過程順利顯示沒有經驗也能完成手術技術宣稱
- 手術剛結束由於新藥副作用戶島發生心室顫動，標準化AI手術無法即時反應病患命懸一線

師徒接續手術 人機互動



- 關鍵時刻世良醫師想起天城特訓時反覆演練特殊繞道手術接手搶救戶島
- 術中AI建議8-0號線，世良依據眼前狀況改用7-0 號線，顯示醫師必須以臨場判斷駕馭 AI，醫療不只是標準化流程，而是結合想像力、技藝與對病人個別狀況負責的醫師判斷

- 用7-0號線
- 它說用8-0號線

人工智慧治理七大原則



兼顧 社會公益、數位平權、
創新研發 與 國家競爭力



AI Airlock 安全閘門 助力臨床導入



- 1 英國首個 AIaMD 監管沙盒**
由 MHRA 與 NHS、DHSC
協作推動。



- 2 試點期間**
2024 年 5 月 – 2025 年 3 月。



- 3 核心目的**
在可控環境探索醫療 AI
監管挑戰。



- 4 應用連結**
把產品測試、法規缺口分析
與臨床情境串接。



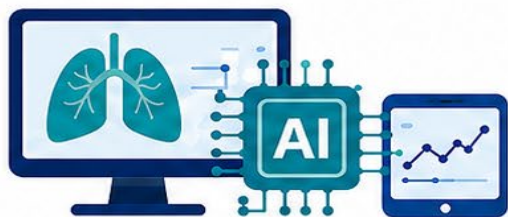
- 5 最終價值**
支持安全部署、政策修正
與公眾信任。



安全閘門智慧醫療應用設計

1 真實產品導向：

以實際 AIaMD 產品
或原型探索監管問題。



2 多方協作：

結合監管者、臨床、
技術、法規與學術觀點。



3 情境化測試：

依產品挑戰設計
測試計畫與
證據生成。



創新測試

在真實世界條件下
測試與驗證



可控測試機制



臨床部署
安全、有效
服務病人



法規學習
產生證據
促進制度改進

4 法規學習回饋：

把測試結果轉化為
指引、政策與制度
改進。



核心目的：在促進創新、維護病人安全與建立公眾信任之間取得平衡。

AI Airlock 治理沙盒智慧技術導入

從實際醫療 AI 應用看沙盒如何連結法規問題



1 Philips Healthcare – AutoImpression



應用領域 放射報告摘要／生成



對應法規議題 合成文字資料之生成、評估與驗證



2 AutoMedica – SmartGuideline



應用領域 臨床指引問答／
工作流程支援



對應法規議題 以 RAG 降低 LLM
幻覺與非決定性



3 OncoFlow – OncoFlow



應用領域 腫瘤學資料分析



對應法規議題 可解釋性與臨床效能
之權衡



4 Newton's Tree – FAMOS



應用領域 放射 AI 安全監測



對應法規議題 模型漂移、資料品質與
自動化偏誤之即時監測



四案共同顯示：醫療 AI 監管問題往往跨越**資料、模型、臨床與監測**多個層面。

AI Airlock 智慧技術導入

Philips AutoImpression



放射科報告產製



生成式 AI / LLM



自動草擬「初步診斷 / 結論」段落



重點：提升效率、減少遺漏、支援放射科醫師

OncoFlow



腫瘤學 MDT 決策支援



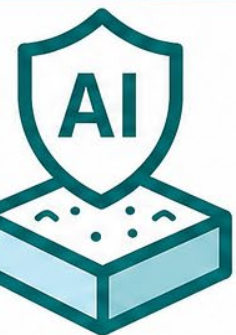
LLM + RAG



非結構化資料擷取 + 治療匹配



重點：提升一致性、透明度與資源運用



AI Airlock 沙盒角色

1



驗證
早期技術



2



釐清
規範議題



3



補足
證據缺口



兩項技術皆屬
早期階段，
主要證據來自

MHRA AI Airlock Sandbox
Pilot Programme Report。

三大測試環境

把監管問題放進不同成熟度與風險情境中驗證

1

模擬 Airlock (Simulation)



- 工作坊／圓桌討論
- 探索可能失效與監管疑義
- 適合概念或早期產品

2

研究／虛擬 Airlock (Research / Virtual)



- 在虛擬環境使用真實或合成資料
- 觀察模型行為、準確性與穩定性
- 類似 test bed 機制

3

真實世界 Airlock (Real World)



- 於醫療場域部署但不影響病人照護
- 觀察工作流程整合與安全性
- 在適當監督下蒐集實務證據



三種場域可單獨或組合運用，
以回應不同生命週期之監管挑戰。

OncoFlow案例如何對應AI基本法

台灣AI基本法	OncoFlow Airlock 對應措施
人本	MDT仍由醫師決策
安全	AI先Airlock再臨床
公平	評估不同資料來源
可解釋	SHAP、LIME、RAG
透明	顯示治療依據
問責	Human Override
持續改善	Real-world feedback

AI安全閘門治理 沙盒(AI Airlock)實例

影像檢查需求增加與影像判讀壓力

📍 北區醫學中心 乳房影像篩檢中心



篩檢量大增

兩年成長近 5 成



醫師人力不足

判讀負荷持續上升



報告產出延後

影響整體作業效率



複檢排程受影響

報告延遲影響通知與安排



**導入 AutoImpression
輔助初步診斷草擬**

協助草擬初步診斷
提升效率與一致性



擷取影像重點



草擬初步診斷



醫師審閱編修



加速報告產出

AI 草擬初步診斷與醫師複核

📍 55歲女性 | 例行雙側乳房攝影篩檢

影像判讀室

AutoImpression
AI 輔助診斷系統



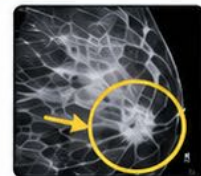
AI 建議初步診斷

AI 草擬初步診斷 (摘要)

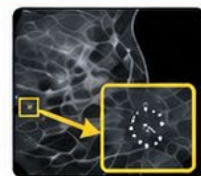
影像重點 (右乳)	右乳見不規則腫塊，合併局部群聚性微小鈣化
BI-RADS 分類	4：高度懷疑惡性
建議處置	建議乳房超音波與切片檢查
關注部位	右乳

備註：本建議未包含病史資訊 (如手術史等)

1 右乳見
邊緣不規則腫塊



2 合併局部群聚性
微小鈣化



3 AI 產生初步診斷：
建議 BI-RADS 4



4 建議乳房超音波
與切片檢查



乳房超音波



切片檢查

5 醫師補充左乳良性手術病史後簽署

☒ 左乳於 2 年前良性腫瘤切除手術

醫師簽署：王○○ 醫師



報告已簽署



AI 用於效率輔助，最終判斷仍由醫師負責

AutoImpression 技術架構與運作流程

雙軌流程：合成資料生成 + 品質評估



Step 1 合成放射報告生成

A. 來源資料



MIMIC-IV
去識別資料庫

2008-2019

逾 28 萬名病人

含量測、診斷、
處置與文字紀錄

B. 生成結構化報告



- ✓ 保留原始結構
- ✓ 導入隨機化臨床內容
- ✓ 移除初步診斷段落

Step 2 合成初步診斷生成



輸入：
結構化報告



輸出：
合成初步診斷

Prompt 設計要點

- ✓ 低溫度
- ✓ 排除正常發現
- ✓ 突顯關鍵異常
- ✓ 納入後續處置建議
- ✓ 批次處理

Step 3 品質評估

A. 自動化評估 (AI 評估)



- 一致性
- 臨床合理性
- 資訊完整性
- 關鍵異常覆蓋

B. 人工審查 (放射科醫師)



- 臨床可讀性
- 臨床準確性
- 遺漏偵測
- 處置合理性

C. 量化指標彙整與分析回饋



AutoImpression 效能結果與規範議題

測試發現與監理重點

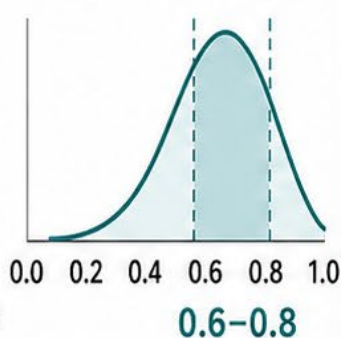


1 效能結果

ICD-10 分布比較



BioBERT 相似度分布



ICD-10 分布與真實資料有差異



相似度多落在 0.6-0.8



部分醫師難分辨合成與真實報告



較短句子通常相似度較高

2 LLM 裁判與 放射科醫師一致性



準確性

105/140



清晰度

117/140



有據性

109/140



完整性

91/140

3 關鍵規範議題



循環性風險：
AI 驗證 AI



概念混淆：
有據性與完整性不同



評分粒度偏低，
且易高分化



法規缺口：
MDR 2002 未明確
涵蓋文字型合成資料

OncoFlow 腫瘤治療智慧決策輔助



1 資料分散

病理、影像、
病歷與指南分散



2 整理耗時

會前準備耗用
大量人力



3 判讀不一致

不同文件格式與
用語造成落差



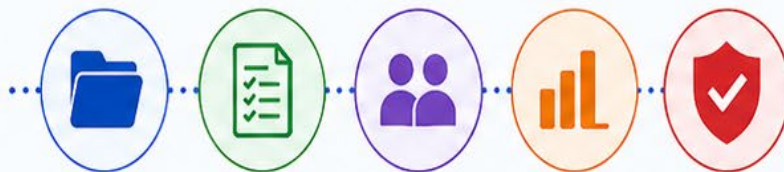
4 討論效率受限

複雜病例難
快速聚焦



5 影響病人照護

可能延誤重要
治療決策



**OncoFlow 將臨床資訊
結構化，協助 MDT
更快聚焦關鍵決策。**

OncoFlow 效能與可解釋性結果



從資料擷取到治療匹配

1 資料擷取結果



傳統規則式 NLP 擅長日期與數字，但難處理複雜臨床推理



LLM 較能處理複雜關係與不同文件風格



最佳配置於相關臨床資料擷取達約 **90%** 準確度



嵌入預先定義臨床規則後，可超越 **80%** 準確度基準

2 治療匹配結果



獨立 LLM (無 RAG)
幻覺風險較高



最佳檢索：
PubMedBERT 嵌入 +
Qdrant 向量資料庫



ESMO / NICE 指南
檢索準確度 **92%**



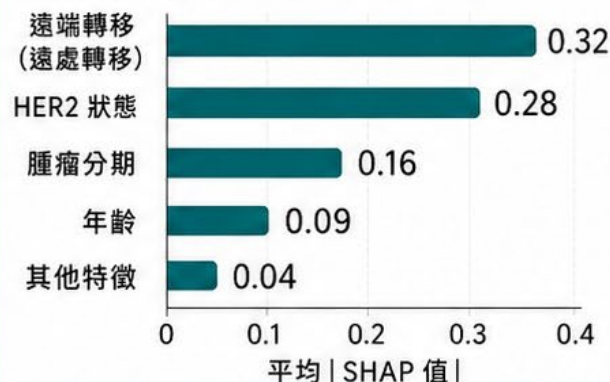
SACT protocol
匹配準確度 **100%**



低幻覺率 **3.4%**

3 可解釋性重點

特徵重要性 (SHAP 值)



SHAP / SHAP-IQ 顯示
「遠端轉移」與「HER2 狀態」
影響最大



仍可能在指南資料缺口下
出現過度概化



最終決策前仍須
臨床醫師驗證



結論： OncoFlow 結合 LLM、RAG 與臨床規則，於資料擷取與治療匹配展現高準確度與低幻覺率，並以可解釋性支撐臨床決策，助力精準腫瘤治療。

兩項智慧醫療輔助工具沙盒導入評估



Philips AutoImpression vs. OncoFlow



兩者皆為早期階段

	Philips AutoImpression	OncoFlow
 技術類型	生成式 AI / LLM	LLM + RAG
 臨床領域	放射科報告產製	腫瘤 MDT 決策支援
 沙盒挑戰	合成資料生成與驗證	可解釋性與效能權衡
 測試環境	虛擬沙盒	虛擬沙盒 + 專家調研
 主要資料	MIMIC-IV (合成 4,000 份報告)	Barts Health (9,000 病理 + 3,000 放射)
 代表性效能	相似度 0.6-0.8 醫師難分合成 / 真實	擷取 90%、指南檢索 92%、 SACT 匹配 100%
 同儕審查發表	皆無	皆無



監理沙盒可
驗證早期技術



可補足
學術證據缺口



後續仍需最佳實踐
與獨立驗證



林庭瑀
博士



陳秀熙
教授



星球永續健康 線上直播



國立台灣大學



林家妤



陳虹彬



許辰陽
醫師



梅少文 主持人



侯信恩主持人



楊心怡製作人



尤翊庭



王斌俞



邱士紘



劉秋燕



嚴明芳
教授



陳立昇
教授



台北醫學大學