衛生福利部資訊處

三大類型智慧醫療中心

技術手冊

113 年 7 月

執行單位 衛生福利部次世代數位醫療平臺專案辦公室

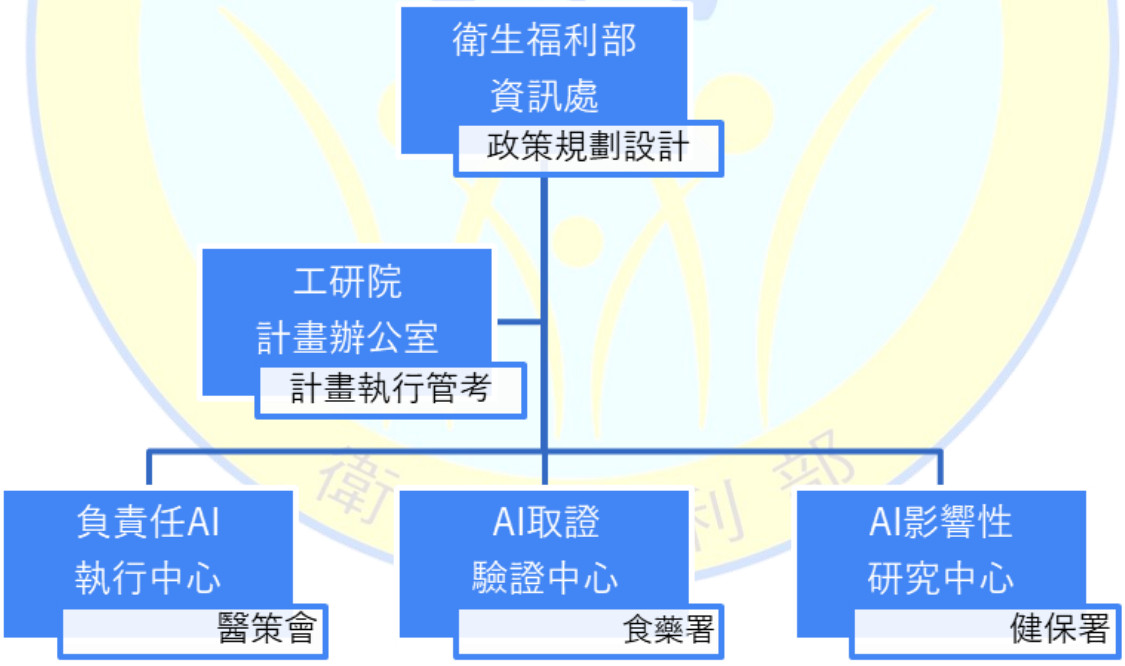
目錄

衛生福利部資訊處三大類型智慧醫療中心技術手冊

前言

本技術手冊將詳細介紹衛生福利部資訊處三大類型智慧醫療推動中心的設置與運作，這些中心將全力推動醫學人工智慧（AI）與資訊技術的發展，目的是提高臨床應用的透明度與效率，增進公眾對於 AI 技術在醫療領域應用的信任與接受度，加速智慧醫療產品在台灣商品化的進程，與建立實證智慧醫療應用，與醫療科技評估的基礎。

圖一：執行架構圖



第一部分：負責任 AI 執行中心

負責任 AI 執行中心的設立主要解決醫師與民眾對於人工智慧技術的疑慮，特別是針對 AI 技術的不透明性與偏差兩大問題。世界衛生組織（WHO）針對在醫療領域中負責任使用人工智慧（AI）提出了六大原則，旨在確保 AI 技術的應用符合倫理、公平，並促進公共衛生。以下是每個原則的簡要說明以及在醫院實行的方法：

一、保護自主權（Protect Autonomy）：AI 應該輔助而不是取代人類的決策，並尊重病患的權利和隱私，醫院應確保 AI 系統提供醫療專業人員決策支援工具，而不是自行做出決策。

二、促進人類福祉和安全及公共利益（Promote Human Well-being and Safety and the Public Interest):AI 系統應優先考慮病患安全，並致力於改善大眾健康，定期評估 AI 系統的安全性和有效性。收集醫療專業人員和病患的回饋，改進 AI 工具。使用 AI 辨識公共衛生趨勢，並支援預防醫學計劃。

三、確保透明度、可解釋性和可理解性（Ensure Transparency, Explainability, and Understandability）：AI 系統應該透明，並提供清晰的解釋其決策過程。使用可解釋的 AI 演算法，確保醫療提供者了解這些系統的運作方式。向病患提供清楚的說明，告知 AI 在其醫療中的應用以及 AI 生成建議的依據。

四、促進責任感和當責制（Foster Responsibility and Accountability）：應明確界定 AI 系統在醫療中的當責機制。醫院可指派專門的團隊或個人負責監控 AI 系統。建立處理 AI 系統錯誤或故障的標準流程，確保當責機制落實。

五、確保包容性和公平性（Ensure Inclusiveness and Equity）：AI 的設計和使用應促進包容性，並解決健康差距問題。研發臨床 AI 產品應使用多元化的資料庫訓練 AI 系統，避免偏見。確保 AI 工具對所有病患可及，不論其社會經濟地位，並監控 AI 驅動的醫療結果中的任何差異。

六、促進可持續的 AI（Promote Sustainable AI）：AI 的開發和部署應環保且資源高效。醫院應選擇節能且環境影響最小的 AI 系統。定期評估 AI 系統的資源使用情況，並尋求減少其生態足跡的方法。

要具體的落實這些原則，美國衛生福利部國家協調辦公室（ONC）最近發布的 HTI-1 規則（健康資料、技術與互操作性）在醫療領域引入人工智能和預測導向的臨床決策平台納入§170.315(b)(11)決策支持介入（DSI，Decision Support Intervention）標準。為了防止預測導向算法中的潛在偏見，ONC 採用了一個稱為 FAVES 的框架，目標是要達成人工智慧醫學決策支持系統能公平、適當、確實、有效、和安全（FAVES, Fair, Appropriate, Valid, Effective, Safe）的做出醫療決策的建議。

公平（無偏見，公正）是指模型在個體或群體的固有或後天特徵上不表現出偏見、偏袒或歧視。使用該模型的影響在相同或不同人群或群體中是相似的。
適當：模型與其應用的具體背景和目標人群相配合。
確實：模型已被證明能夠準確估計目標值，並在內部和外部資料中符合預期。
有效：模型在現實條件下表現出顯著的益處和顯著的結果。
安全：模型的使用可能帶來的益處超過任何可能的風險。
申請本中心的醫院，需要組織管理團隊，建立辦法，落實負責任的 AI 使用，具體負責任 AI 應用的辦法內容必須包括：

1. 符合資安與隱私保護原則的管理辦法
2. 符合透明性原則
3. 遵循 AI 生命週期循環監測有效性原則

符合資安與隱私保護原則的管理辦法

進行適當之隱私保護

1. 建議可以依循美國的 HIPAA（健康保險便攜性和責任法案）或歐盟的 GDPR（通用資料保護條例)移除可辨識個資資料或進行適當資料加密[1,2]。
2. 規範資料存儲地點、期限

3. 規範外部管理人員可接觸資料權限
4. 建立定期查核機制
5. 建立個資外洩應變計畫
6. 對於資料必須上傳院外雲端供 AI 推論運算執行之應用程式必須提高保護措施

資訊安全

1. 對於必須上傳院外雲端或是第三方管理的 AI 應用程式應符合下列資訊安全措施
2. 使用防火牆、入侵檢測/防護系統（IDS/IPS）等網絡安全措施，保護醫院網絡[3,4]。
3. 分割網絡，限制對關鍵系統和資料進行存取。
4. 使用自動化工具掃描 ，檢查安全問題。
5. 規範外部管理人員權限
6. 建立定期查核機制

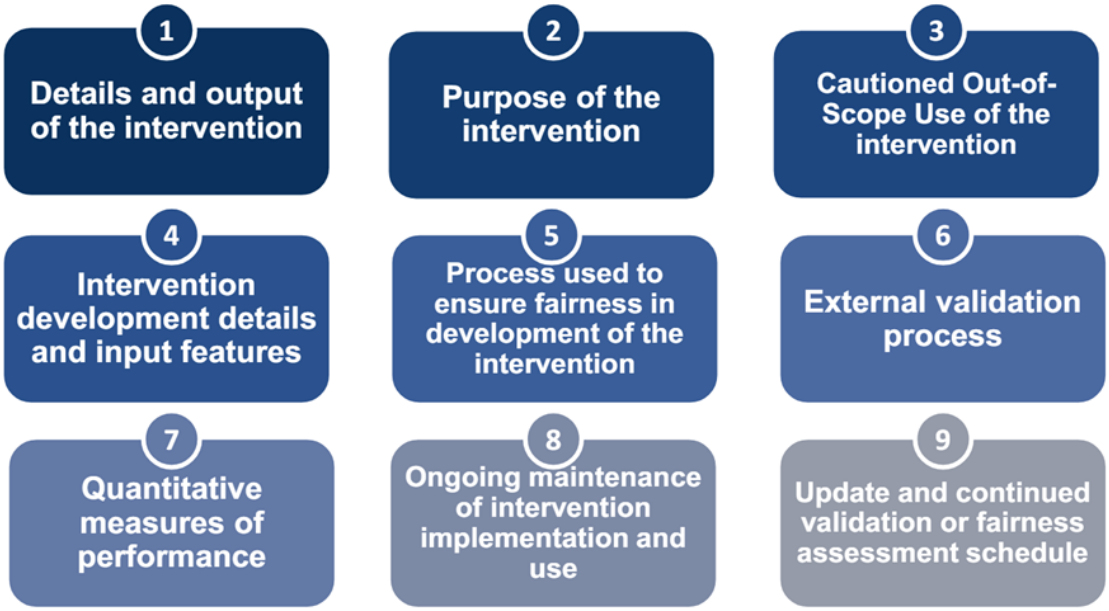
歐盟 GDPR 和美國 HIPAA 規範

GDPR，全名是《通用資料保護條例》，於 2018 年 5 月 25 日生效，是歐盟針對資料保護與隱私制定的法律。它適用於所有處理歐盟居民個人資料的公司，不論公司位於歐盟內外。GDPR 旨在增強個人對其資料的控制，並簡化國際企業的監管環境。HIPAA，即《健康保險可攜性和責任法案》，於 1996 年通過，是美國針對健康資料保護的法律。HIPAA 旨在確保個人的健康資料保密，並制定了標準來保護病人的隱私權，適用於醫療服務提供者、和醫療保險機構，以及他們的商業夥伴。要使資料庫符合 GDPR 或 HIPAA 的要求，需要實施一系列技術和管理措施。首先，必須確保資料在傳輸和存儲過程中被加密。其次，設置嚴格的存取控制機制，僅授權人員才能存取敏感資料，防止未經授權的登錄。資料庫需具備資料刪除功能，確保能徹底移除不再需要的資料。此外，定期進行資料備份，並確保備份資料同樣受到加密保護。在 GDPR 和 HIPAA 的要求下，有多種標識符需要移除，以確保個人資料

的隱私和安全。在 GDPR 中，應移除的標識符包括姓名、地址、電子郵件地址、電話號碼、身份證號碼、駕照號碼、IP 地址、Cookie 識別碼、設備識別碼等任何能直接或間接識別個人的資料。在 HIPAA 中，應移除的標識符主要針對健康資料，包括病人姓名、地理位置、所有日期（如出生日期、入院和出院日期）、電話號碼、傳真號碼、電子郵件地址、社會安全號碼、醫療記錄號碼、健康保險受益者號碼、帳戶號碼、車輛識別號碼、裝置識別號碼、網址和 IP 地址、生物識別標識符等。這些標識符在資料處理過程中必須被移除或匿名化，以保護個人和病人的隱私。

透明度強化：

確保所有使用 AI 工具的醫院必須在一個專門的網站上公開重要資訊，包括 AI 的資料來源、訓練模型、驗證資料、適用的臨床情境、以及 FDA 或其他醫學審核機構的驗證結果等[5,6]。



https://www.healthit.gov/sites/default/files/page/2024-01/DSI_HTI1%20Final%20Rule%20Presentation_508.pdf

揭露項目	示範案例
Details and output of the intervention **介入詳情及輸出**	這套人工智慧系統會分析乳房 X 光片，偵測並標記可能的可疑病灶。輸出的結果如下 a）電腦輔助標記檢測到可疑鈣化或軟組織病變的位置。 b）決策支持通過區域評分提供，範圍從 0 到 100，分數越高表示懷疑程度越高。 c）連接不同乳房視圖中相應區域的鏈接，可用於增強用戶界面和工作流程。 d）檢查評分將檢查分類為 1 到 10 分，分數越高表示患癌可能性越大。該評分經過校準，使得在無癌症的乳房 X 光照片人群中，每個分類大約占 10%。
Purpose of the intervention **介入目的**	這套人工智慧系統旨在輔助放射科醫師辨識乳房 X 光片上的潛在乳癌病灶，幫助提高早期發現的機率，減少診斷遺漏的風險，並提升工作流程。
Cautioned Out-of-Scope Use of the intervention **介入的警告範圍外使用**	這套人工智慧系統不應獨立用作診斷工具。它不是要取代放射科醫師，而是輔助他們進行診斷。請勿將其應用於乳房 X 光片以外的影像，例如 MRI 或超音波圖像，另外如果使用的目標人群和 AI 訓練人群有顯著差異性。也應進行機構正確性評估。
Intervention development details	這套人工智慧系統是基於大量來自不同人群的匿名乳房 X 光片資料進行開發的。乳房 X 光片包含

and input features **介入開發詳情及輸入特徵**	2D 影像(FFDM 全數位乳房攝影)和 3D 影像(DBT 數位乳房斷層攝影)切片，輸入特徵包括可疑的鈣化點和軟組織病變（包括密度、腫塊、結構變形和不對稱）、病患年齡和乳房密度。開發過程中使用了先進的機器學習算法，如卷積神經網絡（CNN）。這些算法使用大量經病理切片證實的乳癌樣本、良性異常樣本和正常組織樣本進行訓練。
Process used to ensure fairness in development of the intervention **確保介入開發公平性的過程**	在開發過程中，對資料集進行了分析，以確保其代表不同年齡、種族和乳房密度的人群。我們使用了檢測和緩解偏差的方法，確保訓練資料的平衡，並進行了公平性審查。
External validation process **外部驗證過程**	這套人工智慧系統在不同醫療機構使用的外部資料集上進行了驗證，這些資料集不包含在初始訓練資料中。這樣的驗證過程確認了系統在各種臨床環境和病患群體中的準確性和適用性。 測試資料包含來自不同製造商的 2D 和 3D 乳房 X 光片（2D：Hologic、GE、Philips、Siemens and Fujifilm，3D：Hologic、Siemens and Fujifilm），代表了常規乳癌篩檢和無症狀患者，這些資料來自七個歐盟國家和美國的多個臨床中心。 測試資料包含 7882 例無癌症檢查和 1240 例有癌症檢查。在無癌症檢查中，4797 例是 2D 影像，3085 例是 3D 影像。在有癌症檢查中，819 例是

	2D 影像，421 例是 3D 影像。總體而言，測試資料中有癌症的檢查中，61%的病灶被歸類為腫塊，33%為可疑鈣化點，6%為結構變形或不對稱。三種主要的組織學癌症類型分別是浸潤性導管癌（60.5%）、原位導管癌（25.9%）和浸潤性小葉癌（9.0%）。病灶範圍中位數（定義為兩維度的最大直徑）在 2D 資料中為 16mm（四分位距：11-24），在 3D 資料中也是 16mm（四分位距：11-25）。
Quantitative measures of performance **模型表現的量化指標**	通過計算至少在一個視角（MLO 或 CC）中正確定位的癌症比例來計算基於檢查的性能指標如下： 準確率：95% 靈敏度：94.7% 特異性：90% 精確度：91% AUC: 0.949 召回率：92% 。
Ongoing maintenance of intervention implementation and use **介入實施和使用的持續維護**	我們有專門的團隊負責在系統部署後的監控，處理任何問題。系統會定期更新，我們也提供使用者支持和故障排除服務。
Update and continued validation or fairness assessment schedule	此人工智慧系統每年會隨機選 300 例以上真實世界影像進行機構獨立性能指標評估，確保系統有穩定的正確性。更新和驗證結果會定期提供給臨

更新和持續驗證或公平性評估計劃	床醫師，當靈敏度低於預設 85%，服務會暫停，系統會補充新影像資料訓練，性能提升至 90%後，重新恢復使用。

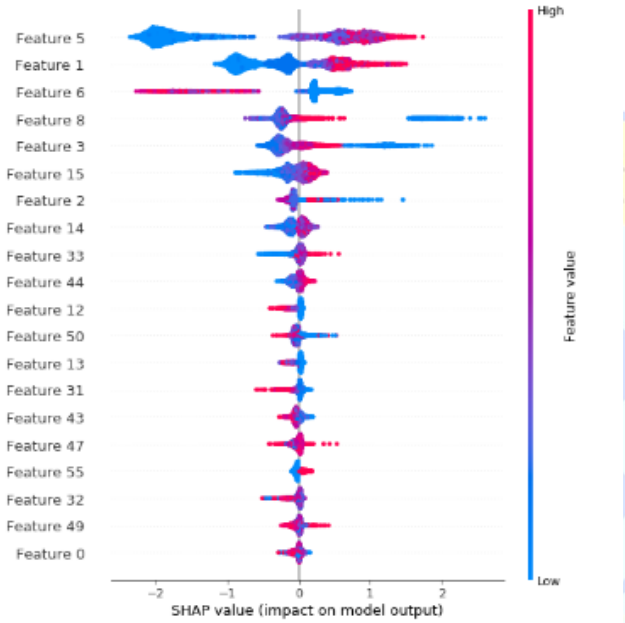
可解釋性分析提升臨床 AI 使用透明性

可解釋性分析在醫療人工智慧中是指用來解釋和理解人工智慧模型如何做出預測或決策的技術和方法。這在醫療領域中至關重要，因為透明性和信任對於人工智慧工具的採用是必不可少的。其目標是提供對人工智慧系統決策過程的洞見，確保臨床醫師能夠理解和驗證其輸出結果。常見的解釋性分析方法，以 SHAP 值和顯著性圖為例。

SHAP 值（SHapley Additive exPlanations）

SHAP 值 是一種強大的方法，源自博弈論，用於解釋機器學習模型所做出的單一預測。它們提供了一種將每個特徵對最終預測的貢獻歸因的方法。SHAP 值基於博弈論中的 Shapley 值，該方法公平地分配報酬給參與者。在機器學習的背景下，特徵被視為“參與者”，對“報酬”（預測）作出貢獻。SHAP 值測量每個特徵對實際預測和整體資料集平均預測之間差異的貢獻程度。對於每個預測，SHAP 值顯示每個特徵（例如年齡、腫瘤大小等）如何影響模型的輸出。正的 SHAP 值表示特徵對增加預測分數（例如，惡性可能性更高）的貢獻，而負的 SHAP 值則表示減少預測分數（圖二)[7]。

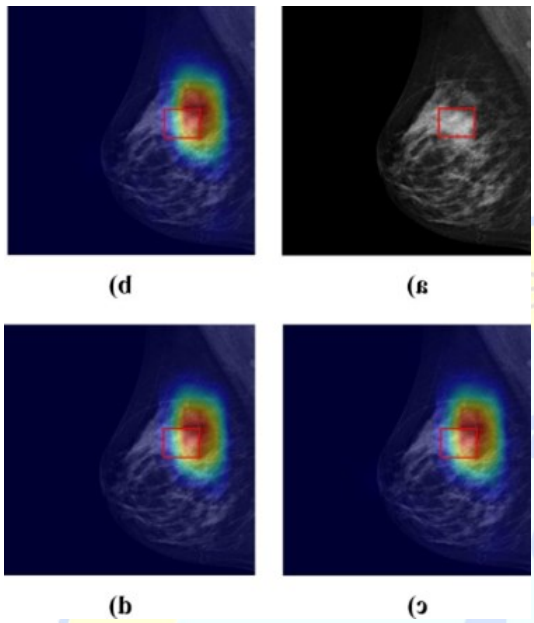
圖二 ： SHAP 值圖



顯著性圖（Saliency Maps）

顯著性圖 是一種視覺解釋方法，突出顯示輸入資料中對人工智慧模型預測貢獻最大的部分。它們通常用於基於圖像的人工智慧模型中。顯著性圖通過計算模型輸出對輸入圖像的梯度來生成。這些梯度顯示了圖像中哪些區域對模型預測有最顯著的影響。在顯著性圖中，較亮的區域表示對模型輸出變化的高重要性或敏感性。這有助於直觀地識別輸入資料（例如乳房 X 光片）中影響模型決策的區域。對於檢測乳房 X 光片中可疑病灶的人工智慧系統，可以為特定的乳房 X 光片生成顯著性圖。該圖將突出顯示模型認為對預測最重要的圖像區域。例如，顯著性圖可能顯示模型集中於特定病灶區域以做出高風險預測（圖三)[8]，幫助放射科醫師驗證人工智慧的重點並更好地理解其推理過程。

圖三: 顯著性圖（Saliency Maps）



Eur J Radiol. 2024;173:111356.

除了 SHAP 值和顯著性圖之外，還有一些其他的解釋性分析方法在醫療人工智慧中也很常用。LIME 提供了局部、可解釋的預測近似。PDPs 和 ALE 圖 提供了對單一特徵影響的深入了解，特徵重要性分數 根據其整體影響對特徵進行排名，反事實解釋 突出顯示了改變預測所需的最小變化，每種方法提供了不同的模型行為視角，使臨床醫師能夠更好地理解、信任和有效利用人工智慧於醫療決策中，以下是這些關鍵技術的描述以及如何解釋其結果：

一、LIME（Local Interpretable Model-agnostic Explanations 局部可解釋模型不可知解釋）

LIME 是一種用來解釋單一預測的方法，它通過在給定預測的局部範圍內，用更簡單、可解釋的模型來近似複雜的人工智慧模型[9]。LIME 的運作方式如下，一、擾動：對於特定預測，LIME 生成輸入資料的擾動版本，通過輕微修改特徵來實現。二、局部模型：然後，它在這些擾動資料點上訓練一個簡

單、可解釋的模型（例如線性回歸），以近似複雜模型在該局部區域的行為。

三、特徵重要性：簡單模型的係數或權重顯示了每個特徵對預測的影響程度。輸出顯示了哪些特徵最影響特定實例的預測[10]。例如，如果 LIME 用來解釋肺癌診斷，它可能會揭示出結節的大小和 CT 掃描中某些模式的存在對預測至關重要。

二、部分依賴圖（PDPs, Partial Dependence Plots）

部分依賴圖 顯示了特徵與預測結果之間的關係，同時保持其他特徵不變。運作方式首先 PDPs 將預測結果與特定特徵的一系列值進行繪圖，同時對其他特徵的值進行平均[11]。視覺化結果圖顯示了該特徵變化如何影響預測，PDPs 幫助視覺化單一特徵對預測的影響。例如，在預測糖尿病風險的模型中，PDPs 可能顯示增加患者的血糖水平如何影響預測的糖尿病風險。

三、累積局部效應圖（ALE, Accumulated Local Effects）

累積局部效應圖 提供了一種替代 PDPs 的方法，通過顯示特徵對模型預測的平均效應，同時考慮特徵之間的交互作用。ALE 圖計算改變特徵值的平均效應，同時調整其他特徵的影響。累積局部效應圖顯示了特徵變化時預測如何變化，這些變化在不同上下文中進行平均[12]。ALE 圖幫助理解特徵在存在其他特徵交互作用時的效果。例如，對於心臟病預測模型，ALE 圖可能顯示年齡對風險的影響如何根據其他因素（如膽固醇水平）而改變。

四、特徵重要性分數(Feature Importance Scores)

特徵重要性分數 提供了基於特徵對模型預測貢獻的整體排名，運作方式使用各種方法，例如排列重要性或模型特定技術（例如基於樹的模型的特徵重要性）來計算每個特徵的分數。然後徵根據其重要性分數進行排名，顯示它們對模型性能的整體貢獻。特徵重要性分數有助於識別哪些特徵在整個資料集上最具影響力。例如，在癌症檢測模型中，像腫瘤大小和患者年齡這些特徵可能會被排名較高，這表明它們對準確預測至關重要[13]。

五、反事實解釋（Counterfactual Explanations）

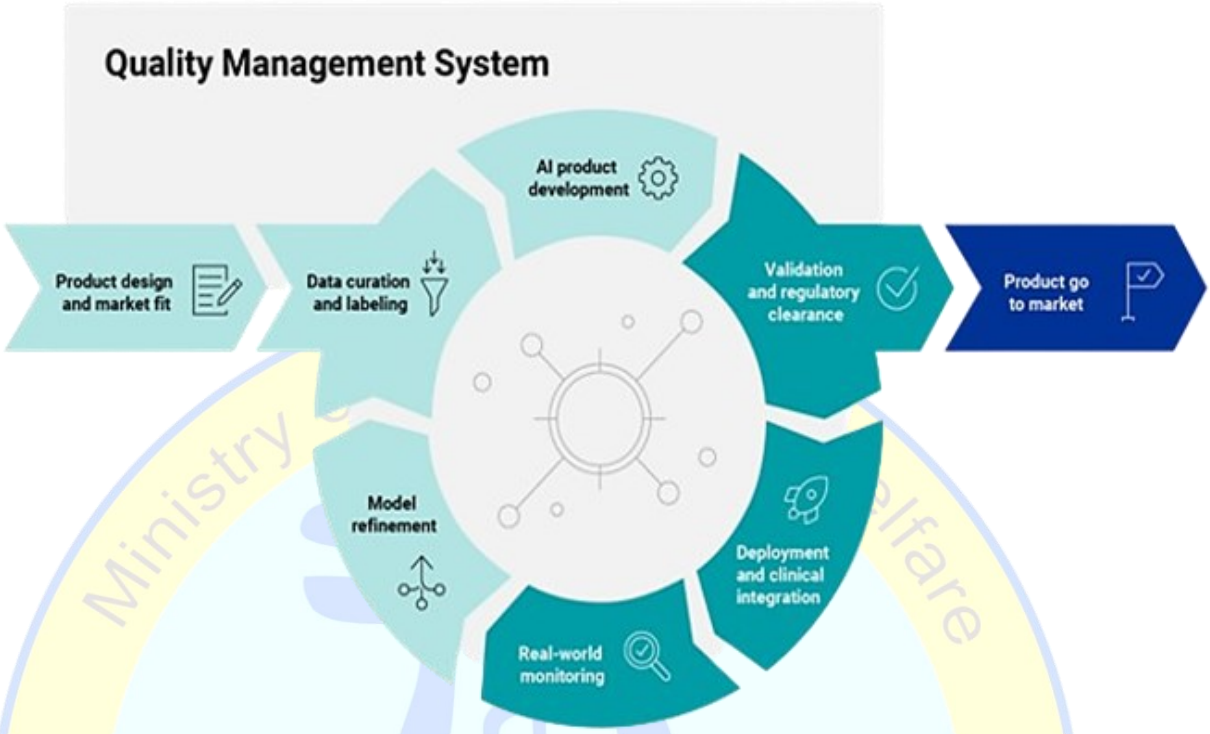
反事實解釋 描述了需要哪些最小變化才能將模型的預測改為另一個結果。場景生成：對於給定的預測，反事實解釋生成假設場景，其中輸入特徵被

稍微調整以改變預測結果，這些場景有助於了解哪些特徵變化會使預測從一個類別轉變為另一個類別。反事實解釋說明了預測距離不同結果的接近程度，以及需要什麼變化才能達到那些結果[14]。例如，如果模型預測中風的高風險，反事實解釋可能會顯示降低血壓和減少膽固醇可能將預測改為較低風險。

遵循 AI 生命週期循環監測有效性：

AI 生命週期循環監測有效性" 在臨床醫學的應用涉及到對人工智慧（AI）系統在整個生命週期中的有效性進行持續的監測和評估。這一過程不僅包括 AI 系統的開發和部署階段，還涵蓋了後續的運行、維護和改進。這樣的監測確保了 AI 系統在實際臨床環境中的表現能夠持續符合預期，並且能夠適應隨時間變化的醫療需求和資料特性，實施定期的性能監控計畫，例如每半年到一年進行一次獨立樣本的性能驗證，這樣的監測確保了 AI 系統在實際臨床環境中的表現能夠持續符合預期，並且能夠適應隨時間變化的醫療需求和資料特性[15]。

有效性監測：包括對 AI 系統的性能進行定期評估，以確保它在處理臨床問題時能夠保持準確性、可靠性和公平性。以影像診斷 AI 系統舉例，考慮一個用於檢測乳癌的 AI 影像診斷系統。這種系統在開發階段可能經過了廣泛的影像資料訓練。在實際臨床環境中部署後，需要持續監測其有效性，包括：再訓練和調整：隨著新的乳房攝影影像資料的進入，對模型進行再訓練，以保持高準確率[16]。此過程也包括對模型偏差的檢查和修正。檢查 AI 系統的診斷結果是否與專家放射科醫師的診斷結果一致，以確保系統的可靠性，收集臨床醫師和放射科醫師的反饋，並根據實際使用中的問題進行系統改進[17]。



Courtesy: https://www.massgeneralbrigham.org/en/research-and-innovation/centers-and-programs/artificial-intelligence

申請醫院資格

1. 申請補助的醫院應為醫院評鑑及教學醫院評鑑合格之醫學中心、區域和地區醫院。
2. 計畫主持人與主要成員完成衛生福利部舉辦 AI 工作坊訓練課程，並取得完訓證明者，審查評分將予以加分。
3. 醫院具備人工智慧醫療軟體落地使用經驗，審查評分將予以加分。

具體工作

1. 成立智慧醫療委員會
2. 制定智慧醫療落地實施管理辦法
3. 成立單一窗口，建立標準化電子化申請試用流程
4. 設計標準化表格審核負責任 AI 使用
5. 建立網頁供使用者可實時查詢九大透明性內容

6. 期中期末申報醫院落地使用之智慧醫療軟體與定期監測 AI 表現指標成果
7. 期中期末申報醫院使用智慧醫療軟體效益
8. 導入 AI 件數、使用場域、服務人數、使用效益（例如:診斷正確率提升、診斷周轉時間縮短）
9. 參加期末成果發表會

申請作業要點

投標資格	計畫主持人與主要成員，需完成 AI 工作坊訓練課程
審查重點	計畫主持人與主要成員 AI 工作坊完訓證明 組建團隊 醫院導入 AI 品質管理系統之規劃 盤點醫院導入 AI 品質管理系統前之醫療 AI 使用概況，包含：導入 AI 件數、使用場域、人數、使用效益等
執行重點	建立單一服務窗口與資訊平台 建構安全使用的醫療 AI 使用環境，並鼓勵醫院加速 AI 使用，透過建立本系統以利收集院內 AI 使用概況 完成醫院導入 AI 品質管理系統後之醫療 AI 使用概況評估報告，評估導入前後，是否具體提升醫療 AI 產品的使用率、使用效益(如使用 AI 件數增加、使用 AI 後可改善醫療流程、改善滿意度、提升診斷速度…等)
結案查核點	建立專責單一服務窗口並於結案前完成線上教育訓練 AI 院內使用成果分享發表 配合政策推動目標，定期登錄台灣醫院使用 AI 概況與效

	益於全國臨床 AI 應用平台

參考文獻：

1. World Health Organization. Ethics and governance of artificial intelligence for health: guidance on large multi-modal models. Accessed July 22, 2024. Available at: https://www.who.int/publications/i/item/9789240084759.
2. Shahab SB, Yarahmadi A, Hsu CC, et al. Application of explainable artificial intelligence in medical health: a systematic review of interpretability methods. *Informatics Med Unlocked*. 2023;40:101286. doi:10.1016/j.imu.2023.101286.
3. Miliard M. HTI-1 final rule now in effect, with an eye on AI. *Healthcare IT News*. Published March 13, 2024. Accessed July 21, 2024. Available at: https://www.healthcareitnews.com/news/hti-1-final-rule-now-effect-eye-ai-and-theres-more-come-says-onc.
4. HealthIT.gov. Health data, technology, and interoperability: certification program updates, algorithm transparency, and information sharing (HTI-1) final rule. Published March 7, 2024. Accessed July 21, 2024. Available at: https://www.healthit.gov/topic/laws-regulation-and-policy/health-data-technology-and-interoperability-certification-program.
5. Ahmad MA, Overman S, Allen C, Kumar V, Teredesai A, Eckert C. Software as a medical device: regulating AI in healthcare via responsible AI. In: *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*. ACM; 2021:4023-4024. doi:10.1145/3447548.3470823.

6. Interpretability of clinical decision support systems based on artificial intelligence from technological and medical perspective: a systematic review. *J Healthc Eng*. 2023;2023:9919269. doi:10.1155/2023/9919269.
7. Lundberg SM, Nair B, Vavilala MS, et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat Biomed Eng*. 2018;2(10):749-760. doi:10.1038/s41551-018-0304-0.
8. Cerekci E, Alis D, Denizoglu N, et al. Quantitative evaluation of Saliency-Based Explainable artificial intelligence (XAI) methods in Deep Learning-Based mammogram analysis. *Eur J Radiol*. 2024;173:111356. doi:10.1016/j.ejrad.2024.111356
9. Choi J, Lee S, Kim Y, et al. Feasibility of local interpretable model-agnostic explanations (LIME) algorithm as an effective and interpretable feature selection method: comparative fNIRS study. *Sci Rep*. 2023;13(1):9123.
10. Graziani M, Palatnik de Sousa I, Vellasco MMBR, et al. Sharpening local interpretable model-agnostic explanations for histopathology: improved understandability and reliability. In: de Bruijne M, et al, eds. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*. Vol 12903. Springer; 2021:576-585. doi:10.1007/978-3-030-87199-4_51.
11. Goldstein A, Kapelner A, Bleich J, Pitkin E. Peeking inside the black box: visualizing statistical learning with plots of individual conditional expectation. *J Comput Graph Stat*. 2015;24(1):44-65. doi:10.1080/10618600.2014.907095.
12. Gkolemisis V, Dalamagas T, Diou C. DALE: differential accumulated local effects for efficient and accurate global explanations. In:

Khan E, Gönen M, eds. *Proceedings of Machine Learning Research*. Vol 189. *Proceedings of the Asian Conference on Machine Learning (ACML)*; December 12-14, 2022; Hyderabad, India. PMLR; 2022:1-15.

13. Singh S, Jangir SK, Kumar M, et al. Feature importance score-based functional link artificial neural networks for breast cancer classification. *BioMed Res Int*. 2022;2022:9606314. doi:10.1155/2022/9606314.
14. Leofante F, Botoeva E, Rajani V. Counterfactual explanations and model multiplicity: a relational verification view. In: *Proceedings of the 19th International Conference on Principles of Knowledge Representation and Reasoning*; September 1, 2023:646-656. doi:10.24963/kr.2023/78.
15. Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology*. 2018;286(3):800-809.
16. Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. *N Engl J Med*. 2018;378(11):981-983.
17. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2019;17(1):195.

第二部分：臨床 AI 取證驗證中心

完成一個臨床 AI 模型後，若要在全台各大醫院使用，必須取得 TFDA 的認證。認證過程中，確保模型在外部資料上的準確性是非常重要的一環。為了顧及台灣各地醫院就診病人的多樣性，通常外部驗證需要跨系統和跨層級的醫院病患資料，在不同的醫療環境中進行外部驗證，以證明模型的準確性，確保其跨體系的適應性和一致性。這一過程同樣適用於國際的臨床 AI 模型。過去，由於缺乏系統性常設機構和專責人員來處理 AI 模型的驗證，整個流程非常緩慢。大部分工作都是在醫院的臨床工作和資訊服務之外的時間完成，而且資料集中、清理、整合的過程，往往是以單一 AI 模型的計畫為中心，缺乏系統性和標準化。這導致驗證資料通常只能一次性使用，累積的資料無法延續，形成行政負擔和浪費。

我們支持在台灣設立多個臨床 AI 驗證中心，以協助來自學界、人工智慧研發團隊及產業界的醫學 AI 產品。這些中心可以整合多家涵蓋不同層級的醫院的醫學資料，包括地區醫院和醫學中心，形成有效的驗證聯盟，使用跨體系資料庫，預先整合資料，完成驗證程序，加快產品取得 TFDA 認證的過程，加速商品化上市的速度。整個過程中，有幾個關鍵機制需要建立。

1、建立跨系統和跨層級的醫院聯盟。

2、建立聯盟共同倫理委員會的機制。

3、建立聯盟整合電子病歷資料庫

4、建立聯盟聯邦學習平台

臨床 AI 驗證中心應成常設的醫院營運單位，而不是由商業機構運作。主要是醫院運作的單位可以更好地保護病人隱私和資料安全。倫理審查重點在於資料集中的過程是否做足夠的資料去識別化、資料儲存和授權管理。這些機制將有助於建立一個高效、可靠的臨床 AI 模型驗證系統。在人力配置方面，中心需組織多學科團隊，包括臨床學家提供相關領域知識；資訊人員協助模型安裝、跨院資料整合，資料科學家協助資料清理、資料標準化、資料分析、聯邦學習、AI 模型調校；流

行病學或生物統計學顧問對驗證性能指標計算進行確認。臨床 AI 驗證中心服務必須透過單一窗口的網站和單一申請程序，使用統一的表單，由專責人員來完成，並提供進度追蹤畫面，以提升過程的透明化。此外，對於新的申請提供一些重要文件的模板，如申請倫理委員會的資料保護和隱私保護措施的文字模板，能加快整個流程的進行。資訊基礎工程部分如果可以建立專屬的資料清理工具，標準化資料萃取、清理、標準化、轉換和上傳整合程序，能夠顯著提升整體流程效率。以下是整個中心的建議服務流程。

圖一：智慧醫材取證驗證中心的服務流程

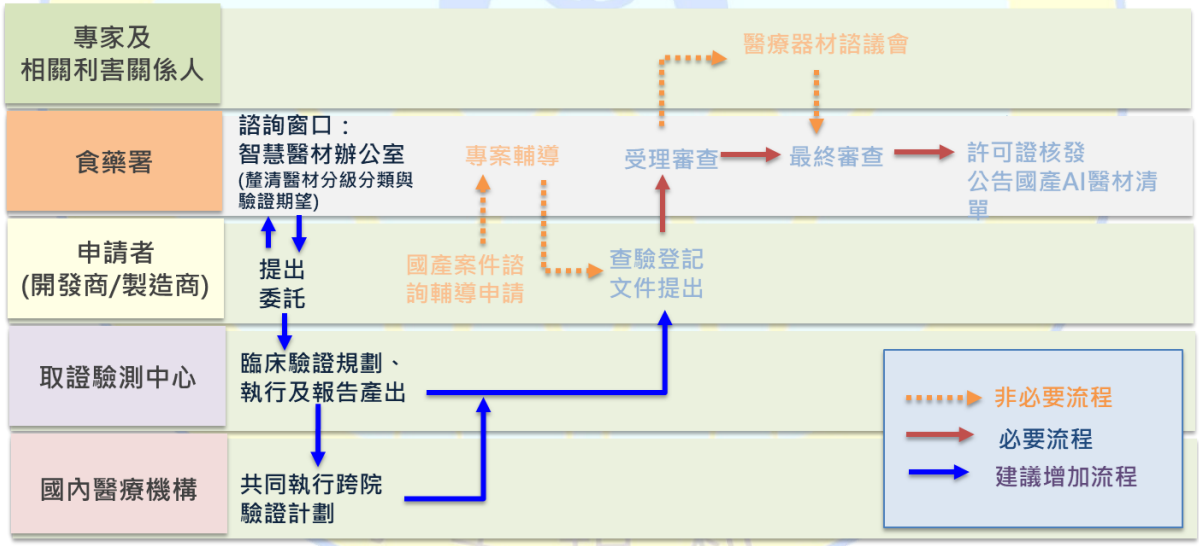
1. 申請者可向台灣的食藥署產品輔導小組提出申請，該小組會提供驗證研究的方向和期望成果。
2. 廠商或學研團隊可依照食藥署建議，到驗證中心申請服務。驗證形式分為資料集中驗證和聯邦學習驗證兩種。聯邦學習適合資料同質性較高的跨中心跨院驗證，如影像資料、實驗室結果等高度結構化資料。對於異質性較高的資料，如病例或自然語言處理應用，建議使用資料集中的方式進行驗證。

3. 中心對於資料驗證的結果，不僅提供驗證報告，對於驗證結果不佳的情況，還需提供一部分去辨識資料，協助申請單位提升模型性能，進行再驗證。
4. 多中心驗證的結果可以產出標準化的驗證報告提供申請單位和食藥署參考，醫院也可以和申請單位合作，共同發表高品質科學論文。

最後，中心在建立多資料、多中心醫學整合資料庫時，應採取不同國際資料模型，如 OP 模型，並使用 FIRE 格式整合資料。這樣的標準化程序能加速和標準化未來不同計畫的資料整合過程。聯邦學習平台常用的開源軟體工具也應預先建立，以便在新計畫進行時，免除相關行政程序，加速跨院聯邦學習及驗證的進行。

更詳細的智慧醫材取證過程，可以由下圖來表示。

圖二：智慧醫材取證流程



臨床驗證中心的主要目的是確保 AI 工具在實際醫療環境中的有效性和適應性。這個中心透過系統化的程序和標準化的資料整合，支持從研發到臨床落地的各階段。以下是根據衛福部資訊處提供的詳細規劃：

1. 建立以 FHIR 基礎的聯盟醫院整合電子病歷資料庫

建立跨醫院的驗證資料庫，是一個重要的資料基礎工程，首先要做的是各個醫院的資料標準化工程，計畫建議所採取的資料標準化的標準是 FHIR（Fast Healthcare Interoperability Resources）[1]，這也是目前也是目前國家在推動的全國電子病歷標準化所採用的標準。FHIR 由國際 HL7 協會訂立，是 21 世紀首個為醫療資料交換建立的標準。雖然前幾代標準如 HL7 v2、HL7 v3 和 CDA 仍會持續存在 5 至 10 年，但 FHIR 代表未來。同時 2023 年 3 月，美國政府頒布了新規定，要求全美的健保計畫合約醫療機構採用國際醫療資料交換標準 FHIR 以提升資料互通性。這項規定不僅影響了醫療機構的資料處理方式，還設立了保險機構和醫療資訊服務業者之間資料互通的新標準[2]。

FHIR 受到重視的原因在於它建立在以往標準的基礎上，強化了資料互通性。醫療資料標準的歷史始於第一次世界大戰前，世界衛生組織（WHO）於 1901 年推出了第一套國際疾病編碼 ICD。隨著科技進步，從紙本到電子化的轉變引發了系統間的互通問題。1987 年，國際 HL7 協會成立，並於 1994 年獲得美國國家標準協會（ANSI）認可，推出了 HL7 v2 和 HL7 v3 等標準。後來，網際網路和行動裝置的普及促使 HL7 設計了 FHIR 標準，於 2011 年推出，現已更新至第四版 R4[3]。FHIR 綜合了前幾代標準的優點，支援更多資料格式，並使用 HTTP 協定的 RESTful API 來提升資料交換的靈活性和效率。

FHIR 由一系列 Resource（資源物件）組成，這些 Resource 像資料庫中的資料表，或像 Excel 中的工作表，用於記錄和管理資料。FHIR R4 標準涵蓋 145 種 Resource，分為基礎、基本、臨床、財務和特殊領域五大類，每類下又細分為 24 類，如專門詞彙、安全性、工作流程等。HL7 協會網站上提供了每種 Resource 的詳細定義和範例，幫助用戶理解和應用這些資源[3]。

FHIR 不僅僅是資料的標準，還有一些用於網路平台上資料轉換的程式工具，所以可以把它看成是不同醫院可以互相做資料交換的轉譯器，所以在跨院驗證的資料庫

能夠採取 FHIR 標準的話，就可以把不同醫院異質性的資料做同質化，整合的過程要，可以透過伺服器常見的伺服器例如 HAPI FHIR Server 、Microsoft Azure API for FHIR，Google Cloud Healthcare API，這些都可以用來作為相關的資料交換[4,5]。在建立驗證中心時，建議利用 FHIR 標準來進行資料整合，這樣可以為未來國際合作或國內醫院間的跨院合作奠定基礎。這不僅能節省大量資料整合的時間，還能提升整體資料品質。資料整合的過程一般包括以下步驟，並有對應的工具，可以由資訊團隊建立相關流程，以逐步實現標準化、自動化和模組化的跨中心資料整合：

(1) 資料模型選擇：決定使用哪些特定的 FHIR 資源。FHIR 提供了以標準化方式電子交換醫療資訊的方法。常見的資源包括病人、就診、觀察、病情等。

(2) 資料萃取 (Extraction)：從每家醫院的電子健康記錄 (EHR) 系統中萃取 FHIR 資料。每家醫院可能有自己的格式和結構，需要映射到 FHIR 標準。

(3) 資料轉換 (Transform)：使用適當的工具或腳本將萃取的資料轉換為 FHIR 資源。這一步確保所有資料都符合 FHIR 標準。

(4) 資料上傳 (Load)：將來自多家醫院的 FHIR 資料整合到中央資料庫或資料儲存庫中。這個資料庫將存儲來自所有參與醫院的 FHIR 資源。

(5) 資料儲存：選擇一個適合的資料庫系統來支持 FHIR 資源。常見的選擇包括：FHIR 伺服器：專門的 FHIR 伺服器 (例如，HAPI FHIR、Smile CDR) 設計用來存儲和提供 FHIR 資源。

當使用 FHIR 整合的資料進行後續驗證時，需要利用不同的資料模型進行進一步的處理。因為在各種研究領域中，會用到許多不同的資料模型 (data model)。例如，OMOP CDM 是臨床觀察性研究中最常見的資料模型，目前在 OMOP 社群內，有許多研究已經開發出來，同時也有不少工具可以用於 FHIR 和 OMOP 資料模型之間的轉換[6]。醫學 AI 研究的內容相當多樣化，因此除了 OMOP 這個常見的資料模型外，

還有許多聯盟開發了適用於不同資料類型和研究目的的工具。以下是一些資料模型範例。

(1) OMOP CDM：全文為 The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM)，主要用於標準化和協調觀察性健康資料，使得多中心研究和比較研究更為便捷[7]。

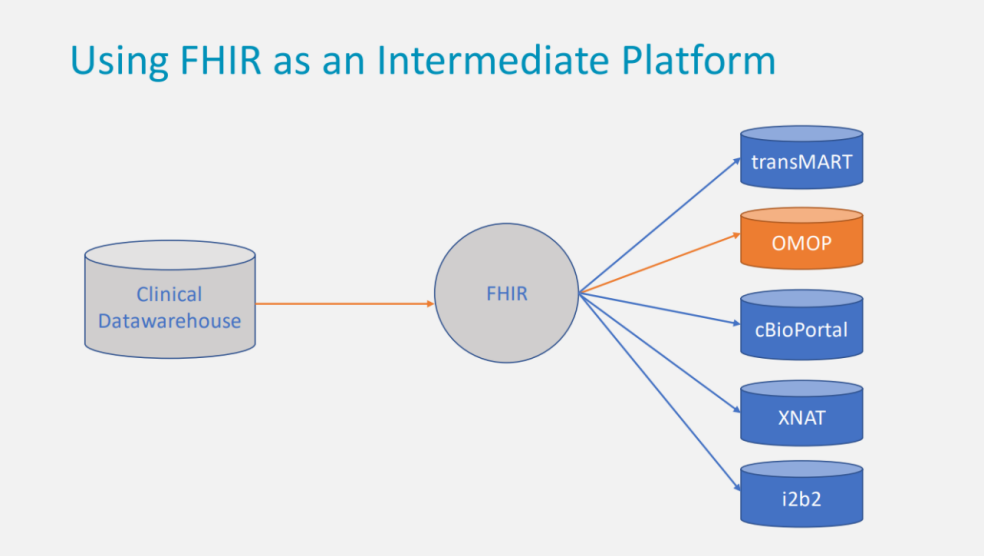
(2) cBioPortal：專注於癌症研究，提供分析和可視化癌症基因組資料的工具，通常會整合使用像 OMOP 這樣的標準化資料集[8]。

(3) XNAT：XNAT 是一個專為影像研究設計的可擴展開源影像資訊學軟體平台，對於影像資料，尤其是神經影像學研究至關重要，並且可以與其他資料類型整合以進行全面分析[9]。

(4) i2b2：i2b2 tranSMART 基金會是一個由會員驅動的非營利基金會，基金會透過分享、整合、標準化和分析來自醫療和研究的異質資料，以及通過動員和參與以生命科學為重點的開源、開放資料社群，促進精準醫療的有效合作。致力於圍繞 i2b2、tranSMART 和 OpenBEL 跨領域研究平台開發開源/開放資料，社支持臨床和組學資料(omics data)的整合和探索，促進轉譯研究並推動合作[11,12]。

所以整個醫學研究資料庫可以使用 FHIR server 當作是一個資料中台，再透過不同的資料模型轉換程式，對接到國際不同領域通用的資料模式，最後再進行的臨床 AI 的跨醫院驗證。雖然轉換這些國際不同聯盟的資料格式跨院驗證並非必要，但是，利用這些資料模式整理資料，可以促進資料的二次應用和國際合作，未來也可以跨聯盟行成龐大的電子病歷資料庫，做進一步的開發和應用。

圖三：FHIR 中介平台整合臨床資料至 transMART、OMOP、cBioPortal、XNAT 和 i2b2 資料模型



建立資料庫需要跨領域的人才，因此人才訓練相當重要，衛福部辦公室也會定期舉辦驗證中心的成果分享，讓大家的經驗得以快速累積。

2. 建立聯邦學習平台

聯邦學習是一種去中心化的機器學習方法，允許多個機構在不共享資料的情況下協同訓練模型。在臨床 AI 驗證中，這種技術特別有利，因為資料隱私和安全性至關重要。聯邦學習的優點包括資料隱私和安全性。聯邦學習允許各機構將資料保留在本地，從而解決隱私問題，並遵守 GDPR 和 HIPAA 等法規。此外，通過協作，機構可以在更多樣化的資料集上訓練模型，提高模型的泛化能力和穩定性。由於資料不會在機構之間傳輸，因此在傳輸過程中資料洩露的風險最小化。機構可以利用自己的計算資源進行訓練，可能減少對集中式高性能計算基礎設施的需求。

然而，聯邦學習也有其缺點。首先，設置和維護聯邦學習系統可能技術上具有挑戰性，需要基礎設施和參與機構之間的協調。資料的異質性，即不同機構的資料格式、質量和分佈差異，可能會使模型訓練變得複雜，並需要高級技術來處理這種變化。此外，聯邦學習需要中央伺服器與本地節點之間的頻繁通信，這可能會消耗大

量頻寬並減慢訓練過程。對於技術專業知識和資源有限的較小機構來說，可能難以有效參與聯邦學習。總而言之，聯邦學習為臨床 AI 驗證提供了一種在資料隱私需求與協作模型訓練的益處之間取得平衡解決方案。

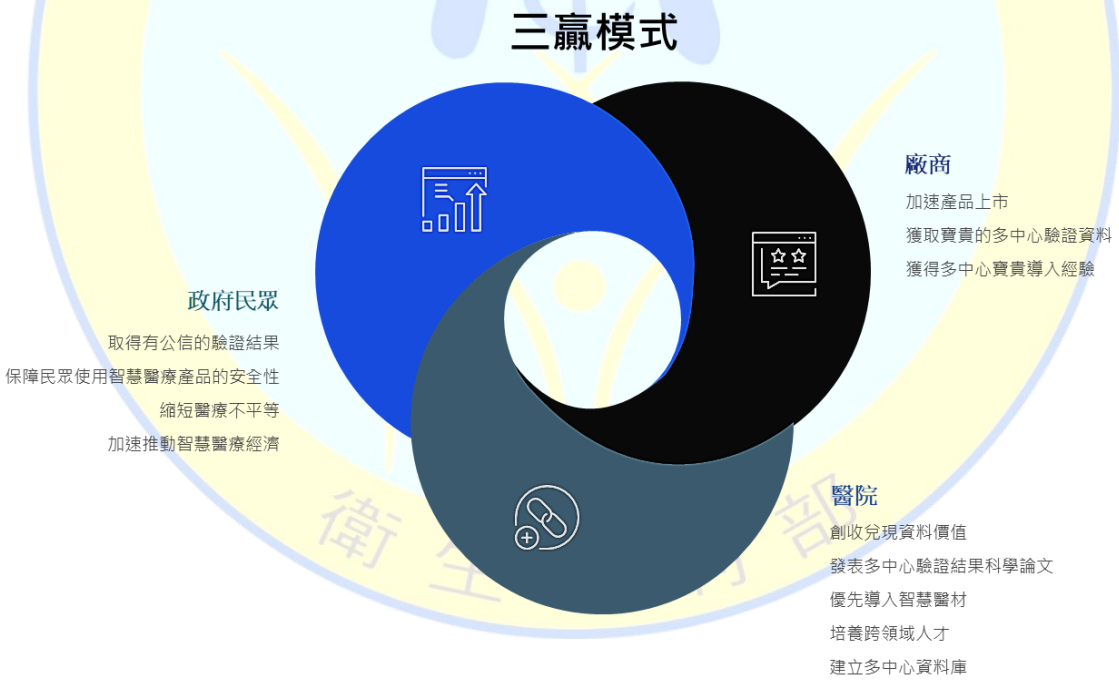
在美國醫療中心中，最廣泛使用的開源聯邦學習平台之一是 FLOWER（Federated Learning Object-oriented Environment for Research）。FLOWER 是一個靈活且功能強大的開源框架，專為聯邦學習的研究和實施而設計。它支援多種機器學習框架，並允許用戶方便地設計、實施和測試聯邦學習算法。FLOWER 的架構能夠有效協調多個醫療機構之間的協作，保障數據的隱私同時實現模型的聯合訓練。另一個廣泛使用的開源聯邦學習平台是 TensorFlow Federated（TFF）。TFF 是由 Google 開發的，提供了一個靈活且可擴展的環境來支持聯邦學習。它允許各機構在不需要直接共享數據的情況下，共同訓練機器學習模型，這對於保護病人隱私尤其重要。

在聯邦學習中，常見的聚合算法包括 FedAvg（Federated Averaging）、FedProx（Federated Proximal）和 SplitNN（Split Neural Networks）。這些算法在實際應用中具有不同的優勢和適用場景。FedAvg 是最基礎的聚合算法之一。其原理是每個本地機構在自己的數據上獨立訓練模型，然後將模型的更新（即權重或梯度）發送到中央伺服器。中央伺服器將這些本地模型的更新進行加權平均，更新全局模型，再將更新後的模型分發到各個本地機構。這種方法特別適用於醫療領域的疾病預測和醫學影像分析，因為它能夠在保護數據隱私的情況下，利用來自不同機構的數據進行協同訓練。FedProx 是 FedAvg 的擴展，旨在解決在聯邦學習過程中出現的數據非獨立同分佈（non-i.i.d.）問題。FedProx 引入了一個正則化項，限制本地模型和全局模型之間的差異，使得本地模型更新更加穩定。這對於數據分佈不均或機構之間存在顯著差異的情況尤其有效。SplitNN 則是一種適合處理複雜模型的方法，它將模型分為前端和後端兩部分。前端部分由本地機構處理，主要負責特徵提取，而後端部分由中央伺服器處理，負責最終的分類或回歸分析。這種方法能夠減少本地機構的計算負擔，同時保持數據的隱私性。SplitNN 常用於大規模神經網絡模型的訓練，特別是在需要大量計算資源的場景下。

這些開源聯邦學習平台和聚合算法提供了靈活而強大的工具，幫助醫療機構在保護數據隱私的同時，進行高效的機器學習模型訓練，推動了醫療人工智慧的進步。

在數位醫療的未來，人工智慧的研發將形成龐大的工作和市場。即早在台灣建立相關的驗證機制，除了能讓病人早期受惠於先進科技的效益，還能促進國內相關產業的發展。對醫院來說，也有很多好處，包括醫師可以參與高品質的科學研究產出，醫院也可以透過驗證過程測試 AI 產品的效益。透過本計畫的支持，前兩年的申請案件可由政府經費協助完成驗證過程。計畫結束後，醫院可透過學習建立整個機制程序，計算整個過程所需的成本，做出合理定價，開始對內外部顧客進行營運。

圖四：驗證過程測試 AI 產品的三贏模式



申請醫院資格

1. 醫院計畫主持人與主要成員，需完成衛福部舉辦 AI 工作坊訓練課程，並取得完訓證明

2. 申請補助的醫院應為經衛生福利部醫院評鑑優等以上、醫院評鑑合格之醫學中心或區域醫院、醫院評鑑及教學醫院評鑑合格之全民健康保險特約醫院
3. 需組建跨體系、跨層級醫院聯盟，由主責醫院申請
4. 符合資通安全責任等級分級辦法§11-全國法規資料庫(moj.gov.tw)。

具體工作

1. 建立對外公開網頁
2. 成立單一窗口，建立標準化申請電子表單
3. 建立標準化驗證流程
4. 建立「中央審查委員會」(CIRB，Central Institutional Review Board)，統一進行倫理審查
5. 建立聯邦學習平台
6. 透過FHIR，以TW Core 為核心進行資料串接，建立跨院電子病歷資料庫
7. 建立電子病歷資料庫驛碼檢索字典
8. 發展資料清理自動化工具
9. 協助建立符合Smart on FHIR規格之AI應用程式
10. 接受學界或廠商委託，依據TFDA驗證諮詢結果，
11. 建立具有人群涵蓋性、足夠樣本數之驗證樣本
12. 依AI應用程式性質選擇聯邦學習或資料集中驗證
13. 驗證成果未達預期，提供去部分去個資料協助調校模型
14. 進行統計分析完成有效性、安全性評估分析
15. 建立中心品質指標，包含各階段行政流程時效，分析方法正確性，分析報告內容完整度
16. 期末建立相關收費機制
17. 配合衛福部收集AI相關商品化成果統計數字

18. 參與期末成果報告

申請作業要點

投標資格	計畫主持人與主要成員，需完成 AI 工作坊相關訓練課程
審查重點	計畫主持人與主要成員 AI 工作坊完訓證明 組建團隊 組建跨體系/跨層級之聯盟醫院 AI 產品取證驗證中心建置與運作規劃 具通過 TFDA 查驗登記之人工智慧醫療產品經驗(加分)
第一年執行重點	建立單一服務窗口與資訊平台 建立 AI 產品取證驗證中心之執行能力 聯盟醫院需建立 CIRB 架構 完成聯盟醫院資料串接，資料需為 FHIR 格式，並符合 TW Core IG 規範 建立專人負責與內部時效管理流程 建立組織營運成本與服務模式，以利獨立營運創收 啟動 AI 產品取證驗證案源 1 案以上
第二年查核點	建立具時效性之運作實績，完成 AI 產品驗證評估報告(若結果不符預期，仍須提供廠商參考以調校優化產品) AI 產品驗證經驗分享 完成臨床學術報告 完成組織營運創收規劃 促成驗證多中心資料庫二次利用

參考文獻：

1. Li Y, Wang H, Yerebakan H, Shinagawa Y, Luo Y. Enhancing health data interoperability with large language models: a FHIR study. Presented at: 2024 AMIA Annual Symposium; 2024. Available at: https://arxiv.org/abs/2310.12989. Published online September 19, 2023. Accessed July 22, 2024.
2. Henke E, Peng Y, Reinecke I, Zoch M, Sedlmayr M, Bathelt F. An extract-transform-load process design for the incremental loading of German real-world data based on FHIR and OMOP CDM: algorithm development and validation. *JMIR Med Inform*. 2023;11 . Published online August 21, 2023. doi:10.2196/47310.
3. Health Level Seven International. Welcome to FHIR. Accessed July 22, 2024. Available at: https://hl7.org/fhir/R4/.
4. Cloudticity. Azure FHIR services vs. Google Cloud Healthcare API: which one is right for you? Accessed July 22, 2024. Available at: https://blog.cloudticity.com/azure-fhir-services-vs.-google-cloud-healthcare-api-which-one-is-right-for-you.
5. Observational Health Data Sciences and Informatics. Standardized data: the OMOP common data model. Accessed July 22, 2024. Available at: https://www.ohdsi.org/data-standardization/.
6. Stanford Medicine. Stanford electronic health records in OMOP. Accessed July 22, 2024. Available at: https://med.stanford.edu/starr-omop.html.
7. cBioPortal for Cancer Genomics. Memorial Sloan Kettering Cancer Center. Accessed July 22, 2024. Available at: https://www.cbioportal.org/.
8. XNAT. Accessed July 22, 2024. Available at: https://www.xnat.org/.

9. i2b2: Informatics for integrating biology and the bedside. Available at: https://www.i2b2.org/. Accessed July 21, 2024.
10. Bierkens M, van der Linden W, van Bochove K, et al. tranSMART. *J Clin Bioinforma*. 2015;5(Suppl 1) . Published May 22, 2015. doi:10.1186/2043-9113-5-S1-S9.
11. The Hyve. Multi-omics data analysis in tranSMART. Accessed July 22, 2024. Available at: https://www.thehyve.nl/articles/multi-omics-data-analysis-in-transmart.

第三部分：AI 影響性研究中心

邁向資料經濟的過程中，有一個關鍵就是要有科學性的定價基礎，以確定 AI 在臨床和醫療經濟上的價值[1]。目前台灣已有三十多個臨床 AI 項目獲得台灣食藥署核發許可證，但是許可證的核發不代表健保給付的保證。如何決定智慧醫材是否可以獲得健保給付，以及健保給付的價格和方式，是世界各國醫療政策的重大挑戰，即使走自費定價路線，如何客觀的評價臨床 AI 的價值，也是一個困難的課題，因為智慧醫材的定價無法採取傳統醫療器材從製造端進行評估成本。

為了解決智慧醫材的定價與其醫療價值的關係，我們將補助建立臨床 AI 影響性研究中心。這個中心將對接健保署，並成立單一窗口，簡化行政流程。我們希望透過跨體系、跨層級的臨床試驗中心聯盟來進行 AI 臨床試驗，透過實驗組和對照組分組的比較，以探討 AI 在臨床上帶來的醫療價值及經濟效益。傳統上，隨機臨床試驗是由藥廠結合 CRO（臨床研究機構）來設計、執行與分析。然而，這個模式在人工智慧醫材領域無法套用。原因在於，AI 醫材的隨機臨床試驗涉及醫院資訊系統的應用程式導入，還有大量病人就醫資料的處理。將這些核心系統和資料交給商業化的 CRO 公司來執行，會對資安和病患隱私造成相當大的風險。

因此，針對 AI 智慧醫材的隨機臨床試驗，最適合由醫學中心組成多中心聯盟，來進行臨床試驗的設計、執行和分析。一個單位若要執行這樣的影響性評估，需要許多專才。首先，需要多中心先行形成聯盟；其次，要建立共同的倫理委員會審查標準；再來，要有流行病學家、生物統計學家和資料科學家等多學科人員組成的隨機臨床試驗方法學諮詢小組，來協助試驗設計和分組；最後，研究結果的分析也需要生物統計學家或流行病學家來進行，並且結果需要醫療經濟學家來做醫療經濟分析。這些資料可以提供健保委員會一個堅實的本土資料，來計算臨床 AI 醫材在台灣醫療環境下，是否能夠確實產生病患或醫療經濟上的效益，從而有科學基礎來做相關的給付決策或定價標準。

FDA 證實準確但是臨床試驗無法證實臨床效益的案例

在臨床 AI 的驗證過程中，第一階段主要是食藥署驗證其正確性。舉例而言，使用人工智慧來偵測乳房攝影早期乳癌病兆的應用已經很成熟，國際上許多產品也獲得美國 FDA 認證，然而大規模的隨機臨床試驗資料直到最近才發表。瑞典 Lund University 放射科研究員 Kristina Lång 2023 年於 Lancet Oncology 發表了一個隨機對照試驗評估臨床 AI 輔助乳房攝影篩檢 AI 系統的醫療價值[2]。研究目的是評估 AI 輔助乳房攝影篩檢方案與標準由放射科醫師進行的雙重讀片相比的臨床安全性和有效性。試驗類型是群體隨機對照（cluster randomized control trial）、以醫師為主要分組對象，醫師隨機分配（1:1）到 AI 輔助篩檢組（介入組）或標準雙重讀片組（對照組），比較 AI 輔助的乳房攝影篩檢與標準雙重讀片的效果。隨機方法使用隨機數生成器在 PACS 系統中隨機分配篩檢影像。參加的受試者年齡在 40 至 80 歲之間的女性包括每 1.5 至 2 年進行一般篩檢，以及對於有中度遺傳性乳癌風險或乳癌病史的女性進行每年篩檢。AI 系統使用 Transpara 1.7.0 版本來生成基於 10 級的惡性風險評分。這些評分被用來將篩檢影像分配為單讀（評分 1-9）或雙讀（評分 10）。篩檢過程在 AI 輔助組中，放射科醫師可以查看 AI 風險評分（針對所有影像）和 AI 輔助檢測標記（針對評分為 8-10 的影像）。結果測量主要終點為癌症檢出率（每 1,000 名篩檢參與者檢出的癌症數）。次要終點為召回率、假陽性率、召回的陽性預測值（PPV）、癌症類型（侵襲性或原位癌）及讀片工作負擔[2]。

在 2021 年 4 月 12 日至 2022 年 7 月 28 日期間招募了 80,033 名女性。分析排除 13 名參與者，共分析 80,020 名參與者。結果 AI 輔助組癌症檢出率每 1,000 名篩檢參與者中檢出 6.1 例癌症（95% CI 5.4–6.9），標準雙重讀片組：每 1,000 名篩檢參與者中檢出 5.1 例癌症（95% CI 4.4–5.8），比例：1.2（95% CI 1.0–1.5；p=0.052）。召回率 AI 輔助組：2.2%（95% CI 2.0–2.3），標準雙重讀片組：2.0%（95% CI 1.9–2.2）。癌症類型，AI 輔助組：244 例癌症中，75%為侵襲性癌，25%為原位癌。標準雙重讀片組：203 例癌症中，81%為侵襲性癌，19%為原位癌，AI 輔助組減少了 44.3%的工作負擔[2]。

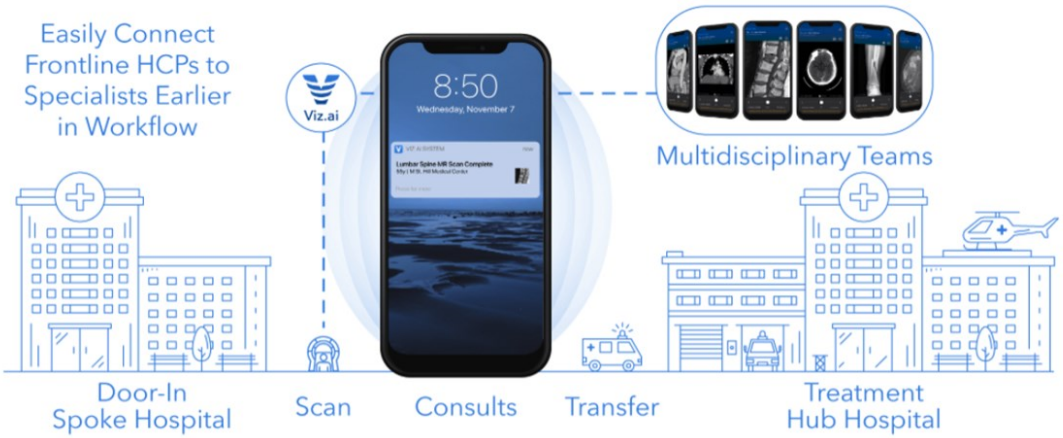
綜合上述的研究結果顯示 AI 輔助組雖然早期癌症偵測率稍高，但與對照組相比，統計上並無顯著差異。然而 AI 輔助組工作負擔減少顯著。這案例顯示 AI 能提升診斷效率和醫師滿意度，但未必能改善病人的預後。

FDA 證實準確同時臨床試驗證實臨床效益的案例

臨床試驗的評估指出，大部分健保項目主要集中在診斷層面。然而，僅有診斷並不足以改善病人的預後，因此常需結合一些介入措施。以美國首例健保給付的智慧醫材案例 Viz LVO ContaCT 這款產品為例，它是一個結合社區和中風中心的解決方案[3,4]。當社區醫院接診到有中風症狀的病人時，病人在社區醫院進行大腦 CT 血管造影，這款人工智慧軟體可以迅速判斷判讀大腦 CT 血管造影影像，Viz LVO ContaCT 產品的準確度與醫師相當，ROC（接收者操作特徵曲線）的曲線下面積（AUC）為 0.91，順利通過食藥署認證。

Viz LVO ContaCT 將懷疑大血管阻塞的通知發送給中風中心的神經血管專家，建議他們檢視這些影像，透過手機應用程式預覽的影像是壓縮過的，僅供參考，不可用於診斷。收到通知的臨床醫師負責在電腦上檢視未壓縮的影像，病人是否存在大血管阻塞。如果確認有大血管阻塞，並且在黃金時間內、病人的神經學症狀符合治療標準的情況下，中風中心立即開啟綠色通道，將病人轉診到中風中心進行溶栓手術。

圖一：智慧醫材案例使用架構



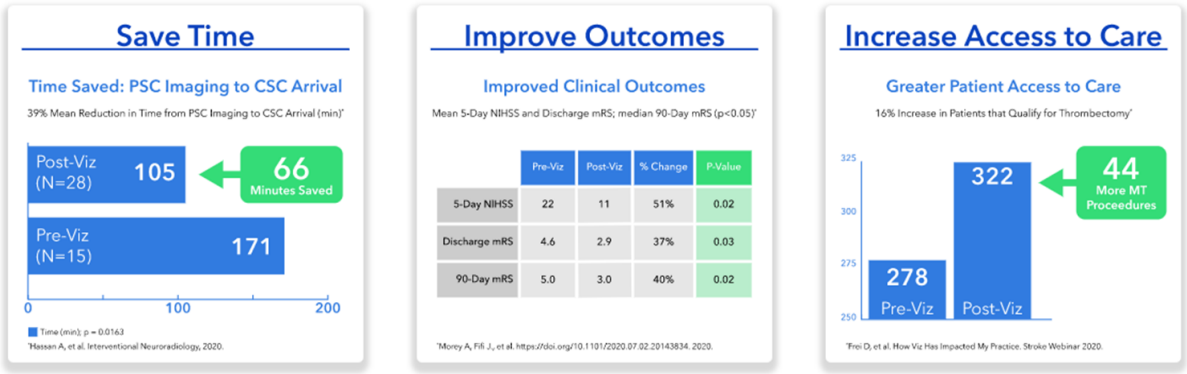
Viz alerts multidisciplinary care teams earlier in the workflow, coordinating care by connecting frontline health care professionals (HCPs) to specialists facilitating efficient communication and coordinating care.

Courtesy: https://www.viz.ai/large-vessel-occlusion

為了申請美國的健保給付，該產品進行了一項臨床試驗，探討導入這一套人工智慧解決方法以後中風病人腦神經預後的改善。根據 2020 年發表於國際中風會議的研究摘要，導入 Viz LVO 後，從地區醫院接診到病人抵達中風中心的時間縮短了 39%，平均減少了 66 分鐘。神經學預後也顯著改善，五天內的神經學評估進步了 51%，出院時的神經學評估提升了 37%，而 90 天後的神經學進步則達到 40%。這些改善在統計上皆具有顯著性[3]。此外，導入前僅有 278 名病人受益於溶栓手術，而導入後受益人數增加至 322 人，增加了 44 人[3]。此外，樣本數更大的臨床驗證研究，發表於 2023 年《Stroke》的研究，導入 Viz LVO 後，從地區醫院接診到神經介入醫師（NIR）通知的時間顯著縮短。這項研究共納入 14,116 名患者，其中 8,557 名使用 AI 平台，5,559 名未使用。使用人工智慧（AI）平台的醫院，患者從到達醫院到 NIR 通知的中位時間為 50 分鐘，而未使用 AI 的醫院則為 89.5 分鐘，兩者之間的差異具有統計學意義（$p < 0.001$）[4]。這些結果意味著導入 AI 平台後，時間縮短了約 39.5 分鐘[4]。此外，使用 AI 平台的醫院在溶栓治療的時間（DTN）上也顯示出改善，DTN 時間中位數為 40 分鐘，而未使用 AI 的醫院則為 44 分鐘（$p = 0.018$）[4]。研究還發現，AI 組進行了更多的高級成像，檢出更多的大血管閉塞（LVO），且接受介入治療的患者比例更高。這些結果顯示，AI 技術

的應用不僅加快了急性中風患者的處理速度，還提高了治療的效率和患者的整體預後[4]。

圖二：智慧醫材案例驗證結果說明



Courtesy: https://www.viz.ai/large-vessel-occlusion

完成這些臨床試驗後，Viz LVO 獲得了美國 CMS 健保局的首個給付案，即每個使用 Viz LVO 的案例可獲得 1040 美元的額外健保給付。這樣的給付被納入 DRG 包裹式健保給付制度中，意即中風病人在治療過程中若使用這款人工智慧軟體進行早期診斷，並結合後續的轉診中風中心的流程，若能改善關鍵品質指標，即可獲得這項給付。而這種給付方式是滾動調整的，若未達到設定的品質指標，給付將無法生效。

此案例為我們提供了良好的示範，顯示診斷工具並非無法進行臨床試驗。事實上，診斷工具能夠及早治療並改善病人結果的介入方案，在臨床試驗中若能取得良好成果，就有機會獲得健保給付。

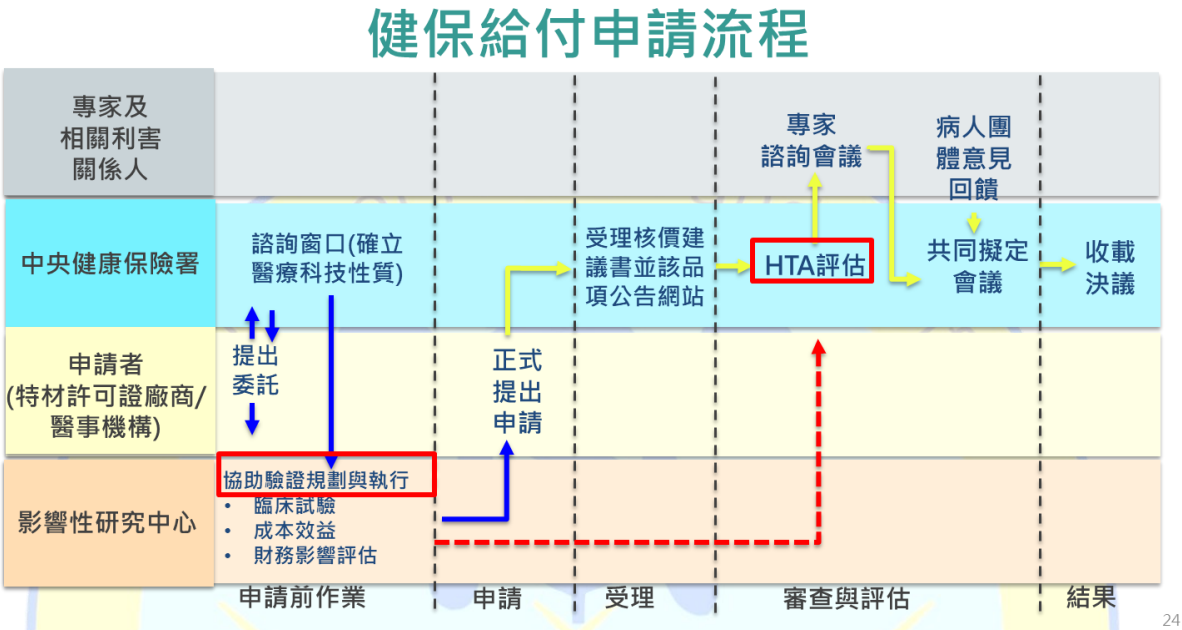
AI 影響性研究中心的服務流程

我們的試驗流程和健保署整合，如果一個醫材已經取得證照且要進行商業化，產品取證後可先到健保署諮詢，健保署的統一諮詢窗口會為該產品邀集專家，或由內部專家給予相關建議方向，根據健保署的建議回饋，再到影響性研究中心申請服務，協助進行臨床試驗設計和執行來做臨床效益評估。申請方主要工作是提供這項智慧醫材軟體的性能表現資料和臨床適用情境。申請方可以指定影響性評估中心體系內的相關領域專家，或由中心推薦一個該領域的專家作為總主持人，邀集其他相關領域的協同主持人，組成研究團隊來進行臨床試驗設計、倫理委員會申請和執行，在試驗結果完成後，會撰寫相關的科學報告，進行醫療經濟評估，並將報告交給申請單位。

圖三：影響性研究中心的試驗流程和健保署整合

更詳細的過程，可以由下圖來表示。

圖四：智慧醫材健保給付申請流程



AI 的臨床試驗方法

AI 的臨床試驗方法與新藥臨床試驗有所不同。首先，AI 臨床試驗通常對整個系統進行隨機分配，而非個人[5]。其次，AI 主要用於診斷，若無輔助介入措施，難以達到臨床預後的重大改善[5]。為此，我們補助 AI 影響性研究中心，負責臨床研究設計的顧問，並提供設計服務，與臨床醫師合作，結合診斷和介入流程進行系統性介入。適用於 AI 介入的臨床試驗主要有三大類方法：前後比較、群體隨機臨床試驗和楔形階梯設計。

對照前後試驗（Controlled Before-After Trial, CBAT）

旨在比較介入前後的結果，並設有對照組來進一步檢驗介入效果。在這種設計中，一部分群體在介入前後進行評估，而另一部分群體作為對照組，只進行常規診斷而不接受介入[6]。選擇多個醫院或診所作為研究群體，將群體隨機分配到介

入組（在介入前後進行評估）和對照組（僅接受常規診斷）。確保介入組在實施 AI 診斷工具前後都有明確的觀察時間段，對照組則在相同時間段內進行觀察。

參與研究的患者不知自己所屬的群體，以減少患者行為對結果的影響。 雙盲法：如果可能，醫護人員和資料分析者都應盡量保持盲目，不知患者是否屬於介入組或對照組，以減少分析偏差。樣本量計算應考慮效果大小、群內相關性和研究設計特點。根據預期的診斷準確率提高、敏感性或特異性變化來估算效果大小。使用適合前後試驗設計的樣本量計算公式，並考慮群內相關性（Intra-cluster correlation）和設計效果，以確保樣本量足以檢測到統計顯著的效果。

主要終點分析

Difference-in-Differences (DiD) 分析方法

Difference-in-Differences（DiD）是一種計量經濟學方法，用於評估政策或介入措施的效果。DiD 通過比較介入前後兩組（實驗組和對照組）之間的變化，來估計介入的因果效應[7]。

設置回歸模型，將結果變數（如診斷準確率、敏感性或特異性）作為因變數，將群體類別（實驗組或對照組）、時間段（前或後）以及兩者的交互項作為自變數[8]。估計效應：交互項的係數代表介入的平均處理效應（Average Treatment Effect, ATE），即 AI 診斷工具的影響。

$$Y_{it} = \alpha + \beta_1 \text{Post}_t + \beta_2 \text{Treatment}_i + \beta_3 (\text{Post}_t \times \text{Treatment}_i) + \epsilon_{it}$$

Yi_t：表示第 i 個觀察單位在第 t 個時間點的結果變數（例如診斷準確率、健康指標等）。

a ：回歸模型的常數項。

$Post_t$：時間指標變數，用於區分介入前（0）和介入後（1）的時間段。

$Treatment_i$：群體指標變數，用於區分實驗組（1）和對照組（0）。

$Post_t x Treatment_i$：介入效果的交互項，即時間指標和群體指標的乘積項。

ε_{it}：誤差項，表示其他未觀察到的因素對結果變數的影響。

$\beta 1$ 表示在介入後，對照組的結果變數變化。這是介入後時間段對結果變數的影響，但不考慮群體是否接受介入。

$\beta 2$ 表示在介入前，實驗組和對照組之間的差異。這反映了群體間在介入前的基線差異。

$\beta 3$ 是 Difference-in-Differences 的關鍵，稱為介入效應。它表示在介入後，實驗組相對於對照組的變化。簡單來說，這個參數衡量了介入對結果變數的額外影響，即介入的因果效應。

中斷時間序列分析（ITS, Interrupted Time Series Analysis）

中斷時間序列分析（Interrupted Time Series Analysis, ITS）是一種統計方法，用於評估介入措施在特定時間點前後對一系列資料的影響[9]。ITS 通過分析介入前後的趨勢變化，來估計介入的效應。

應用於 AI 診斷工具的前後試驗

在介入前後的多個時間點收集實驗組和對照組的結果資料，分析介入前的趨勢並將其延伸至介入後，以此作為預期趨勢。設置回歸模型，將時間、介入變數（介入前或介入後）以及兩者的交互項作為自變數，結果變數（如診斷準確率、敏感性或特異性）作為因變數，分析介入後的趨勢變化和水平變化，以評估 AI 診斷工具的影響。

$$Y_t = \alpha + \beta_1 \text{Time}_t + \beta_2 \text{Post}_t + \beta_3 (\text{Time}_t \times \text{Post}_t) + \epsilon_t$$

Y_t：表示第 t 個時間點的結果變數（例如診斷準確率、敏感性、特異性等）。

α：回歸模型的常數項，表示基線水平。

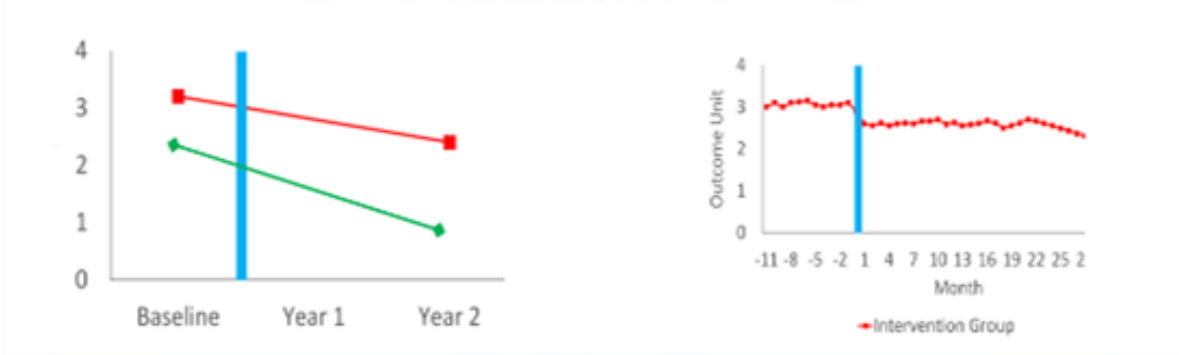
Time_t：時間指標變數，表示時間的順序，例如從介入前的時間點 1 到介入後的時間點 t

Post_t：介入指標變數，介入後的時間點為 1，介入前的時間點為 0。

$Time_t \times Post_t$：時間和介入的交互項，表示介入後趨勢的變化。

ϵ_t：誤差項，表示其他未觀察到的因素對結果變數的影響。

DiD 分析和 ITS 分析均可應用於 AI 診斷工具的對照前後試驗，以評估其效應。DiD 分析側重於比較實驗組和對照組的變化（下圖左），而 ITS 分析側重於介入前後的趨勢變化(下圖右)。這兩種方法能夠提供不同角度的證據，幫助全面評估 AI 診斷工具的有效性。



群集隨機對照試驗

群集隨機對照試驗（Cluster Randomized Controlled Trial, CRCT）是將整個群體（如醫院或診所）隨機分配到不同的試驗組別，而不是將個別患者隨機分配[10]。這種設計適用於 AI 診斷工具的評估，因為能夠減少交叉污染，並更真實地模擬實際應用情況[10]。

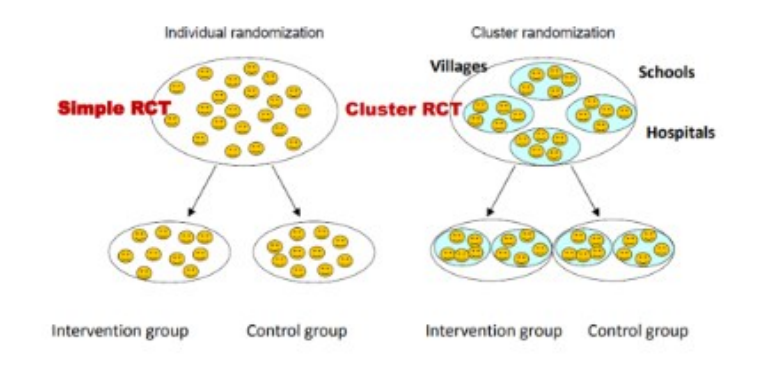
首先選擇多個醫院或診所作為群集，利用隨機的過程，將這些群集隨機分配到實驗組（使用 AI 診斷工具）和對照組（使用傳統診斷方法）。隨機化可以使用隨機數生成器或電腦軟體來進行。隨機化過程要確保各組別在基線特徵（如醫院規模、地理位置、病人特徵等）上平衡相似，以減少偏倚。參與研究的患者不會知道自己所屬的群集是實驗組還是對照組，從而減少患者行為對結果的影響。如果可能，醫護人員也應該盲於患者所屬的組別，這樣可以減少醫護人員的行為對診斷結果的影響。然而，由於 AI 工具的明顯特性，完全雙盲可能難以實現。

樣本量計算應考慮群內相關性（intracluster correlation coefficient, ICC）以及效果大小。常用的計算方法包括：計算個體樣本量：使用常規樣本量計

算公式。調整群內相關性：使用設計效果（Design Effect, DE）來調整樣本量，即調整後的樣本量 = 常規樣本量 * [1 +（平均群集大小 - 1）* ICC]。

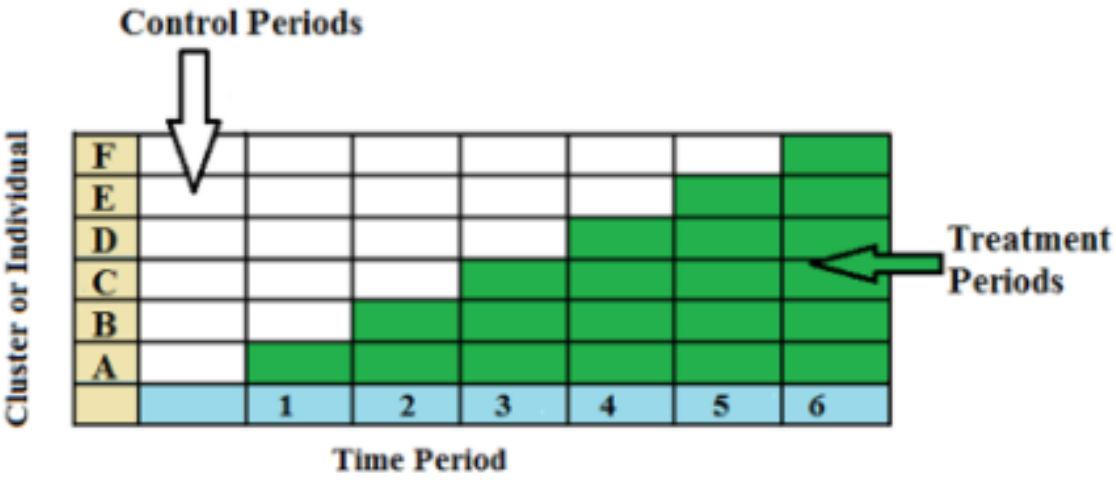
可以設計期中分析用於評估試驗的進展及安全性，可以提前停止無效或有害的試驗。預定分析時間點設置在試驗進行到一半或三分之二時。設定基於效益、中立性或安全性的提前停止規則。常用的統計方法包括 O'Brien-Fleming 或 Pocock 法[11]。

主要終點分析要確定 AI 診斷工具的主要評估指標，如診斷準確率、敏感性、特異性等。使用混合效應模型或廣義估計方程（GEE）來分析群集隨機對照試驗資料，這些方法能夠考慮群內相關性，另外應該要採用意向性分析，所有隨機分配到各組的患者均應納入分析（intention-to-treat analysis），以保持隨機化的優勢並減少偏倚。



階梯楔形務實對照試驗（Step Wedge Pragmatic Controlled Trial）

階梯楔形務實對照試驗是一種設計方法，其中每個群體（如醫院或診所）在不同時間點依次轉換到介入組[12,13]。這種設計特別適合於實施新技術（如 AI 診斷工具），因為它允許所有參與群體最終都能接收到介入，同時能夠觀察其影響。



https://www.statisticshowto.com/stepped-wedge-design/

首先選擇多個醫院或診所作為群體，其次將這些群體隨機分配到不同的轉換時間點。每個時間段，都有新的群體從對照組（使用傳統診斷方法）轉換到介入組（使用 AI 診斷工具）。轉換安排應確保轉換的過程是隨機的，以減少系統性偏差。可以使用電腦生成的隨機數或其他隨機化方法。由於階梯楔形設計中介入的時間點明確，患者可能知道自己是否處於介入組或對照組。因此，完全盲法可能難以實施，但可以嘗試盡量減少患者對試驗設計的了解。盡量讓資料分析者對群體的介入狀態保持盲目，這有助於減少分析偏差。樣本量計算需要考慮到重複測量的設計特點以及群內相關性。使用適合階梯楔形設計的樣本量計算公式，如 Hussey 和 Hughes 方法。群內相關性調整：考慮群內相關性（ICC）和時間點數量，確保樣本量能夠檢測到統計顯著的效果[13]。

期中分析旨在評估試驗進展、效益和安全性，分析時間點可以在所有群體完成一半或三分之二的轉換後進行。設定基於效益、中立性或安全性的提前停止規則。可以使用如 O'Brien-Fleming 或 Pocock 方法。

確定 AI 診斷工具的主要評估指標，如診斷準確率、敏感性、特異性等。統計方法：使用混合效應模型或廣義估計方程（GEE）來分析階梯楔形設計資料，這些方法能夠考慮重複測量和群內相關性。所有隨機分配到各群體和時間點的患者均

應納入分析（intention-to-treat analysis），以保持隨機化的優勢並減少偏倚。

設計 AI 診斷工具的對照試驗需要謹慎考慮隨機化方法、盲法實施、樣本量計算、中期分析及主要終點分析[14]。通過這些步驟，可以最大程度地減少偏倚，確保試驗結果的可靠性和有效性，同時允許所有參與群體最終都能受益於新技術。

倫理委員會的考量

如果臨床 AI 群體層級介入是針對整個團體進行的，個別群體成員無法避免這種介入[15]。如果群體成員無法避免接觸這項介入，那麼拒絕同意基本上就沒有意義[15]。只要這項介入風險很小，使用同意豁免是適當的[15]。一般來說，當資料收集發生在個體層級時，必須考慮取得知情同意。

健康經濟分析

健康經濟分析是一種評估醫療技術或介入措施的成本效益的方法，旨在確定其對於資源配置的價值。這種分析有助於決策者在有限的資源下選擇最具成本效益的選項，從而改善整體健康成果，醫療經濟學家將上訴臨床試驗結果換算成品質調整生命年（Quality Adjusted Life Years，QALYs）來做醫療經濟評估[16]。品質調整生命年是一種用來衡量健康介入效果的指標，它不僅考量生命年數，還考量生活品質。對於 AI 診斷工具來說，QALYs 可以幫助我們評估這項技術對病人健康狀況的影響，並衡量它是否值得投資[16]。

QALYs 的計算方式如下：

1. **生命年數**：首先，計算病人因使用這項 AI 診斷工具而能夠增加的生命年數。
2. **生活品質調整**：其次，根據病人的生活品質來進行調整。這是因為即使病人活得更久，如果生活品質不佳，這些額外的生命年也可能不如預期。生活品質通常以 0（死亡）到 1（完全健康）的數值來表示。

將生命年數與生活品質調整數值相乘，就可以得到品質調整生命年數。公式如下：

QALYs=生存年數x生活品質指數

例如，假設一名病人使用 AI 診斷工具後，預期能多活 3 年，但這三年中的生活品質指數為 0.8，那麼這名病人因 AI 診斷工具所獲得的 QALYs 就是：

QALYs=3 年x0.8=2.4 QALYs

這表示病人因使用這項工具，獲得了 2.4 個品質調整生命年。透過這種方式，醫療經濟學家能夠評估 AI 診斷工具的效益，並與其他治療方案或健康介入進行比較，以決定是否值得廣泛採用。

健康經濟分析中常使用決策樹（Decision Tree），決策樹是一種圖形化工具，用於描述和分析不同治療方案的可能結果[16]。它以樹狀圖的形式展示各種決策和結果的機率，幫助評估每種方案的成本和效果。樹狀圖結構：包括決策節點、機率節點和終端節點。決策節點代表選擇不同治療方案的點，在我們 AI 的臨床試驗中就是代表 AI 輔助組和傳統對照組的選擇，機率節點表示各種結果的機率，終端節點則是最終結果及其成本和效益。每條路徑的成本和效益通過加權平均來計算，其中權重是各結果的機率。

蒙地卡羅模擬（Monte Carlo Simulation）

由於每一種決策的途徑，都有不同的機率，所以醫療經濟學透過蒙地卡羅模擬預測複雜系統的行為[17]。這種方法通過反覆隨機抽樣，模擬不同情境下的結果，進而估算可能的成本和效益範圍。設定模型參數的概率分佈，通過多次模擬生成大量可能結果。每次模擬都會隨機選擇參數值，然後計算結果，通過分析多次模擬的結果分佈，確定成本和效益的不確定性範圍。

增量成本效益比（ICER, Incremental Cost-Effectiveness Ratio）

ICER 是評估新技術相對於現有技術的成本效益指標。它計算每增加一單位效果所需的額外成本[17,18]。

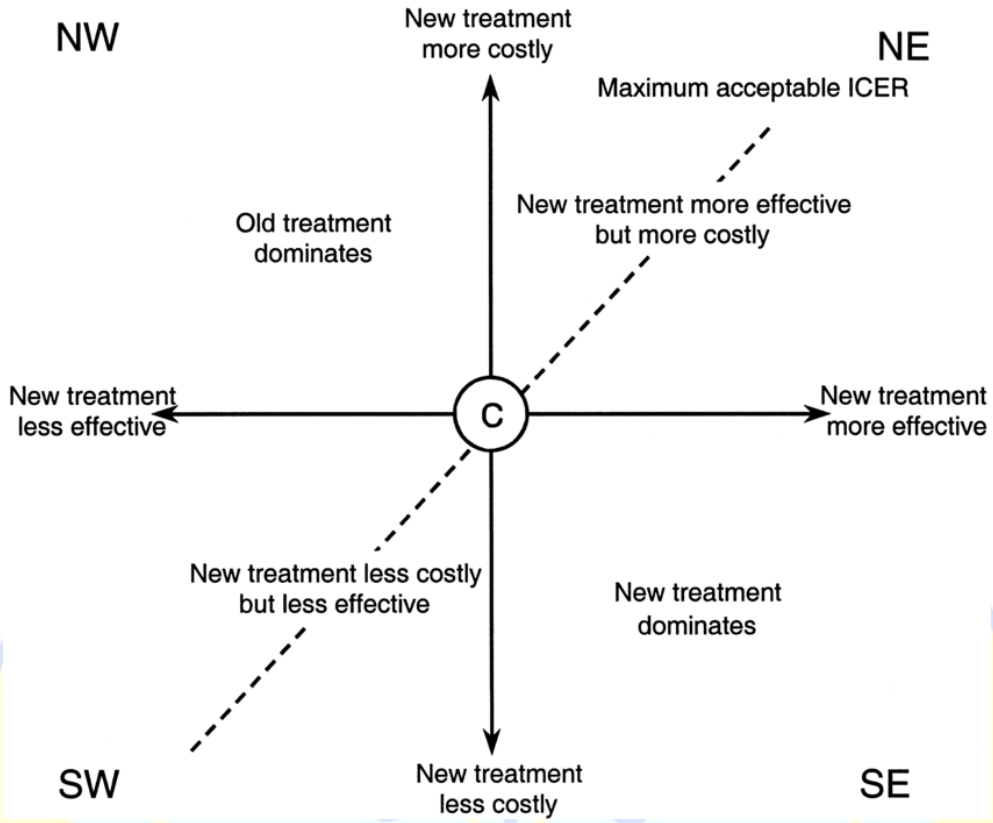
計算公式：

$$\text{ICER} = \frac{\text{新技術的總成本} - \text{現有技術的總成本}}{\text{新技術的效果} - \text{現有技術的效果}}$$

ICER 越低，表示新技術的成本效益越高。如果 ICER 小於某個預設的成本效益閾值，則該技術被認為是經濟上可接受的。根據世界衛生組織（WHO）的標準，根據世界衛生組織（WHO）的標準，一個國家是否能負擔新的 AI 技術通常取決於該技術的 ICER 相對於該國的 GDP 每人收入的比例。WHO 建議，若 ICER 低於人均 GDP 的三倍，該技術可被認為是成本效益合理的[19]。

成本效益平面分析（Cost-Effectiveness Plane）

成本效益平面分析是一種視覺化工具，用於展示不同治療方案的成本和效益[17,18]。平面圖 X 軸代表效果（如健康品質調整年數），Y 軸代表成本。根據不同治療方案的位置，可以快速判斷其相對於其他方案的成本效益。



Therapeutic Innovation and Regulatory Science 35(4)May 1998

- 象限分析：
  1. 第一象限：高成本、高效益（可能需要進一步分析是否值得）。
  2. 第二象限：低成本、高效益（理想的方案）。
  3. 第三象限：低成本、低效益（可能需要重新考慮）。
  4. 第四象限：高成本、低效益（通常不建議選擇）。

這些分析方法共同幫助評估 AI 診斷工具的經濟價值，以便在醫療資源有限的情況下做出最有效的資源配置決策。

所有申請影響性評估中心服務的頭兩年將由政府經費支持，幫助醫學中心組成跨學科、跨領域的人才團隊，建立專門的服務流程機制。在計畫結束後，希望這些中心能夠持續提供服務，形成醫院的創收單位，加速台灣智慧醫材的使用和落地。這些計畫的完成，將幫助醫學中心除了傳統的臨床服務之外，也能夠在智慧醫療價

值鏈上做出貢獻，並融入智慧醫療經濟產業中，取得收益，提升醫院的營運效率。我們希望通過這樣的模式，達成三贏局面：廠商加速產品取得健保或自費的科學證據，醫院提升診斷效率和醫療品質，政府確保健保資源有效運用，縮短城鄉差距，推動智慧醫療經濟。

圖五：政府、醫院和廠商在智慧醫療價值鏈中的三贏模式

申請資格

1. 計畫主持人與主要成員，需完成衛福部舉辦 AI 工作坊訓練課程，並取得完訓證明
2. 申請補助的醫院應為經衛生福利部醫院評鑑優等以上、醫院評鑑合格之醫學中心或區域醫院、醫院評鑑及教學醫院評鑑合格之全民健康保險特約醫院
3. 跨體系三家醫院以上組建聯盟，以執行多中心臨床試驗
4. 已具備通過 TFDA 查驗登記之人工智慧醫療產品進行試驗(加分)
5. 已有成功 AI 臨床試驗發表經驗(加分)
6. 已有成功的 AI 產品醫療經濟評估發表(加分)

團隊組成

1. 醫院管理階層副院長以上擔任計畫主持人
2. 醫療資訊室主任
3. AI 醫療資訊長（醫師背景）
4. 臨床試驗中心主任或具國際臨床試驗經驗主持人
5. 資料科學家或電腦科學家
6. 流行病學家或生物統計學家
7. 醫療經濟學家(加分)

組織架構

1. 需有專人專責與專用辦公室之 AI 影響性評估中心

具體工作

1. 建立對外公開網頁
2. 成立單一窗口，建立標準化申請電子表單
3. 建立標準化 AI 影響性評估委託研究流程
4. 建立「中央審查委員會」（CIRB，Central Institutional Review Board），統一進行倫理審查
5. 接受學界或廠商委託，依據健保諮詢結果，由委託者提供領域主持人或是由中心推薦主持人，中心方法學專家使用下列方法協助設計臨床試驗

   甲、Before-and-after trial（比較 AI 導入前後的結果來評估其影響）

   乙、Cluster randomized controlled trial（測試實際臨床環境中 AI 干預的效果）

   丙、Stepped-wedge RCT 群體隨機試驗，其中 AI 干預按隨機順序逐步引入到所有群體中
6. 中心領域專家、方法學專家、與委託者推薦專家共同制定研究終點
7. 中心提供研究設計計畫書由主持人申請倫理委員許可

8. 中心訓練臨床研究護理師協同資訊室工程師與執行臨床試驗
9. 委託者須提供符合 SMART 之應用程式，資訊室工程師將所需資料以 FHIR 串接
10. 中心進行統計分析完成臨床試報告
11. 中心提供去辨識臨床試驗資料，在倫理委員會規範下由委託者進行其他二次分析
12. 建立中心品質指標，包含各階段行政流程時效，分析方法正確性，分析報告內容完整度
13. 期末建立相關收費機制
14. 配合衛福部收集 AI 相關成果統計數字
15. 參與期末成果報告，經驗分享

申請作業要點

投標資格	計畫主持人與主要成員，需完成 AI 工作坊相關訓練課程
審查重點	計畫主持人與主要成員 AI 工作坊完訓證明 組建團隊 組建跨體系/跨層級之聯盟醫院 AI 產品影響性研究中心建置與運作規劃 已有成功 AI 臨床試驗發表經驗(加分)
第一年執行重點	建立單一服務窗口與資訊平台 建立 AI 研究特殊性臨床試驗方法論之執行能力 聯盟醫院需建立 CIRB 架構 建立專人負責與內部時效管理流程 建立組織營運成本與服務模式，以利獨立營運創收 啟動 AI 產品取證驗證案源 1 案以

第二年執行重點	完成 AI 產品影響力評估報告 1 份 持續啟動新 AI 產品落地試驗案源 1 案以上 必須應用進階 AI 產品影響力評估方法執行 完成臨床學術報告 AI 產品落地試驗有效性案例實績成果分享發表 完成組織營運創收規劃 促成驗證多中心資料庫二次利用

參考文獻

1. Wolff J, Pauling J, Keck A, Baumbach J. The economic impact of artificial intelligence in health care: systematic review. *J Med Internet Res*. 2020;22(2)
2. Lång K, Dustler M, Dahlblom V, et al. Artificial intelligence-supported screen reading versus standard double reading in mammography screening: a randomized, controlled, non-inferiority trial. *Lancet Oncol*. 2023;24(8):936-944. doi:10.1016/S1470-2045(23)00298-X.
3. Hassan AE, Ringheanu VM, Rabah R, Preston L, Qureshi AI. Early experience utilizing artificial intelligence shows significant reduction in transfer times and length of stay in a hub and spoke model. Abstract presented at the *International Stroke Conference; 2020 Feb 19-21; Los Angeles, CA.*
4. Sevilis T, Figurelle M, Avila A, Boyd C, Gao L, Heath GW, Ayub H, Devlin T. Validation of artificial intelligence to limit delays in acute stroke treatment and endovascular therapy (VALIDATE). *Stroke*. 2023;54(Suppl 1):WP81. doi:10.1161/str.54.suppl_1.WP81.

5. Han R, Acosta JN, Shakeri Z, Ioannidis JPA, Topol EJ, Rajpurkar P. Randomised controlled trials evaluating artificial intelligence in clinical practice: a scoping review. *Lancet Digit Health*. 2024;6(5).
6. National Heart, Lung, and Blood Institute. Study Quality Assessment Tools. 2013. Available at: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools. Accessed July 22, 2024.
7. O' Neill S, Kreif N, Grieve R, Sutton M, Sekhon JS. Estimating causal effects: considering three alternatives to difference-in-differences estimation. *JAMA*. 2016;316(16):1718-1730. doi:10.1001/jama.2016.10487.
8. Daw JR, Hatfield LA. Matching and regression to the mean in difference-in-differences analysis. *Health Serv Res*. 2018;53(6):4138-4156.
9. Lopez Bernal J, Cummins S, Gasparrini A. Interrupted time series regression for the evaluation of public health interventions: a tutorial. *Int J Epidemiol*. 2017;46(1):348-355. doi:10.1093/ije/dyw098.
10. Martinez-Gutierrez JC, Kim Y, Salazar-Marioni S, Tariq MB, Abdelkhaleq R, Niktabe A, Ballekere AN, Iyyangar AS, Le M, Azeem H, Miller CC, Tyson JE, Shaw S, Smith P, Cowan M, Gonzales I, McCullough LD, Barreto AD, Giancardo L, Sheth SA. Automated large vessel occlusion detection software and thrombectomy treatment times: a cluster randomized clinical trial. *JAMA Neurol*. 2023;80(11):1182-1190.
11. Oikonomou EK, Thangaraj PM, Bhatt DL, et al. An explainable machine learning-based phenomapping strategy for adaptive predictive enrichment in randomized clinical trials. *npj Digit Med*. 2023;6(1):217. doi:10.1038/s41746-023-00963-z.

12. Heagerty PJ. Developing statistical methods to improve stepped-wedge cluster randomized trials. Washington, DC: Patient-Centered Outcomes Research Institute (PCORI); 2021 Aug. Available from: https://www.ncbi.nlm.nih.gov/books/NBK604414/. Accessed July 22, 2024.
13. Hemming K, Taljaard M, McKenzie JE, Hooper R, Copas A, Thompson JA, Dixon-Woods M, Aldcroft A, Doussau A, Grayling M, Kristunas C, Goldstein CE, Campbell MK, Girling A, Eldridge S, Campbell MJ, Lilford RJ, Weijer C, Forbes AB, Grimshaw JM. Reporting of stepped wedge cluster randomized trials: extension of the CONSORT 2010 statement with explanation and elaboration. *BMJ*. 2018;363.
14. Zhou Q, Chen ZH, Cao YH, Peng S. Clinical impact and quality of randomized controlled trials involving interventions evaluating artificial intelligence prediction tools: a systematic review. NPJ Digit Med. 2021;4(1):154.
15. Nix HP, Weijer C, Brehaut JC, Forster D, Goldstein CE, Taljaard M. Informed consent in cluster randomised trials: a guide for the perplexed. *BMJ Open*. 2021;11(9). doi:10.1136/bmjopen-2021-054213.
16. Tseng AS, Thao V, Borah BJ, Attia IZ, Medina Inojosa J, Kapa S, Carter RE, Friedman PA, Lopez-Jimenez F, Yao X, Noseworthy PA. Cost effectiveness of an electrocardiographic deep learning algorithm to detect asymptomatic left ventricular dysfunction. *Mayo Clin Proc*. 2021;96(7):1835-1844. doi:10.1016/j.mayocp.2020.11.032.
17. Rossi J, Gomez J, Rojas-Perilla N, Krois J, Schwendicke F. Cost-effectiveness of artificial intelligence as a decision-support system applied to the detection and grading of melanoma, dental caries, and diabetic retinopathy. JAMA Netw Open. 2022;5(3):e220269. doi:10.1001/jamanetworkopen.2022.0269.

18. Mital S, Nguyen HV. Cost-effectiveness of using artificial intelligence versus polygenic risk score to guide breast cancer screening. *BMC Cancer*. 2022;22(1):501.
19. Thokala P, Ochalek J, Leech AA, et al. Cost-Effectiveness Thresholds: The Past, the Present and the Future. *PharmacoEconomics*. 2018;36(5):509-522. doi:10.1007/s40273-017-0606-1.