

台灣聯合學習產業大聯盟TAIFA
聯合學習工作小組專家諮詢會議 會議摘要

時間: 2021/6/28(Mon.)10:00 ~ 12:00 線上會議

主持人: 杜奕瑾 會長

報告事項

一、 聯合學習與研究倫理(報告人：臺灣大學 蔡甫昌 教授)

1. 聯合學習精神在於以模型共享取代資料共享，透過在技術突破可以解決研究倫理遭遇的一些問題，特別是醫院資料倒出所衍生資安、資料管理、擁有權等。
2. 當前AI研究以使用大數據資料進行AI輔助診斷與流程管理為主，對於大規模健康資料的收集與使用常見研究倫理問題有：
 - i. 知情同意及退出機制；
 - ii. 隱私保護相關，保護規格、去識別標準、跨越不同資料庫的串聯，如跨院同主題資料集，或串聯不同的資料來源健保、死亡、癌登、氣候、空汙等；
 - iii. 研究方法與程序:醫院的管理規範；
 - iv. 合作研究相關: 智慧財產、資料所有權、攜出問題。

這些老的問題，未來如何新瓶裝老酒，在這些既有的基礎上往前推進。在這方面可以參考英國 2018 年「數位、文化、媒體暨體育部」所頒布<數位倫理框架>，從數據的生成、分析及傳播所衍生的倫理問題進行整體性的闡述，包括: 數據倫理、演算法倫理、實作倫理。

3. 聯盟推動參考台灣臨床試驗聯合倫理審查機制(NRPB-IRB&C-IRB)。透過聯合審查可以避免重複審查與標準不一的問題，希望大家在同樣的臨床研究資訊系統PTMS基礎上進行，提高審查效率。整套機制從議定書簽訂、作業要點、聯盟擴張、說明會及共識會議都有完整流程，以及建立專案辦公室、計畫主持人、審查人之間的運作系統，讓審查時間從幾個月降到7天。建議可以借鏡聯合審查機制解決多中心合作問題，透過主審與副審模式，在各試驗聯盟讓開發完成的模型可以在不同醫院做測試，加快AI開發與使用的進程。

二、 聯合學習與個人資料隱私(報告人：中央研究院 邱文聰 主任)

1. 資料再利用過程中利害關係人主要有：資料當事人、資料原始蒐集者/再利用之資料提供者、資料利用者。在最原始模式底下是由資料蒐集者向當事人徵求同意，再回來提供給這個資料利用者，但在這之下資料利用者與資料當事人無直接關係。面對大規模資料再利用時，將會面臨當事人同意成本高與資料利用搜尋成本高的困境。
2. 現行資料治理策略常採用的方法：
 - i. 去識別化取代再同意：以資料「去識別化」，取代再利用前資料當事人之「再同意」，作為再利用之合法基礎；

- ii. 分散式資料運算：以「結果」共享，取代「資料」共享；
- iii. 建立資料管理中介者角色。

前述仍無法確保資訊自主，真正提高公眾信賴。在資料再利用上應同時考量當事人對資料處理與利用的自主權及資料處理利用過程的資料安全兩個面向。

3. 權利管理機制 (Rights Management System, RMS)：借助資訊科技，建構低成本的資訊隱私與資訊自主的權利管理機制，協助處理多邊的關係，可以降低權利行使成本、提高跨域資料搜尋效率、促成個資當事人參與的資料分享框架。主要有五個規劃原則可以參考：

- i. 資料當事人以RMS為其個資權利代理人
- ii. RMS不知資料當事人的實際數據資料內容
- iii. 資料蒐集者（原則上）不知資料當事人個資利用偏好
- iv. 資料管理者不知資料當事人身分
- v. 資料利用者只會接觸到申請利用範圍的去識別化資料

建議可以藉由此機制補足聯合學習在資料當事人角色上處理的問題。

三、 聯合學習與AI共享(報告人：臺北醫學大學 李崇僊 教授)

- 1. 在聯合學習的新資料利用模式下，有許多過去沒有討論過的層面會出現，值得進一步探討。首先需要對於資料本質進行界定(包括人格性與財產性)，並建構出資料價值創造過程，資料合作、資料利用最重要的課題在價值鏈，在過程中有貢獻的人需要得到什麼回報，決定了他們願不願意去參與貢獻，利用關係架構圖有助於設計利於資料分享流通的誘因模式。
- 2. 聯合學習的資料利用特性：
 - i. 資料之利用並不限於對資料作物理性移轉，在資料原保存地點進行AI Model訓練，仍是一種資料利用
 - ii. 聯合學習僅是降低了資料人格性面向的疑慮，仍有資料財產性面向的課題挑戰
 - iii. 聯合學習之利用成果分享問題仍是一大挑戰
 - iv. 資料的財產性並不限於機構性資料保管者才可主張，資料當事人也可能主張
 - v. 例如NIH Precision Medicine就建立起對資料當事人回饋其資料分析報告之模式，2020年12月已有27萬人參與，其關注更多資料財產性問題，當事人可因參與得到有價值資訊，未來可以從這方向來建立分享誘因。
- 3. 聯合學習的可能模式設計
 - i. 由最具代表性或最有管理經驗之機構擔任盟主，與其他資料提供機構簽約，第三方AI Model開發者為契約履行輔助人，進行AI訓練與成果回饋應用或利益分

享，相關條件由盟主與盟友們自行約定。如此聯盟可在同一合約架構下跟不同AI開發者進行合作，提高各機構維護資料品質之動機。

- ii. 除前述模式外，建議可以廣思可能的模式，再確認有利於未來利益分享與誘因動機的模式，在進一步處理後續IRB、權利分配等問題

四、 專家與談(報告人：臺北醫學大學 許明暉 數據長)

1. 兼顧RMS資料人權隱私自主保障與健康資料善加利用，借助技術發展，可以有更多創新彈性的資料運用模式，這是很好的開始。
2. 以「國際觀察性健康數據科學與資訊學聯盟」(OHDSI)機制為例，強調common data model建立的重要性，同時在符合各場域的規範下，說明資料共享模式的實際運作方式，可當作未來大聯盟建立資料共享機制的參考範例。OHDSI精神在於「one study- one script」以及「result sharing instead of data sharing」，與大聯盟推動精神一致，希望未來國際已經有類似的標準或是平台，值得借鏡和應用。
3. 美國NIH透過PCORNET機制，開始鼓勵知名大學以區域為基礎共享區域的健康資料，也開始在討論IRB與opt-out等機制建立議題，值得持續關注，台灣在這部分可以有更多國際角色扮演。

五、 主持人回饋

1. 資料治理目前是在全世界很重要探討的議題，我們很高興聯合學習產業大聯盟可以說是初試啼聲，大家有一些共同的想法，我們也希望在我們大聯盟成立之後，我們在這裡定期開會討論，然後我們把我們在台灣的遊戲規則把它訂出來，也希望有需要立法的需求，或者是每個醫院IRB裡面內部可以去介接的這個部分，大家一同來推動。
2. 蔡老師剛才講到的就是一個從問題C-IRB request為導向去找解決方法，這個機制其實他可以確保資料的使用跟落地是在他的intending use，不會因為用A的名義資料收集之後，在B的地方去使用，那目前大部分的科技公司的資料收集方式，如何像剛才說的intending use和實際上的使用是不符合的問題，那TCTC的辦公室是未來會跟AILabs這邊，我們一起來想看看聯盟的部分，我們有一套機制可以確保資料的應用跟儲存使用，是在醫院當初規劃的範圍裡面。
3. RMS的思考有一個很重要的觀念是把個資的權利管理跟資料管理分開，資料管理可能是醫院也可能是國網中心，但是當資料在使用的時候，資料的擁有者怎麼去授權資料得以再被利用，這個其實可以利用比較聰明的系統的方式來解決。
4. 其實資料不是只有我們講的人權、個資隱私，其實還有人格跟財產的面向，所以說財產的面向要怎麼有經濟上或公益上incentive，但個人或機構是可以加入聯合學習，保障個人機構合理的分潤，其實我們在聯盟的設計跟制度上面我們就要思考到這個面向，這個也會列入到我們未來的討論。

綜合討論

略

結論

1. 建立跨機構聯合審查機制，借鏡TCTC、NRPB-IRB及c-IRB經驗，同時納入模型共享模式下隱私風險評估，讓聯合學習研究發起更簡易更可行。
2. 權利管理系統（Rights Management System, RMS），強化資料利害關係人溝通，建立可信賴的資料利用模式與AI成果公益性回饋。
3. 接軌國際建立共通資料模型（Common Data Model），以AI開發需求導向建立高品質可互通的資料集(one project one script)，參與國際聯合學習。
4. 研擬聯合學習回饋誘因機制，建立可行的合作架構與商業模式，持續建置高品質資料集與開發助益於人類社會之AI模型。
5. 借助創新科技，引領國際資料使用、隱私保障與AI倫理作法，如:區塊鏈、智能合約 (smart contract)、非同質化代幣(NFT)、資料品質管理、差分隱私技術 (differential privacy)、非平衡資料處理 (data imbalance) 。