醫療 AI：在資料利用與隱私保護之間起舞

Medical AI: A Dance between Data Utilization and Privacy Protection

王　玥

Wang Yue

Abstract

At present, the development of AI depends on three core elements: high-quality data, accurate algorithms and sufficient computing power. New technologies represented by big data, cloud computing and AI are exerting a significant impact on traditional data protection. Individuals' control over their personal data is weakening, data protection is becoming more difficult, and traditional measures of privacy protection are at risk of failure. These are the most representative problems in the conflict between the development of new technology and privacy protection. A new legal and ethical

王　玥，西安交通大學法學院副教授，中國西安，郵編：710049。
Wang Yue, Associate Professor, School of Law, Xi'an Jiaotong University, Xi'an, China,710049.

《中外醫學哲學》XVII:2 (2019 年) ：頁 117-121。
International Journal of Chinese & Comparative Philosophy of Medicine XVII:2 (2019), pp. 117-121.

framework that values humans' physical safety, health and dignity should be established and deeply integrated into the entire life cycle of the design, production and application of medical AI. Based on this premise, effort should be made to promote the development of medical AI for the benefit of mankind.

人工智能並不是一個新鮮的概念。自從 1956 年美國達特茅斯學院的夏季研討會上首次提出後，半個多世紀以來，人工智能發展經歷了多次起伏。現階段，人工智能有兩條基本的技術路徑，一條是人腦智慧，對人腦的認知與運行來進行模擬，進而實現智慧；另一條，則是掀起現階段人工智能熱潮的資料智慧，其發展依賴於三個核心要素：高品質的資料、精準的演算法和充足的算力。醫療領域是目前人工智能應用的熱點領域。Robert Sparrow 對於人工智能應用在醫療領域的美好前景和潛在風險，做出了較為全面的分析。其中，他對於人工智能應用於醫療領域的隱私挑戰問題的論述，十分具有啟發性。

史派羅和哈瑟利的論文〈人工智能醫學應用的前景與風險〉對於醫療 AI 與隱私的問題，從三個方面展開了討論，闡述了醫療 AI 對隱私風險的加重、分析了技術手段保護隱私的乏力，以及隱私與潛在醫療福利之間的權衡。這幾個問題正好是目前新技術發展與隱私保護衝突中最具代表性的問題。目前，以大資料、雲計算、人工智能為代表的新技術正對傳統的資料保護造成重大衝擊：

第一，資料主體對其個人資料的控制力不斷被削弱。新技術改變了處理、存儲和訪問個人醫療健康資料的方法。現在，僅依靠一份唾液或血液的生物樣本可以揭示人體的整個基因組序列，通過創建一個基因組序列的電子記錄而無需實際長久保存一份生物樣本，然後將其存儲在“雲”中，生成一個個人基因的永久記錄，並且可以從多點進行訪問和無限複製。與紙質記錄或物理樣本可以被刪除、粉碎或銷毀不同的是，個人醫療健康資料的電子序列創建後，一旦被共用，便很難再對其進行控制和支配。在某些情況下，協力廠商還可以保存和處理這些資料，從而創建額外

的訪問點。即使收集資料的實體從伺服器上刪除了主記錄，其他訪問者也可能已經下載並共用了副本，從而創建了一個難以跟蹤且幾乎不可能完全收回的資訊網路，資料的跨境流動則更加劇了這一問題的嚴重性。

第二，資料保護的難度較之於傳統時代已經空前增大。數位時代的一切表徵都可以泛化為資料，對資料的有效收集、分析和使用構成了社會運行的基本範式。移動互聯網、大資料、雲計算、人工智能等技術應用模糊了傳統物理網路的邊界，“海量資料+複雜結構”的資料模型對隱私提出了新的保障要求。由醫療 AI 技術利用需要而產生的資料聚集和膨脹效應，使資料類型和資料結構由簡單的資料集發展為海量資料，再發展為包含複雜資料結構的大資料，通過資料收集、分析與挖掘，資料的控制者和處理者表現出的資料整合與控制力已經遠超以往，這對隱私保護帶來了空前挑戰。此外，隨著資訊的數位化程度越來越高，個人資料的收集、存儲、處理、傳輸都依賴資訊系統和網路完成。但是，這些技術的使用也與新的風險相生相伴，網路安全風險也已經成為隱私保護的又一重大挑戰。

第三，傳統的隱私保護手段面臨失效風險。此外，考慮到電腦性能、演算法能力和資料的可用性在未來將繼續增長，傳統隱私保護中最重要的手段——匿名化，應用在大量個人醫療資訊的可行性將面臨失效。在過去，研究人員認為只要不將個人醫療資料與任何能夠可直接識別的個體的資訊（例如姓名、出生日期、唯一識別碼等）儲存在一起，那麼這些資料在本質上就可以被認為是匿名的。一旦人工智能技術普遍使用，可能會增大了個人醫療資料與外部可用的公共資料匹配的可能性更大，從表面上匿名的個人醫療資訊中重新識別出個體的可能性增加了。因此，資料分析技術使個人醫療資料的匿名性很難再得到保證，隱私保護比以往任何時候都更加脆弱。更值得注意的是，傳統上認為的可以完全匿名化個人資料，現在可以通過重新識別大型匿名資料庫中

的資訊來確定特定個體，而即使是最好的資料加密技術也容易受到攻擊。國外的研究人員已經做了幾項試驗，證明有可能在大型基因研究中重新識別研究參與者：一項研究表明已經能夠通過姓氏推斷識別個人基因組，公開訪問的網路資源，以極高的概率追溯基因測序專案中參與者的身份；而另一研究發現，隱私洩露技術可以通過交叉引用兩條或更多條資訊來獲取有關個人或其家人的新的且可能有害的資料，身份追蹤技術可以利用 DNA 資料或中繼資料中的准識別字來揭示未知資料集的身份。這意味著，新技術條件下，即使對個人醫療資訊像一般資料一樣作匿名化處理是幾乎不可能完成的。

鑒於本輪的人工智能浪潮本身就是由資料和演算法驅動的，資料是人工智能的基礎，是驅動人工智能創新發展的關鍵資源。人工智能提供的服務越智慧，就越依賴高品質資料對於演算法的訓練和檢驗，但同時也意味著對隱私侵害的風險就越大。鑒於人工智能的美好前景，目前的核心問題是：在促進釋放資料潛能的同時，如何最大化地減輕資料洩露和濫用帶來的隱私侵害的種種後果？因此，不論是美國通過 Facebook 事件來深刻反思其較為寬鬆的資料治理的基本理念，還是歐盟在通過《通用資料保護條例》空前加強資料保護力度的同時，探索促進非個人資料共用，都是摸索如何在資料利用和隱私保護之間取得最大的優化平衡，去釋放人工智能時代資料的無限潛能。因此，正如史派羅和哈瑟利在文章最後指出的，為了應對醫療 AI 的潛在風險，需要來自多個相關領域的專家進行綜合研究和共同努力。

而從這個角度來看，史派羅和哈瑟利的文章對於現在中國醫療 AI 發展具有很好的啟示意義。目前中國 AI 技術領域發展火熱，關於 AI 發展中衍生的法律和倫理問題也正在逐步引起關注，醫療作為人類生命和健康守衛的最終環節，在醫療 AI 發展和應用過程，應當特別關注以下要素：第一，保障人的生命安全和健康應當成為醫療 AI 發展過程當中首要的考慮因素，對於醫療 AI 可能

帶來的對人生命和健康的潛在風險應當作出最大程度的消減。第二，隱私等人類尊嚴的價值，應當在醫療 AI 的設計、生產、應用等全生命週期當中得到充分的考慮。第三，在保障了人類的生命健康和尊嚴價值的基礎上，應當著力推動和促進醫療 AI 的發展，造福人類的福祉。

參考文獻 References

羅伯特·史派羅、約書亞·哈瑟利：〈人工智能醫學應用的前景與風險〉，《中外醫學哲學》，2019 年，第 XVII 卷，第 2 期，頁 79-109。Sparrow, Robert and Joshua Hatherley. “The Promise and Perils of AI in Medicine,” *International Journal of Chinese & Comparative Philosophy of Medicine* XVII:2 (2019), pp.79-109.