

## 世界衛生組織關於人工智慧醫療的倫理與治理指引：AI 醫療的關鍵倫理原則與其對台灣的啟示\*

A Study of the WHO Guidance on the Ethics and Governance of Artificial Intelligence for Health: Key Ethical Principles of AI for Health and Their Enlightenment to Taiwan

劉宏恩 Hung-En Liu\*

### 摘要

人工智慧（AI）在醫療領域的快速發展，普遍被認為將會對醫學與臨床醫療照護帶來許多改變。但 AI 醫療也引發了傷害風險、偏見、可能違反平等、不透明、難以究責等倫理、法律和社會方面的許多爭議及問題。如何使它既能夠促進公共利益，又能夠減輕潛在危害和意外後果，並取得社會信賴與支持使用，便成為 AI 醫療發展時無論是業者、醫療人員或是政府主管機關都需要認真加以考慮並試圖解決的挑戰。針對這些問題，世界衛生組織（WHO）發布的「人工智慧醫療倫理與治理指引」是目前國際上關於 AI 醫療的倫理與治理，最具完整性也是最具權威性的指導文件。WHO 指出：要充分發揮 AI 醫療的益處，就必須正面直視並設法解決它的倫理挑戰。倫理考量和人權思考並不是在 AI 醫療技術先發展出來之後，才需要考量的事項，而是一開始就應該把它視為 AI 醫療的設計、開發和部署使用的核心問題。本文深入分析並舉例評釋 WHO 本指引所提出的，AI 醫療相關人員皆應遵循、政府主管機關於治理時應考量的六項關鍵倫理指導原則。本文並指出：我國政府於推動 AI 醫療時似乎沒有足夠重視 WHO 此一指引，並至今尚未參考其指導原則訂定我國的 AI 醫療倫理與治理指引，這可能使台灣的 AI 醫療產業發展仍缺乏指導性的遊戲規則而平添相關業者與人員的不確定性，並有國際接軌問題，實應儘速訂定。

\* 本文為國科會專題研究計畫 113-2410-H-004-062-MY2 之部分研究成果。

\* 政治大學生醫倫理與醫事法律研究中心主任（Director, Center for Bioethics and Medical Law, ChengChi University），政治大學法律科際整合研究所副教授（Associate Professor, Institute of Law and Interdisciplinary Studies, ChengChi University）。

## Abstract

The rapid development of artificial intelligence (AI) in the medical field is generally believed to bring many changes to medicine and clinical care. However, AI in medicine has also raised many ethical, legal and social issues and problems, such as the risk of harm, bias, possible violations of equality, lack of transparency, and difficulties in accountability. How to enable it to promote the public interest while mitigating potential hazards and unintended consequences, and gain the trust and support of society to use it, has become a challenge that needs to be seriously considered and attempted to be solved by both industry and medical personnel or government authorities in the development of AI for health. In response to these issues, the “Guidance on the Ethics and Governance of Artificial Intelligence for Health” issued by the World Health Organization (WHO) is currently the most complete and authoritative international guidance document on the ethics and governance of AI in healthcare. The WHO points out that ethical considerations and human rights thinking are not matters that need to be considered after AI for health is first developed, but should be considered from the beginning as core issues in the design, development and deployment of AI for health. This article provides an in-depth analysis and examples to explain the six key ethical guiding principles set out in the WHO guidance that should be followed by those involved in AI for health and considered by government authorities in governance. This article also points out that: it seems that the Taiwanese government has not attached sufficient importance to this WHO guidance when promoting AI in healthcare, and has not yet formulated its own ethical and governance guidance for AI in healthcare with reference to the WHO’s guiding principles. This may result in the development of Taiwan’s AI medical industry still lacking guiding rules of the game, which will add uncertainty for related businesses and personnel, and may cause problems with international integration. A Taiwanese guidance on the ethics and governance of AI for health should be formulated as soon as possible.

**關鍵詞：**人工智慧、世界衛生組織、指引、倫理、治理、偏見、公平、風險、透明、可解釋性、信賴、責任、課責、永續性

**Key Words:** Artificial Intelligence, World Health Organization, Guidance, Ethics, Governance, Bias, Equity, Risk, Transparency, Explainability, Trust, Responsibility, Accountability, Sustainability

## 壹、緒論：AI 醫療的蓬勃發展伴隨全球性的倫理與治理問題

人工智慧 (AI) 在醫療保健領域的快速發展和日益廣泛的應用，普遍被認為將會對醫學與臨床醫療照護帶來許多改變，從疾病預測、診斷到提供治療建議，以及公共衛生監測與疫情控制，都可看到 AI 技術的實際應用與進一步發展的潛力。人工智慧醫療被期待將可大幅提高醫療服務的效率、準確性和醫療服務的可及性。而在產業規模及市場方面，預計全球醫療保健領域的人工智慧市場將從 2023 年的 224 億美元增長到 2030 年的超過 2,082 億美元，從 2024 年到 2030 年的年均複合增長率將達 37.5%<sup>1</sup>。

但人工智慧在醫療領域的應用，也引發了倫理、法律和社會方面的許多爭議及重大問題。例如：人工智慧醫療領域的重要先驅產品之一，IBM 公司的「華生腫瘤學系統」(Watson for Oncology) 使用 AI 演算法來評估病患病歷中的訊息並建議醫師治療選項，但於 2018 年時卻被發現：該公司的專家於內部文件中承認該系統經常給出「不安全、不正確」的癌症治療建議，很可能是由於該系統的訓練資料有所偏誤<sup>2</sup>。又例如：研究發現美國許多醫院普遍使用的一個 AI 演算法，該演算法用來預測哪些病患可能需要額外的醫療照護，但由於其測量偏差 (measurement bias)，其演算結果出現系統性的偏誤，使得少數族裔經常被認為毋需如同白人病患同樣的額外醫療照護，造成系統性的歧視<sup>3</sup>。此外，COVID-19 新冠肺炎疫情期間，世界各地許多國家及醫療系統迅速採用人工智慧工具進行接觸者追蹤、疾病監測，甚至評估哪些病人應得到優先治療，或呼吸器與疫苗不足時的優先分配計算，但它們未必充分考慮到隱私保護或其對於平等權的影響，造成社會大眾難以信任以及許多人權與法律爭議<sup>4</sup>。

另外一個例子是關於將 AI 技術運用於心理健康或心理治療領域的問題。2023 年的一篇研究顯示：數據資料處理不當，可能會導致偏差、對結果的不準確解釋以及對 AI 系統性能的過度樂觀。該研究發現，此領域的 AI 應用對於統計資料的處理方法經常存在重大瑕疵，對於資料的驗證不足，關於偏差的風險的評估也做得相當少。而且由於 AI 模型的不透明，也破壞了它的可重現性 (replicability) 與可驗證性。上述問題影響了 AI 應用的安全性以及實際上的有

---

<sup>1</sup> GRAND VIEW RESEARCH, GLOBAL AI IN HEALTHCARE MARKET SIZE & OUTLOOK 1-3 (2024).

<sup>2</sup> Casey Ross & Ike Swetlitz, *IBM's Watson Supercomputer Recommended 'Unsafe and Incorrect' Cancer Treatments, Internal Documents Show*, STAT, July 25, 2018, <https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments> (last visited Jan. 25, 2025).

<sup>3</sup> Ziad Obermeyer et al., *Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations*, 366 SCIENCE 447 (2019).

<sup>4</sup> Carmel Williams, *AI's Role in COVID-19 Poses Human Rights Risks Beyond Privacy*, University of Essex Blogs, Feb. 02, 2022, <https://www.essex.ac.uk/blog/posts/2022/02/04/ais-role-in-covid19-poses-human-rights-risks-beyond-privacy> (last visited Jan. 25, 2025). 另可參考：人工智慧與法律規範學術研究群，以人工智慧輔助 COVID-19 防治之法律議題初探，臺灣人工智慧行動網，2020 年 7 月 16 日，<https://ai.iias.sinica.edu.tw/ai-assisting-covid-19-prevention/> (瀏覽日期：2025 年 1 月 25 日)。

效性，該研究指出：新的 AI 模型往往在還沒有完成評估其在現實世界中是否可行與能否可靠應用的情況下，就被提早過度加速了對於它們的推廣與使用<sup>5</sup>。此外，近年來極為熱門的生成式 AI 聊天機器人，其對於使用民眾的心理健康或精神疾病症狀的影響，也有嚴肅的風險問題。例如，2025 年 2 月知名的科技媒體報導了相當多位民眾的實際經驗：在他們與 AI 聊天機器人對話、表述自己的心情苦悶及生活不如意後，AI 建議他們可以去自殺，甚至詳細提供了自殺方法的可能選擇，以及具體實踐步驟<sup>6</sup>。

誠如以上所述，將人工智慧應用於醫療領域一方面有相當大的現實應用價值與潛力，但另一方面它面臨一些已知或未知的風險，以及倫理、法律與社會疑慮<sup>7</sup>，因此，如何平衡這兩者之間的可能衝突，使它既能夠促進公共利益，又能夠減輕潛在危害和意外後果，並取得社會信賴與支持使用，便成為 AI 醫療產品的研發設計者及廠商、部署採用者（如醫療機構）、醫事人員、病人或大眾，以及尤其政府主管機關都需要認真加以考慮並試圖解決的挑戰。有哪些議題與挑戰是原本的相關規範及傳統醫學倫理原則不足以完全因應，需要有額外的規範架構，甚至需要全新的治理理念，而傳統的醫學倫理原則要如何含括、延伸甚至修正，以處理 AI 醫療帶來的新的或程度有異以往的醫學理論與臨床實務問題，是亟需加以釐清並論證的重要問題。考量到 AI 醫療技術與產品經常跨國使用，甚至發展之始便希望以全球為推廣市場，而且開發 AI 技術與產品的公司通常是高資本密集、高技術密集的跨國大型公司（如 Google、Meta、IBM、OpenAI），上述倫理、法律與社會議題其實是世界各國政府與社會同時都要面對的挑戰，需要有全球性觀點與可能協調一致的治理觀點與原則。

世界衛生組織（World Health Organization, WHO）是國際上最大的政府間公共衛生與健康政策組織，為聯合國下的專門機構之一，COVID-19 新冠肺炎疫情更加證明了它在跨國醫療問題與政策上面的重要性。即使它經常受到批評質疑，但即使是批評它的人恐怕也難以否認：它的確是目前全世界最具權威性與重要性的國際醫療政策與公衛治理機構。WHO 對於人工智慧用於醫療領域是否有公布任何國際性的倫理治理或規範架構的建議與指引，無疑應該是前述議題與挑戰最具參考性的重要指標之一。我國政府及民間近三十年來，不論執政黨為何，皆曾將台灣參與 WHO 列為重要的外交與衛生政策，積極推動不遺餘力<sup>8</sup>。然而推

---

<sup>5</sup> Roberto Tornero-Costa et al., *Methodological and Quality Flaws in the Use of Artificial Intelligence in Mental Health Research: Systematic Review*, 10 JMIR MENTAL HEALTH 42045 (2023).

<sup>6</sup> Eileen Guo, *An AI Chatbot Told a User How to Kill Himself—But the Company Doesn't Want to "Censor" It*, MIT TECHNOLOGY REVIEW, Feb. 06, 2025, <https://www.technologyreview.com/2025/02/06/1111077/nomi-ai-chatbot-told-user-to-kill-himself/> (last visited Feb. 7, 2025).

<sup>7</sup> 蔡甫昌、胡嘉輝，人工智慧醫療時代的倫理思維，月旦醫事法報告，58 期，2021 年 8 月，7-19 頁

<sup>8</sup> 林正義、林文程，台灣加入世界衛生組織的策略，<http://www.taiwanncf.org.tw/tforum/18/18-02.pdf>（瀏覽日期：2025 年 1 月 25 日）；衛生福利部國際合作組，WHO 簡介，<https://dep.mohw.gov.tw/OOIC/fp-7151-61301-119.html>（瀏覽日

動台灣加入 WHO 不能僅將之視為一種國際政治或外交上的訴求，而忽略了重要的實質內涵與價值：我國政府是否真的深入研究並依照 WHO 的重要建議與指引來擬定相關政策與規範，是否推動符合 WHO 希望各國可以具協調一致性的倫理與治理準則。不過似乎有些遺憾，雖然我國政府已將積極推動人工智慧醫療列為國家重要衛生政策<sup>9</sup>，但是在 WHO 自 2021 年起便已陸續公布數份具有前瞻性，且兼顧理論與應用性、相當完整並縝密的人工智慧醫療倫理與治理指引的情形下，至今沒有看到我國政府在對之加以深入研析之後，參照 WHO 的 AI 醫療國際倫理與治理指引，訂定屬於我國的較為完整且明晰的 AI 醫療倫理指引與治理架構。這對於台灣的 AI 醫療研發設計者及廠商、部署採用機構、醫事人員、病人或大眾來說，一方面可能落入某種不確定性或遊戲規則不明的狀態中，另一方面對於相關產業與產品可能未來會有國際接軌的議題，實在必須認真並審慎地加以考量。

在世界衛生組織已公布的 AI 醫療倫理與治理指引當中，最詳盡且重要的基礎文件，是 2021 年 6 月公布的「世界衛生組織人工智慧醫療倫理與治理指引」（Ethics and Governance of Artificial Intelligence for Health: WHO Guidance）<sup>10</sup>。該國際指引長達 160 餘頁，不但十分完整詳盡，而且相當具有前瞻性，例如當時生成式 AI 尚未掀起全球 AI 浪潮與各界矚目，這份指引便已提出相關可能的倫理與治理問題。但或許也是因為其詳密程度與份量長度，台灣、甚至中文文獻至今似乎除了對之加以簡介的文章，還沒有足夠完整詳盡的評析文獻。本文因此不揣淺陋，希望針對這份在 AI 醫療領域極重要的國際倫理與治理指引做較為深入並完整的評析，從其目的、脈絡、架構到重要內容加以評釋，以提供人工智慧業者、醫療機構、醫事人員、社會大眾以及更重要的是政府主管機關，在發展與使用 AI 醫療時能有與國際接軌的倫理與治理思考，追求可負責的與可信賴的人工智慧醫療發展。

## 貳、世界衛生組織人工智慧醫療倫理與治理指引之目的、脈絡與架構

對於公布 AI 醫療倫理與治理指引的目的，WHO 首先指出：為了讓 AI 對公共衛生和醫學產生有益影響，必須將倫理考量和人權思考置於醫療 AI 技術的設計、開發和部署使用的中心。要充分發揮人工智慧的益處，就必須解決醫療衛生系統、從業人員以及醫療和公共衛生服務受益者面臨的倫理挑戰。人工智慧的機遇和挑戰，兩者是不可分割的。設計開發與部署使用 AI 醫療時，應該確保病人和社會大眾他們的權利和利益，使之不會從屬於 AI 技術公司的強大商業利益，或是從屬於政府在監視人民和社會控制方面的利益。將人工智慧納入醫療衛

---

期：2025 年 1 月 25 日)。

<sup>9</sup> 衛生福利部，衛生福利部三大 AI 中心啟動記者會，2024 年 10 月 7 日，<https://www.mohw.gov.tw/cp-16-80155-1.html> (瀏覽日期：2025 年 1 月 25 日)。

<sup>10</sup> World Health Organization, Ethics and Governance of Artificial Intelligence for Health: WHO Guidance, June 28, 2021, <https://www.who.int/publications/i/item/9789240029200> (last visited Jan. 25, 2025).

生系統的目的是用來促進「人」的自主性和尊嚴，不能反而將「人」排除在醫療決策的中心之外<sup>11</sup>。

對於人工智慧醫療可能引起了跨國的倫理、法律、商業和社會問題，WHO 指出：這已經對傳統的醫療管理及醫學倫理帶來了額外的、新穎的倫理與治理挑戰，而這些挑戰已超出了傳統監管機構和醫療衛生系統成員有辦法處理的範圍。倘若我們希望將人工智慧廣泛應用於改善人類健康、維護人類自主性並確保這些技術能夠被公平獲取，就必須正面且充分地去應對這些倫理挑戰，而且不能夠採取「技術樂觀主義」(techno-optimism) 的思維，因為這種思維可能會讓問題變得更糟<sup>12</sup>。所謂的「技術樂觀主義」是指一種相信科技可以解決各種問題的思考方式，不但認為科技可以帶來創新進步，並且認為即使科技自身所帶來的風險或問題也可以再用更多的科技來予以解決。批評這種思考方式的論者會認為：很多風險或問題，涉及人的態度或行為、人性或慣性、情境脈絡、社會文化與政經條件等因素，並不是科技都能夠事先充分預測或事中事後做控制，更遑論總是能夠解決<sup>13</sup>。關於 AI 醫療的倫理與治理問題，WHO 明顯是採取對於「技術樂觀主義」較為批判或保留的立場。

WHO 的此一倫理與治理指引歷經兩年的討論與研擬過程，由 20 位醫學、公共衛生、法律、人權、科技、倫理等領域的專家組成工作小組。經過廣泛與深入分析各項相關資料，他們指出：國際上一直都還沒有基於倫理規範和人權標準訂定的全面性的 AI 醫療指引。大多數國家亦欠缺法律或規則來規範人工智慧技術於醫療領域中的使用，即使有予以規範亦往往不充分或不夠具體。WHO 指出：一份對於開發、使用或監督此類技術的不同實體的共同觀點的倫理指引，對於建立對這些技術的信任、防範負面或慢性毀壞的影響，以及避免各國間矛盾準則的擴散至關重要。因此 WHO 認為有必要訂定一個全球視角的關於 AI 醫療設計與實行，使之可能具有協調一致性的倫理指導原則。而且訂定時應該將整個醫療衛生系統中所有利益相關者都納入考量，尤其應該包括病人，因為使用這些技術雖然可能賦予生病的人權力，但也可能使他們變得脆弱，反而獲得更少的服務和保護。WHO 指引希望標定出 AI 醫療引發的各種倫理問題，並提供相關的倫理與治理框架來解決這些問題，以促使 AI 醫療能夠在全球範圍內被負責任地使用<sup>14</sup>。

對於「人工智慧」的定義，WHO 指引參考經濟合作暨發展組織 OECD 的定義：人工智慧系統是一種基於機器的系統，能夠針對人類設定的目標，進行預測、建議、或者決策，以影響真實或虛擬的外界環境。人工智慧系統被設計為具有不同程度的自主性運作。不同類型的 AI 技術，包括機器學習應用，如模式識

---

<sup>11</sup> *Id.* at v-xi.

<sup>12</sup> *Id.* at 1.

<sup>13</sup> Peter Königs, *What is Techno-Optimism?* 35 *PHILOSOPHY & TECHNOLOGY* 1 (2022); Seyram Avle & Jean Hardy, *What Is “Techno-Optimism”? 2 Technology Scholars Explain the Ideology That Says Technology Is the Answer to Every Problem*, *THE CONVERSATION*, Apr. 24, 2024, <https://theconversation.com/what-is-techno-optimism-2-technology-scholars-explain-the-ideology-that-says-technology-is-the-answer-to-every-problem-222668> (last visited Jan. 25, 2025).

<sup>14</sup> World Health Organization, *supra* note 10, at 2-3.

別、自然語言處理、信號處理和專家系統，皆可含括在此定義中<sup>15</sup>。在這裡，本文認為值得一提的是：絕大部份目前已經在市場上應用的人工智慧醫療產品，仍然是鑑別式 AI (discriminative AI) 的產品，而非近兩年來廣受媒體與社會大眾矚目的生成式 AI (generative AI) 的產品，我們不能因為近兩年來生成式 AI 極為熱門就誤以為生成式 AI 是人工智慧技術的代表，或誤以為生成式 AI 已經在醫療領域獲得廣泛應用<sup>16</sup>。

針對 AI 醫療的現況，WHO 的專家經過調查後發現：雖然人工智慧已應用於診斷，但很少有這樣的系統在前瞻性臨床試驗 (prospective clinical trials) 中被做過評估。很少有研究提供外部驗證的結果或使用相同樣本，來比較深度學習模型和醫療專業人員的表現。另一個問題是：人工智慧的性能是否可以推廣到臨床實踐中做實施，以及在一個背景下訓練的人工智慧是否可以在不同的地理區域或背景條件下仍然準確和安全地使用，其實並不確定。許多 AI 系統開發的原型在測試中表現良好，但它們經常無法轉化至較大規模的施行、商業化或部署使用。另一個問題是電腦和資訊技術會持續不斷變化，系統會變得過時或時間久了變得不穩定，而原先開發系統的公司或團隊可能會消失，無法再維護或更新<sup>17</sup>。

WHO 指出：AI 醫療應用的增加，醫療專業工作者將不得不大幅調整他們的臨床實踐方式。雖然理論上人工智慧可以自動化許多任務，讓醫生可以空出時間聽取病患的意見，並安撫他們的恐懼和擔憂，詢問與疾病沒有直接相關的社會或家庭因素。但是醫師也將因此不得不培養自己的新能力，必須理解 AI 醫療的原理才能與病患溝通風險，解釋 AI 的建議，並與病患討論風險效益等權衡取捨，且醫師需要關注 AI 醫療技術的倫理與法律問題。WHO 並特別警示：臨床醫師不能落入「自動化偏差」(automation bias) 中，讓人工智慧取代自己的臨床上盡職診察，否則很容易會不再關注技術的實施是否真的能符合自己醫療專業及病患的需求<sup>18</sup>。這裡所說的「自動化偏差」是指一種不加省思地接受電腦或自動化設備生成的結果的傾向。自動化偏差的可能原因是對於自動化結果基於依賴或習慣而不再去思索或審查，也有可能是因為自己對於自動化系統如何生成結果不知如何或無從審查，或是自動化結果並沒有給自己選擇的可能性。AI 系統比傳統的自動化設備可能更複雜難以理解或解釋，或更令人感到神奇，因此更容易加深自

---

<sup>15</sup> *Id.* at 4.

<sup>16</sup> Anand Mahurkar, *Why Discriminative AI Will Continue to Dominate Enterprise AI Adoption in a World Flooded with Discussions on Generative AI*, FAST COMPANY, July 25, 2023, <https://www.fastcompany.com/90927119/why-discriminative-ai-will-continue-to-dominate-enterprise-ai-adoption-in-a-world-flooded-with-discussions-on-generative-ai> (last visited Jan. 25, 2025); Paisan Ruamviboonsuk, *Discriminative, Generative Artificial Intelligence, and Foundation Models in Retina Imaging*, 14 TAIWAN JOURNAL OF OPHTHALMOLOGY 473 (2024).

<sup>17</sup> World Health Organization, *supra* note 10, at 6-7. 例如：本文「壹」提及的 IBM 「華生腫瘤學系統」(Watson for Oncology)，便於 2020 年終止計畫，見 Steve Lohr, *IBM Is Selling off Watson Health to a Private Equity Firm*, NEW YORK TIMES, Jan. 21, 2022, <https://www.nytimes.com/2022/01/21/business/ibm-watson-health.html> (last visited Jan. 25, 2025).

<sup>18</sup> World Health Organization, *supra* note 10, at 7-8.

動化偏差的問題<sup>19</sup>。

除了對於醫師及其臨床實務的影響，WHO 觀察到 AI 醫療將可能大幅改變病患於醫療照護中的角色，以及其接受醫療照護的場域、方式與權益。人工智慧可能改變病患自我管理自身健康與疾病治療狀況的方式，並在慢性病的情形尤為明顯。雖然轉向以病患自己為中心的照護，可能被認為是一種對於病患的賦權、對其有益，但它也可能讓病患承擔額外的責任而令其感到壓力，形同被轉嫁負擔，甚至可能會縮限某些人原本獲取正規醫療照護服務的機會。而且某些行動裝置的「健康應用程式」(Apps) 雖然涉及病患的健康與醫療照護，但是卻處於灰色地帶主張自己並非醫療器材軟體因此不須受到衛生主管機關較為嚴格的監管審查，WHO 認為這可能對某些病患帶來風險，尤其發展與提供這些應用程式的實體經常並非醫療機構，而是不屬於正式醫療照護系統的科技業者，這將會使這方面的擔憂更為加劇<sup>20</sup>。

尤其值得注意的是：這種病患角色的轉變，與藉由 AI 技術來自我照護醫療狀況的方式，將不再需要依賴傳統醫療體系，也可以不需要前往醫療機構場域，但它意味著可能需要對人們進行近乎持續性的監測，即使他們沒有生病、並不是「病患」。它允許對過去本應保持未知或不會蒐集的過多健康數據，透過穿戴裝置等方法進行近乎持續的監視和蒐集，形成一種不斷擴張的「生物監視」(biosurveillance)。這造成了重大的倫理和法律問題，例如這些資料可能被用於未明確取得同意的醫療或非醫療目的，或者政府或公司可能將這些資料用於非健康目的（譬如刑事司法或移民系統）。WHO 認為這一類的資料，應該與在正式臨床照護環境中蒐集的個人資料，受到相同等級的個資保護和安全性的拘束<sup>21</sup>。

在以上的背景脈絡中，WHO 認為：人工智慧醫療是否能夠超越過往尚屬狹隘的適用範圍，取決於人工智慧科學技術狀態之外的許多因素，以及醫療服務提供機構、病人和醫療專業人員對於人工智慧技術的信賴。之所以討論技術、法律、安全和倫理上面臨的挑戰問題，並非要阻礙人工智慧於醫療領域的進一步應用，反而恰恰相反是希望確保人工智慧能夠實現其巨大的潛力與前景<sup>22</sup>。在論述 WHO 訂定本指引的目的、AI 醫療的現況及問題等背景脈絡之後，本指引在架構上首先檢視現有的國際上相關規範與政策，並指出其零散、欠缺一致及不足等問題；接著 WHO 論述 AI 醫療在理念價值上應該追求與遵循的關鍵倫理原則；然後，進而討論若要實現這些倫理原則，於現實上必須面對與因應的具體挑戰；接著，WHO 論述於因應上述挑戰時，應該採取哪些較明確的方法與途徑；再接下

---

<sup>19</sup> Stefan Strauß, *Deep Automation Bias: How to Tackle a Wicked Problem of AI?* 5 BIG DATA COGN. COMPUT. 18 (2021); Rohan Khera et al., *Automation Bias and Assistive AI: Risk of Harm from AI-Driven Clinical Decision Support*, 330 JAMA 2255 (2023).

<sup>20</sup> World Health Organization, *supra* note 10, at 8. 另可參考劉宏恩，論智慧裝置上行動醫療應用程式之法律管制：以 Apple Watch 相關功能之爭議為例，月旦醫事法報告，84 期，2023 年 10 月，132-154 頁。

<sup>21</sup> World Health Organization, *supra* note 10, at 10.

<sup>22</sup> *Id.* at 16.

來，WHO 以專章討論使用 AI 醫療技術若是對病人造成傷害，其賠償責任歸屬制度的問題；最後，綜合上述討論，WHO 提出一套 AI 醫療的治理架構應含括的項目與子架構，並提出一份 AI 醫療於設計、部署及使用時應具體考量的項目清單。整份指引的架構與論述邏輯十分完整且清晰，從目的與背景脈絡、抽象理念與原則，到需要因應的具體問題以及對之可採取的具體方法，再到未來建置一套治理架構時的應有建制內容。這不是一份僅是宣示原則的指引，而是一份堪稱 AI 醫療之倫理治理的使用手冊的具體指導與建議準則。

WHO 表示本指引的主要訴求對象雖然為各國衛生主管機關，但它也適用於其他政府機構(如監管人工智慧或個資保護的部門)、使用 AI 醫療技術的機構(如醫院)或人(如醫師及病人)，以及設計或投資發展 AI 醫療技術的人員與實體。而且本指引的落實需要所有利害相關者的集體行動。包括病人、民間組織和公民社會也應該能夠要求政府和業者承擔責任，並讓他們共同參與技術和規則的設計，參與制定新的標準和方法，要求和尋求 AI 醫療技術的透明度，以滿足他們自己、社區和醫療衛生系統的需求<sup>23</sup>。

由於此一 WHO 指引的內容十分詳實深入且篇幅極長，本文以下主要針對 WHO 基於人性尊嚴及健康人權視角所提出的 AI 醫療關鍵倫理原則，來進行詳細分析並做評釋，最後並說明 WHO 建議的 AI 醫療於設計、部署及使用時應具體考量的項目清單。至於 WHO 提出的 AI 醫療於實務上面臨的具體倫理挑戰、因應這些挑戰做治理的具體方法與取徑、AI 醫療賠償責任歸屬及治理架構建制等議題，作者日後將再以專文做深入評析。

### 參、人工智慧醫療基於人性尊嚴及健康人權視角的關鍵倫理原則

WHO 認為：在討論 AI 醫療治理的具體問題、應採取的措施及架構之前，必須先提出價值理念上的重要倫理原則，做為上位概念指導 AI 醫療開發者、使用者和監管者發展、改進和監督此類技術的設計和使用。最上位的概念為：人性尊嚴和人固有的價值，是所有其他倫理原則所奠基的核心價值。WHO 從傳統醫學倫理被公認的四原則<sup>24</sup>，亦即不傷害原則(“do no harm”/nonmaleficence)、行善原則(beneficence)、正義原則(justice/fairness)及尊重自主原則(respect for persons/autonomy)出發，首先說明：由於健康對於個人福祉基於核心地位的重要性，也由於病人高度依賴醫療專業人員提供有關其診斷、預後以及可用治療或預防方案的相對優缺點的資訊，醫學倫理向來認為醫療服務提供者對於病人負有特殊的義務來促進這些倫理原則與價值；如果醫療服務提供者或醫療人員使用人工智慧系統來執行臨床照護任務，將原本承擔的臨床任務委由人工智慧系統行使，那麼我們就應要求設計開發這些人工智慧技術的程式工程師等人員也應熟悉

<sup>23</sup> *Id.* at xv.

<sup>24</sup> Tom Beauchamp & James Childress, *Principles of Biomedical Ethics: Marking Its Fortieth Anniversary*, 19 AMERICAN JOURNAL OF BIOETHICS 9 (2019)；蔡甫昌，生命倫理四原則方法，醫學教育，4 卷 2 期，2000 年 6 月，12-26 頁。

並遵守這些倫理原則。這些倫理原則應該內化到 AI 醫療的專業和技術標準中，我們必須對開發和使用此類技術的人員（而非僅醫療專業人員）提供倫理與治理的持續教育，並將倫理原則的考量納入系統的評估與外部稽核的標準<sup>25</sup>。WHO 進而提出了 AI 醫療技術應做「基於價值理念的設計」(design for values) 以及「設計時即包含倫理」(ethics by design) 的指引<sup>26</sup>。

基於 AI 醫療的性質與實際應用潛力，WHO 提出了相關人員皆應遵循、政府主管機關於治理時應考量的六項 AI 醫療關鍵倫理原則：(1) 保護人的自主性；(2) 促進人類福祉、人類安全和公共利益；(3) 確保透明度、可解釋性和可理解性；(4) 強化責任和課責制度；(5) 確保包納性和公平性；(6) 促進具回應性和可永續的人工智慧。WHO 並指出：衛生主管機關及醫療機構為醫事人員提供 AI 教育訓練時，必須包括使其瞭解 AI 技術的倫理問題以及其對於病患人權的影響，以之作為 AI 素養能力培育的一部分<sup>27</sup>。

## 一、保護人的自主性

人工智慧技術應該被設計成協助人類而非取代人類，人類應該維持保有完全的控制。應確保臨床醫師能夠推翻人工智慧系統做出的決定，而且機器的自主性應能受到限制並屬於本質上可逆的 (intrinsically reversible)。同時，本原則不僅強調要保護醫師的自主性，也強調必須保護病人的自主性，這代表一方面需要保護其隱私與祕密，不能未經其同意使用其資料，另一方面涉及病人是否有權知道自己接受了 AI 醫療技術的診察。WHO 認為：若診斷、預後和治療計畫中使用機器學習演算法，則應將之納入病人知情同意的過程，使之能理解 AI 技術扮演的角色。如果病人不同意，亦不得限制或拒絕對之提供基本醫療服務，政府或私人也不應提供額外的獎勵或誘因給同意使用 AI 醫療的病人<sup>28</sup>。

## 二、促進人類福祉、人類安全和公共利益

人工智慧醫療應遵循「不傷害」原則。在部署使用之前，應確保其滿足安全性、準確性和有效性的監管要求。於部署使用之後，廠商、開發人員和使用機構／人員有持續的義務來衡量和監測人工智慧演算法的性能，以確認其是否實際按照設計運作，並評估其是否對個別病人或群體產生任何不利影響。值得注意的是：即使 AI 醫療確實按照原先設計而運作，並能得出具準確性的結果，也並不代表其使用就不會造成傷害。倘若 AI 醫療提供了診斷或警告，但實際上醫學尚無法提供治療照護的方案，或是病患個人因經濟或資源等緣故無法獲得或負擔得起治療照護，那麼此時 AI 醫療技術的使用可能對之反而造成心理上的傷害，或

<sup>25</sup> World Health Organization, *supra* note 10, at 23-24.

<sup>26</sup> *Id.* at 68; Philip Brey & Brandt Dainow, *Ethics by Design for Artificial Intelligence*, 4 AI AND ETHICS 1265 (2024).

<sup>27</sup> World Health Organization, *supra* note 10, at 144.

<sup>28</sup> World Health Organization, *supra* note 10, at xii, 25-26.

是可能因此使其因健康狀況而遭受污名化或歧視。必須謹慎。<sup>29</sup>

### 三、確保透明度、可解釋性和可理解性

AI醫療使用的人工智慧技術應該是開發者、使用者和監管者可以理解的。一般而言，確保可理解性（intelligibility）的兩種方法是提高人工智慧技術的透明度（transparency）和可解釋性（explainability）。透明度要求在人工智慧技術於設計和部署之前記錄或公布足夠的資訊，包括技術的假設和局限性、操作協議、資料屬性（包括資料蒐集、處理和標註的方法）以及演算法模型的發展。並且，在人工智慧技術獲得使用許可後，仍應定期及時繼續記錄和公布此類信息。透明度可以提高系統的品質並保護病人和公眾的健康安全。例如系統評估人員需要透明度才能識別出錯誤，政府主管機構也必須依靠透明度才能進行適當、有效的監督管理。在可解釋性方面，對於人工智慧技術應在可能的範圍內，根據接受解釋的對象的能力來進行解釋。然而演算法的完全可解釋性將影響其準確性提高的可能性，兩者間可能需要權衡，WHO指引在之後AI醫療須面對的具體挑戰的章節，將進一步論述此問題，包括AI應解釋到什麼程度或解釋方式的討論。關於此原則，WHO並且指出：醫療機構、醫療系統和衛生主管機關應定期公布關於其如何決定採用人工智慧技術、如何定期評估技術、AI的使用情形與已知局限、AI的決策角色等相關資訊，以促進外部審視與監督<sup>30</sup>。另一方面，落實本原則，亦有助於提升醫師／醫療機構、病人及社會大眾對於AI醫療的信賴與使用意願<sup>31</sup>。

### 四、強化責任和課責制度

爲了確保醫療服務提供者可以負責任地使用人工智慧技術，對於AI醫療系統需要明確、透明地說明該系統可以執行的任務範圍，以及在什麼條件下有辦法實現預期的性能水準。AI醫療的利害相關者有責任確保它是在設定的任務範圍內，並且是在適當的條件下被使用。關於責任（responsibility），WHO認爲可以透過「人類擔保」（human warranty）加以確保，亦即：必須在人工智慧技術的開發和部署中，由病人和臨床醫師進行評估。應該在演算法的上游和下游建立人工監督點，確保演算法其機器學習發展保持在醫學上有效、可以被檢驗質疑、在倫理上負責的路徑上。值得注意的是：WHO指引特別強調應該讓病人及公眾在這過程中可以參與，而非僅由醫師及主管機關進行評估及監督，雖然其落實應該由主管機構進行驗證。而關於課責制度（accountability），WHO指出：當AI醫療造成傷害，或演算法決定對個人和群體造成不利影響時，必須確保有機制讓他們能夠進行糾問和獲得補救。其中應該包括從部署使用人工智慧技術的政府和公司獲得及時、有效的補救和補償。補救應包括賠償、協助康復、恢復原狀、必要時的制裁

---

<sup>29</sup> *Id.* at 26.

<sup>30</sup> *Id.* at 27.

<sup>31</sup> Victoria Tucci et al., *Factors Influencing Trust in Medical Artificial Intelligence for Healthcare Professionals: A Narrative Review*, JOURNAL OF MEDICAL ARTIFICIAL INTELLIGENCE, Mar. 30, 2022, <https://jmia.amegroups.org/article/view/6664/html> (last visited Jan. 25, 2025).

以及不再重複的保證。爲了避免責任分散，發生「每個人都好像有點問題但最後卻變成無人負責」的現象，WHO認爲可以採用無過失責任或集體責任的模式，對所有參與人工智慧技術開發和部署的行爲主體都要求共同負責，這可以鼓勵所有行爲主體都審慎誠信行事，將傷害降至最低。由於AI系統的複雜性及參與主體的多樣性，這是一個仍在不斷發展中的具挑戰性的問題，WHO指引在之後亦將以專章論述AI醫療的賠償與救濟制度<sup>32</sup>。

## 五、確保多元包納性和公平性

WHO於其憲章前言中揭示：「健康是基本人權，是普世價值，不因種族、宗教、政治信仰、經濟或社會情境而有所分別」<sup>33</sup>。於本AI醫療指引中，WHO也清楚闡釋此一健康人權的訴求<sup>34</sup>。基於包納性（inclusiveness）的要求，AI醫療技術的設計及部署使用應盡可能使其可以廣泛、公平的被使用和獲取，而不分年齡、性別、種族、收入或其他特徵條件。特別值得注意的是：WHO認爲這不僅是技術或設計層次的問題，而也是人員組成的問題，所以指引相關機構（例如AI公司、主管機關、醫療系統）於設計與處理AI醫療事務時應僱用來自不同背景、文化和學科的員工來開發、監測、評估與部署人工智慧，因爲人員組成欠缺多元性的結果，於設計、監管、評估和使用時就不易呈現多元視角，難以反應不同群體的經驗與需求<sup>35</sup>。WHO並指引：AI醫療技術的設計和評估應該自始積極納入未來需要使用該系統或將受其影響的人，包括醫療服務提供者和病人，而且這些參與者的背景應足夠多樣化。

而在公平性方面，WHO從全球性的視角指出：AI醫療技術不僅應在高收入國家，或是高收入環境的背景中可以使用，而且還應適應低收入和中等收入國家的設備類型、電信基礎設施和資料傳輸能力。AI醫療公司和政府應努力確保國家內部，以及國與國之間的「數位落差」（digital divide）不會因該技術而擴大。此外，必須特別注意人工智慧技術經常有偏見（bias）的問題，除了本文緒論舉例的由於測量偏差（measurement bias），致使演算結果出現系統性的偏誤與歧視的情形之外，由於訓練樣本代表性所造成的偏誤亦十分常見，這往往是因爲AI模型是在不具多元性、或不具少數群體代表性的樣本中做訓練或評估的。例如，脈搏血氧飽和度檢測儀（pulse oximeter device）測量皮膚對光的吸收，並將得到的資料通過演算法近似計算病人的真實血氧水準，但最近的研究發現：它在黑人病患身上的準確度很差，這可能是因爲它主要是在白人病患身上進行訓練的，當初訓

<sup>32</sup> World Health Organization, *supra* note 10, at 28.

<sup>33</sup> World Health Organization, Constitution, <https://www.who.int/about/governance/constitution> (last visited Jan. 25, 2025).

<sup>34</sup> 關於健康人權的概念、實踐與挑戰，可參考吳全峰，健康人權挑戰與展望簡析，月旦醫事法報告，34期，2019年8月，7-21頁。

<sup>35</sup> 可參考：陳亭瑋，人工智慧複製人類偏見 AI與性別議題間的拉扯，國科會科技大觀園，2021年2月26日，<https://scitechvista.nat.gov.tw/Article/c000003/detail?ID=386e5d7f-bd9c-406c-a3fd-f84313443fb9>（瀏覽日期：2025年1月25日）。

練資料中的黑人樣本資料不足<sup>36</sup>。這些偏誤或系統性的偏見不但涉及公平性的問題，而且可能會對病人造成危害。WHO指引要求AI醫療設計開發者、部署使用者及政府監管部門必須理解這些偏誤並不罕見，並應極力避免並識別AI系統可能的偏見，透過過度抽樣（oversampling）等方法加以緩解，不能使AI醫療技術加劇原本即已處於少數或弱勢的群體的健康風險，或使他們承受更多偏見和歧視。此外，WHO特別提及「權力」差距的問題：AI醫療技術應盡最大可能去減少醫療提供者和病人之間的權力差距，或是創建和部署AI醫療技術的公司與使用或依賴這些技術的人員之間的權力差距，即使這些差距可能是難以完全避免的。<sup>37</sup>

## 六、促進具回應性和可永續的人工智慧

回應性（responsiveness）的原則要求 AI 醫療的設計者、開發者和使用者系統性地、透明地去持續檢查與評估該人工智慧技術，以確定它是否在使用環境中充分且適當地回應並符合原本的期望和要求。當 AI 技術無效、造成風險或引發不滿時，回應的義務要求機構建立解決問題或終止使用該技術的程序。在其與永續性（sustainability）的要求共同做考量之下，AI 醫療技術只有在其能夠回應既有的醫療系統需求、被完全整合到醫療照護系統中並能夠繼續維持時，才應引入人工智慧技術，不應使 AI 醫療只是在短期熱度或短期資源之下引入，特別是在資源有限的醫療系統中，新技術若只是短期引入卻無法維持長期使用，或是日後難以得到維修或更新，將會浪費原本可以投入到其他常規、已獲驗證的醫療措施中的稀缺資源。此外，AI 醫療系統的設計應最大可能地減少其生態足跡（ecological footprints）並提高能源效率，使人工智慧的使用減少對於地球環境、生態系統和氣候的影響<sup>38</sup>。人工智慧對於環境與生態的影響可能相當巨大，例如訓練一個大語言模型相關的排放量相當於大約 300,000 公斤二氧化碳，或紐約市與北京之間 125 次飛機來回航程的排放量；而一個深度學習模型 GTP-3 的單次訓練所需能量，就相當於 126 個丹麥家庭的年消耗量，並產生相當於汽車行駛 700,000 公里的碳足跡<sup>39</sup>，十分驚人。

永續性的原則還要求政府和業者解決可預期的對於醫療工作職場的影響，包括應培訓醫療專業人員適應人工智慧的使用的能力，並因應 AI 醫療引入對於工作人員可能造成的就業損失<sup>40</sup>。最後這一點其實也是許多醫療專業人員的重要關切，而且已經有醫療專業人員提出強烈的訴求主張，甚至進行集體的抗議行動。例如美國全國護理師聯合會（National Nurses United）就在 2025 年 1 月號召全國

<sup>36</sup> Michael W. Sjoding et al., *Racial Bias in Pulse Oximetry Measurement*, 383 NEW ENGLAND JOURNAL OF MEDICINE 2477 (2020).

<sup>37</sup> World Health Organization, *supra* note 10, at 29-30.

<sup>38</sup> *Id.* at 30.

<sup>39</sup> *Id.* at 64. See also Sarah DeWeerd, *It's Time to Talk About the Carbon Footprint of Artificial Intelligence*, ANTHROPOCENE, Nov. 10, 2020, <https://www.anthropocenemagazine.org/2020/11/time-to-talk-about-carbon-footprint-artificial-intelligence> (last visited Jan. 25, 2025).

<sup>40</sup> World Health Organization, *supra* note 10, at 30

性的遊行、抗議和集會，抗議醫療機構與醫療系統爲了降低成本而引入 AI 等自動化技術，但一些 AI 醫療技術並未經過充分驗證而可能傷害病人，且破壞護理師的專業與照顧品質<sup>41</sup>。

## 肆、代結論：WHO 指引對於台灣推動 AI 醫療的重要提醒

「世界衛生組織人工智慧醫療倫理與治理指引」是目前國際上關於 AI 醫療的倫理與治理，最具完整性也是最具權威性的指導文件。WHO 雖然去年針對近來引發廣泛熱潮的生成式 AI 的醫療應用提出了補充性的建議，但仍然是以此一指引作爲其基礎，並在此指引提出的關鍵倫理原則與治理架構之下做補充論述，且其篇幅與內容亦遠不及此一指引詳細完整。此外，目前絕大多數的 AI 醫療應用並非屬於生成式 AI 的類型，而是鑑別式 AI 的類型，因此「世界衛生組織人工智慧醫療倫理與治理指引」仍然是關於 AI 醫療的倫理與治理最重要、最基礎、最詳盡的國際指引與治理架構建議，對於 AI 醫療的倫理與政策研究，不宜跳過對此指引的理解掌握而直接參閱 WHO 關於生成式 AI 的補充建議。

我國政府及醫界、科技業於積極推動 AI 醫療的同時，往往有「法規應鬆綁」、「不要有太多管理和限制」才有辦法讓此產業蓬勃發展的迷思。但是，爲了讓我國的 AI 技術及產品的發展能與國際接軌、符合國際上有共識的「負責任的及可信賴的 AI 醫療」的要求，並考量醫療領域因爲直接牽涉生命健康與公共衛生，其風險性質與管理方式向來與一般科技產業有所不同，且病患及社會大眾的信賴尤爲重要，我國政府及醫界實在應該積極重視 WHO 此一重要的國際指引，儘速研議並參考其建議特別針對醫療領域提出我國 AI 醫療發展的倫理與治理指引及管理架構，而非如同目前狀況僅公告一般性不區分領域與性質的 AI 科技研發或應用指引。例如我們亞洲的鄰國新加坡，其政府便於 WHO 公布此一指引之後，迅速於 2021 年 10 月公告其「人工智慧醫療指引」(Artificial Intelligence in Healthcare Guidelines, AIHGle)<sup>42</sup>；又例如印度政府，亦於 2023 年公告「人工智慧應用於生醫研究及醫療照護之倫理指引」(Ethical Guidelines for Application of Artificial Intelligence in Biomedical Research and Healthcare)<sup>43</sup>。我國衛生福利部至今沒有公告類似的 AI 醫療倫理指引，不僅對於病患權益可能有所影響，事實上對於醫療機構、醫事人員及技術業者而言，亦可能會由於遊戲規則不明，反而可能對其造成不確定性或採用與否時的猶豫。「官方指引」由於其「軟法」(soft

<sup>41</sup> National Nurses United, *Nurses March Nationwide in Support of Safe Staffing, Patient Protections Against A.I.*, Jan. 14, 2025, <https://www.nationalnursesunited.org/press/nurses-march-nationwide-in-support-of-safe-staffing-and-patient-protections-against-ai> (last visited Jan. 25, 2025).

<sup>42</sup> Ministry of Health, Singapore, *Artificial Intelligence in Healthcare Guidelines*, Oct. 2021, [https://isomer-user-content.by.gov.sg/3/9c0db09d-104c-48af-87c9-17e01695c67c/1-0-artificial-in-healthcare-guidelines-\(aihgle\)\\_publishedoct21.pdf](https://isomer-user-content.by.gov.sg/3/9c0db09d-104c-48af-87c9-17e01695c67c/1-0-artificial-in-healthcare-guidelines-(aihgle)_publishedoct21.pdf) (last visited Feb. 05, 2025).

<sup>43</sup> Indian Council of Medical Research, *Ethical Guidelines for Application of Artificial Intelligence in Biomedical Research and Healthcare*, 2023, <https://www.icmr.gov.in/ethical-guidelines-for-application-of-artificial-intelligence-in-biomedical-research-and-healthcare> (last visited Feb. 05, 2025).

law) 的性質，雖然沒有法律上的強制性，但描述了政府及專家對於特定主題的意見與法令解釋適用上的原則，將之予以透明化，給予業者及相關人員於法規及倫理遵循上的建議；它們一方面具有教育業者及相關人員的功能，可協助他們進行法規與倫理遵循，減輕他們於遊戲規則不明的情況下去自行摸索作法或猜測主管機關見解的成本與困擾，另一方面對於主管機關所屬職員於執法或審查時的判斷標準也有統一的作用。軟法性質的官方指引於世界各主要國家，包括台灣，向來對於醫療實務及醫藥產業發展具有相當大的實質重要性<sup>44</sup>。而且，面對快速發展中的新技術或新產業，它是一種可能兼顧「提供遊戲規則」與「避免強制規定可能阻滯發展」的作法。我國衛生福利部至今沒有公告 AI 醫療倫理與治理指引，平添醫療業者、醫事人員及科技業者的不確定性，甚且可能造成「沒倫理意識的人什麼都敢做，有倫理意識的人不確定該怎麼做」的「劣幣驅逐良幣」的結果，並可能對病患造成風險，實在令人感到著急。

WHO 此一指引清楚指出：AI 醫療的發展應同時以全球視角及在地視角加以審視，不但必須考量國與國之間的技術水準、基礎設施及政府監管能力的差異，也必須考慮一國之內不同群體與各個利害相關者的不同利益及「數位落差」的問題。此外，WHO 指引於一開始便揭示：要充分發揮 AI 醫療的益處，就必須正面直視並設法解決它的倫理挑戰。倫理考量和人權思考並不是在 AI 醫療技術先發展出來之後，才需要考量的事項，而是一開始就應該把它視為 AI 醫療的設計、開發和部署使用的核心問題，WHO 並提出了「基於價值理念的設計」(design for values) 以及「設計時即包含倫理」(ethics by design) 的指引，並強調欲取得公眾信賴、避免偏見或偏誤，AI 技術不僅在訓練資料來源上必須能夠具有不同群體的資料代表性，而且自始其設計人員及軟體工程師的背景便應具有多元性，並且應該讓倫理及個資隱私專家從設計開發到上市後監測評估的整個產品生命週期都扮演重要角色。此外，AI 醫療的設計開發者必須遵循透明性原則，這不僅有助於信任，而且藉由資料、變項及方法的揭露才有可能讓其受到真正的審查監督與評估，避免對病患造成傷害；另一方面，沒有透明度則其也難以被理解和解釋。WHO 並要求：採用 AI 醫療的醫院或醫事人員必須自己有能力理解並向病人解釋該 AI 技術的基本原理、限制和風險，若是自己不能理解或解釋，甚至根本不知道該 AI 技術的限制及風險，則根本不應該採用。

WHO 並特別批判了「誤以為所有問題，包括科技的風險問題及倫理問題，都可以用更多的科技來解決」的「技術樂觀主義」的觀點，強調許多風險與倫理問題是「人」或「情境脈絡」及其他條件的問題，而且無論如何，在臨床醫療的領域仍應保留「人」(醫師或病患)的主導決定權，AI 醫療的診斷和建議必須是「人」可以取捨和推翻的，不能讓 AI 技術取代了人的自主性。醫院及醫師必須避免「自動化偏差」，亦即對於 AI 得出的結果不做審慎檢視，或是習慣性地予以接受；而倘若醫院或醫師並沒有足夠能力檢視或挑戰 AI 得出的結果是否無誤，則根本一開始就不應該採用 AI。而且，AI 醫療應該只有在確認它比過去已驗證

---

<sup>44</sup> 劉宏恩，同註 20。

的傳統醫療措施更符合對病患的風險效益評估的情況下，才加以採用，不能「爲了 AI 而 AI」或「爲了熱門而 AI」，更不應該僅基於短期的計畫補助或經費而採用 AI，於計畫或補助結束了就無法再長期或永續發展，甚至因爲使用 AI 醫療而排擠到過去已被驗證有效且更爲便宜的傳統醫療措施的預算，對病患及公共健康造成不利。

WHO 指引所指出的這些問題及提供的指導原則，對於我國發展 AI 醫療提供了許多具體且實際的提醒。例如：本文作者於參與臨床試驗審查時，便不只一次發現有醫師在自己沒有充分理解 AI 技術的基本原理與風險的情況下，或是於並不確定自己是否有充分能力檢視出 AI 的錯誤的情況下，便希望使用 AI 醫療於病患身上<sup>45</sup>。至於有沒有醫療機構在並未確認 AI 醫療真的比過去已驗證的傳統醫療措施更符合對病患的風險效益評估的情況下，就去採用該 AI 醫療技術，是否有「爲了 AI 而 AI」的情形，也值得我們反省和檢視。然而，由於 WHO 此一指引的架構及內容十分完整詳盡，篇幅極長且論述的議題複雜多元，本文僅就 WHO 指引的目的、脈絡、架構與其提出的關鍵倫理指導原則做深入討論與評析，至於 WHO 指出的具體倫理挑戰與因應對策、應有的課責賠償制度及治理架構設計，作者還會再以專文加以評釋分析，希望能藉之填補台灣及中文文獻對此指引仍欠缺長篇深入研究的空白，並提供政府主管機關及 AI 醫療相關業者與專業人員於發展該技術與產業時做國際接軌的參考。

---

<sup>45</sup> 一個具體的例子是：當某醫師被問到生成式 AI 會不會發生幻覺（hallucination），編造出錯誤的醫療照護資訊給病患時，他竟表示「合作的廠商說只要透過微調（fine tune）就可以完全解決幻覺的問題」，這樣的回答顯示他自己似乎並沒有真正理解生成式 AI 的運作原理和限制，甚至根本不知道微調只可能減少幻覺，而不可能完全解決幻覺問題。至於有些醫院和醫師因爲 AI 技術太過複雜，往往會被廠商的說法牽著鼻子走的現象，是另外一個值得探討的問題。