人工智慧醫療應用與倫理準則

蔡甫昌、胡嘉輝
國立臺灣大學醫學院 醫學教育暨生醫倫理學科暨研究所

摘要

本文介紹人工智慧之定義與醫療研發應用之範疇，探討歐盟及我國科技部所頒佈之人工智慧倫理指引，期能提供國內AI醫療研發者於進行研究時能建立清晰的倫理思維，而一般讀者亦可據此來關心其應用並進行監督。（澄清醫護管理雜誌 2020；16（2）：4-8）

前言

近年來人工智慧（Artificial Intelligence, AI）的研發與應用突飛猛進，且有超越人類之趨勢。例如 Google DeepMind 所研發的下棋程式屢次擊敗棋王、IBM Debater 也在劍橋聯合會（Cambridge Union）辯論會中，獲得 51.22%聽眾的支持擊敗人類團隊 [1]。許多資源也投注在 AI 醫療照護研發與應用，期能提升人類的醫療與健康福祉。

然而 AI 的研發及運用將帶來各種倫理、法律、社會衝擊，例如人們質疑 AI 決策機制的黑箱問題（Black-box Problem）、AI 決策可能反映或激化人類偏見、AI 造成傷害時責任如何歸屬、AI 將取代人類造成失業潮…等。AI 倫理學（AI Ethics）因而受到學界與業界之重視，各國政府與國際組織（例如歐盟及經濟合作暨發展組織 OECD）亦投注資源研究 AI 之倫理法律社會問題，並且建立政策與指引 [2]。本文介紹 AI 之定義、AI 醫療研發之類型、探討歐盟及臺灣之 AI 倫理指引，期能提供國內 AI 醫療研發者於研究進行時能建立清晰的倫理思維。

通訊作者：蔡甫昌
通訊地址：臺北市中正區仁愛路一段一號
E-mail：fctsai@ntu.edu.tw

AI醫療運用概述

一、AI之定義

世界醫師會（World Medical Association）認為：「AI 由大量的電資方法（Computational Methods）所實現，經由這些方法所架設的系統可以執行與人類難以區分（Indistinguishable）的智能行為」[3]。然而，世界醫師會認為就算機器擁有上述能力，也無法全面取代醫師及醫病關係，提出『擴增智能』（Augmented Intelligence）的替代概念，以增進人類智能為前提，扮演輔助人類之角色 [3]。歐盟人工智慧高級專家小組（High Level Expert Group on Artificial Intelligence, HLEG）則定義：AI 泛指以模仿人類智能為導向的技術，或具備執行類似人類智能行為能力的軟體或軟硬體系統。

二、AI之醫療運用

AI 之醫療運用可簡要歸納為以下範疇 [5]：

（一）輔助診斷：基於深度學習在圖像辨識的成效顯著，又能夠快速處理大量訊息，使得 AI 在醫療影像診斷廣泛被使用。電子健康記錄（Electronic Health Records, EHRs）使得 AI 在即時性臨床診斷上更具潛力。AI 被廣泛認為可以大幅降低醫師在診斷工作上的負擔，並減少醫療錯誤 [5]。

（二）健康追蹤及個人化醫療：隨著物聯網（Internet of Things, IoT）技術之普及，智慧式穿戴裝置能即時與健康應用程式及社交平台串聯，使大量健康相關資訊得在醫療場域外被收集。例如精神科運用 AI 針對憂鬱症者之情緒開發電子追踪工具，緩解精神科人力不足之壓力 [5]。將 EHRs、通過 IoT 所收集的各式病人資訊、家庭史、DNA 等資料進行結合判讀與研發，乃當前醫療 AI 趨勢，推升個人化醫療。

（三）分流（Triage）及資源分配：AI 可融入繁忙的臨床工作流程（Clinical Workflow），將醫療資源

進行有效率而公平的分配。AI 可依據不同參數對病況進行預後預測，包括各種症狀及併發症的發生率、致死率（Mortality）、回診率（Readmission）等，有助於資源分配之考量 [5]。

（四）臨床研究及新藥開發：AI 可加速研究效率、設計及製造新藥物、預測藥物脫靶效應及毒性等。劍橋大學及曼徹斯特大學聯合開發的機器人『伊芙』（Eve），發現牙膏中可用來製造抗癌疾藥物成分而引起藥廠興趣 [5]。

（五）其他：如照護機器人、長期照護、病房監控的 AI 應用。

AI倫理指引

AI 倫理學屬於應用倫理學的領域，專注於研發（Development）、部署（Deployment）及運用（Usage）AI 時所產生的倫理議題；AI 倫理學所討論對象不只涵蓋技術本身，也探討可能參與之利害關係者（Stakeholders），如公司、研發人員、政府機關、公民社會團體、個人等應當扮演的角色及所具備的倫理敏感度 [4]。以下介紹歐盟《可信任 AI 倫理準則》（Ethics Guidelines for Trustworthy AI, 以下簡稱歐盟準則）[4] 以及我國科技部之《人工智慧科研發展指引》[6]，這些文獻闡述 AI 科研人員所應遵守之倫理原則。

一、歐盟《可信任AI倫理準則》

（一）背景及基本架構

為提升公私部門對 AI 的投資及使用、因應 AI 所帶來的社經變遷及確立一個能強化歐洲價值（Europeans Values）之倫理及法律框架，歐盟委員會（European Commission）於 2018 年 6 年成立 High-Level Expert Group on Artificial Intelligence, HLEG（以下簡稱 HLEG）。HLEG 由企業、學術及公民團體共計 52 名成員所組成，負責訂定 AI 倫理指引（AI Ethics Guidelines）及政策與投資建議（Policy and Investment Recommendations）。2019 年 4 月 8 日，HLEG 發布一般性指引（Guidance in General），確立可信任的 AI (Trustworthy AI）之願景。

歐盟準則認為，經由民主、法制及基本權利（Fundamental Rights）的支持，AI 可在此環境下持續精進及保衛民主文化，使創新及負責任的競爭（Responsible Competitiveness）得以實現 [4]。可信任的 AI 有三項組成要件（Components）[4]：

1. 合乎法律的 AI（Lawful AI）：AI 的運作須合乎所有適用的歐盟、國家及國際層級的法律與管制條例，及各個特定領域（Domain-specific）的條例。
2. 合乎倫理的 AI（Ethical AI）：法律或許無法及時跟上科技之快速發展，故 AI 須符合倫理規範（Ethical Norms）。
3. 強健的 AI（Robust AI）：即使 AI 能夠合乎倫理地運作，也不能保證不會帶來無意的傷害（Unintentional Harms）。技術上 AI 不僅應避免可預見的傷害，也應依據脈絡及環境因素來運作。

歐盟準則以「AI 系統生命週期（AI System Life Cycle）」作為 AI 倫理學的主軸，包含 AI 的發展、部署及運用 [4]。歐盟準則的前提是應用於 AI 系統生命週期的法律權利與義務都是強制性且必須被遵守的。準則提出四項「倫理原則（Ethical Principles）」及由此導出之七項「倫理要件（Ethical Requirements）」；倫理要件之下有「技術性方法（Technical Methods）」及「非技術性方法（Non-Technical Methods）」作為實踐方針，最後有「評估表（Assessment List）」做為檢核之用。

二、以人為中心進路（Human-Centric Approach）及四項倫理原則

歐盟準則主張 AI 倫理學必須以人類基本權利作為論述基礎，依據歐盟憲章（EU Charter）、歐盟協議（EU Treaties）及國際人權法律，歸納五項基本權利：（一）尊重人性尊嚴；（二）個體自由；（三）尊重民主、正義及法制；（四）平等、非歧視性及團結；（五）公民權利 [4]。人性尊嚴乃所有人類都具有之內在價值（Intrinsic Worth），不能被其他價值或事物（包括 AI）所減損、妥協或壓制。據此，尊重人性尊嚴意味著將每位人類視為道德主體（Moral Subjects）看待，讓人類享有基本權利，也尊重人類所獨享、無可剝奪之道德地位（Moral Status）及人性尊嚴 [4]。基於歐盟憲章乃具有法律約束力，伸張基本權利的 AI 也具備『合乎法律 AI』的組成要件。

以人為中心進路的 AI 系統生命週期乃是合乎上述價值，也重視與人類生態系統（Human Ecosystem）息息相關的自然環境及其他生物，現

今的人類必須採取永續性進路來保障未來世代的福祉 [4]。以人為中心進路的 AI 提出四項倫理原則：

（一）尊重人類自主原則（The Principle of Respect for Human Autonomy）：在人類與 AI 互動的過程中，人類必須能擁有完全及有效的自主。AI 應該擴增（Augment）、補充及鼓勵人類之認知、社交及文化技能；不可脅迫、欺騙、操控人類，或將人類從屬於其下。人類及 AI 所能分別從事的任務必須遵從『以人為中心進路』—人類必須被賦予有意義的選擇（Meaningful Choice）。此原則也要求 AI 之運作必須要有人類之監督（Human Oversight）。AI 也應該在各種職業環境中支持人類，以創造有意義的工作（Meaningful Work）。

（二）預防傷害原則（The Principle of Prevention of Harm）：AI 不可造成及加劇傷害，或執行會危害人類的任務。這表示 AI 技術上必須強健、抵禦惡意之使用，以保護人性尊嚴及人類身心之完整性。在 AI 系統生命週期中，應注意易受傷害族群（Vulnerable Persons or Groups）及因資訊或權力不對等所造成的不欲後果。此原則也適用於避免對自然環境和所有生物造成傷害。

（三）公平原則（The Principle of Fairness）：公平性區分為實質向度（Substantive Dimension）及程序向度（Procedural Dimension）。前者乃確保對利益與負擔之平等與公正的分配、使個人及群體免於不公平之偏差、歧視與污名，在手段與目的之間遵守比例原則（Principle of Proportionality Between Means and Ends），謹慎地在競爭之利益與目標間尋求平衡。後者要求人們得以對 AI 之決策及其操控者能提出有效的異議與糾正；決策負責者應該能夠被識別，決策過程必須能夠被解釋。

（四）可解釋性原則（The Principle of Explicability）：透明性（Transparency）意味著 AI 所具備的能力與製造目的必須經過公開的協商；AI 之決策生成機制也要能夠對直接與間接受到 AI 決策所影響的對象進行解釋。然而「黑箱問題」使這種解釋不會一直存在，而應考慮使用其他解釋機制（例如可回溯性、可稽核性及針對系統所擁有的能力進行公開透明的溝通）。在選擇採納何種解釋機制時，應視脈絡及後果之嚴重性而決定。

歐盟準則強調上述四項原則並不存在位階關係（Hierarchy），實際運用時不免會遭遇衝突或張力，準則建議三項調和張力的方法：（1）通過公開的民主參與機制，以進行負責任的審議（Accountable Deliberation）；（2）AI 所帶來長遠好處必須超越任何可能帶來的風險；（3）在缺乏倫理上可接受之代價產生時，必須堅守某些特定、絕對的基本權利，如人性尊嚴 [4]。

三、七項倫理要件及實作方法

為實現可信任的 AI，各個參與 AI 系統生命週期的利害關係者（Stakeholders），包括開發者（Developers）、部署者（Deployers）、終端使用者（End-Users）與所有可能被 AI 所直接及間接影響的人們，應被賦予不同的倫理要求。歐盟準則按照上述四項倫理原則，進一步歸納七項倫理要件，不僅制定了 AI 系統應持有之系統層面（Systemic）要求，更著眼將個人（Individual）及社會（Societal）的考量亦納入 [4]。這七項倫理要件與倫理原則之對應關係如表一。歐盟準則並不認為只有單一或若干利害關係者需要接受倫理制約，而是盡可能涵蓋所有可能的參與者；而每項倫理要件之內涵都與原則相互呼應。

表一 倫理原則與倫理要件之對應關係

倫理原則	倫理要件
1. 尊重人類自主性原則	維繫人性及監督（Human Agency and Oversight）
2. 預防傷害原則	技術強健及安全性（Technical Robustness and Safety）
	隱私及數據治理（Privacy and data Governance）
	社會及環境福祉（Societal and Environmental Well-Being）
3. 公平原則	多元、非歧視及公平性（Diversity, Non-Discrimination and Fairness）
	可問責性（Accountability）
4. 可解釋性原則	透明性（Transparency）

歐盟準則也提出五項技術性方法及七項非技術性方法以實踐可信賴的 AI。技術性方法包括：可信任的 AI 之架設（Architectures for Trustworthy AI）、倫理及法制之設計（Ethics and rule of law by Design）、解釋方法（Explanation Methods）、測試及驗證（Testing and Validation）、設立服務品質指標（Quality of Service Indicators）。非技術性方法包括：管理規範（Regulation）、行為守則（Codes of Conduct）、標準化（Standardisation）、認證制度（Certification）、管制架構確立可問責性（Accountability via Governance Frameworks）、教育及認知提升以培育倫理思維（Education and Awareness to Foster an Ethical Mind-Set）、利害關係者之參與及社會對話（Stakeholder Participation and Social Dialogue）、具多元及包容性之設計團隊（Diversity and Inclusive Design Teams）。歐盟準則強調，基於 AI 乃在動態的環境中運作，針對應滿足何種倫理原則及落實何種倫理要件，需由利害關係者們針對 AI 不同之運用及脈絡而提出。當在考慮應實施何種實作方法時，也必須做出持續性（Ongoing）的評估及證成，以顯示 AI 的系統設計、部署及使用上皆能滿足先前訂定之倫理原則及要件 [4]。

四、臺灣科技部《人工智慧科研發展指引》

本土 AI 倫理指引進展方面，我國科技部於 2019 年 9 月發布《人工智慧科研發展指引》[6]。此文件指出 AI 科研人員必須與其他利害關係者共同打造符合普世期望、促進人機合作及值得信賴的 AI 環境。為擴增 AI 之益處，並消弭 AI 可能帶來的歧視、偏見、濫用及排除的風險，此份文件提出三個 AI 所遵從之價值，輔以八項 AI 科研人員所需遵從的指引 [6]。

AI 科研人員必須與其他利害關係者密切合作及對話，以冀 AI 科研可滿足三**大價值**：

（一）以人為本：提升人類生活、增進人類福祉、及尊重人性尊嚴、自由與基本人權。

（二）永續發展：追求經濟成長、社會進步與環境保護間之利益平衡。

（三）多元包容：積極啟動跨領域對話機制，普惠全民對 AI 的理解與認知，以創建及包容多元價值觀與背景之 AI 社會。

針對實作指引方面，AI 科研人員應施行以下八項：

1. 共榮共利：致力於多元文化、社會包容、環境永續，及保障人類身心健康、創建全體人民利益、總體環境躍升之 AI 社會。
2. 公平性與非歧視性：確保 AI 之決策平等尊重所有人之基本人權與人性尊嚴，避免產生偏差與歧視等風險，並建立外部回饋機制。
3. 自主權與控制權：AI 係以輔助人類決策的工具，並能讓人類能擁有完整且有效的自主、自決與控制的權利。
4. 安全性：AI 之安全性應包括但不限於穩健性、網路與資訊安全、風險控管與監測。
5. 個人隱私與數據治理：注意個人資料蒐集、處理及利用符合相關法令規範，並針對 AI 系統內部之個人資料架構有適當的管理措施。
6. 透明性與可追溯性：AI 之運作機制需進行最低限度的資訊提供與揭露，包括但不限於對於模組、機制、組成、參數及計算等。AI 技術之發展與應用須遵循可追溯性要求，對於決策過程中包括但不限於資料收集、資料標籤以及所使用的演算法進行適當記錄，並建立相關紀錄保存制度，以利救濟及事後釐清。
7. 可解釋性：除了致力權衡決策生成之準確性與可解釋性，並應盡力以文字、視覺、範例進行事後的說明、展現與解釋。
8. 問責與溝通：問責制度應包括但不限於決策程序與結果的說明、使用者與受影響者之回饋管道。

由此可見，科技部以普世價值的刻畫為起點，輔以較明確的實作方向，期望在科研階段時盡量減少 AI 可能產生的風險；並基於科研人員擁有專業知識，應扮演各個利害關係者中介之角色，以期將這些倫理價值實踐於 AI 之研發與應用中，以增益人類福祉。

結語

AI 之醫療研發及應用已廣泛開展，相關倫理法律社會挑戰亦接踵而至，本文介紹歐盟及臺灣發佈之 AI 倫理指引，期能提供給 AI 研發人員、利害關係者、醫療機構、政策制訂者等，於面對相關倫

理議題時可參考依循，進而提升 AI 研發應用之倫理品質，保障民眾權益與社會福祉。

參考文獻

1. Moskvitch K: Augmenting humans: IBM's project debater AI gives human debating teams a hand at Cambridge. 2019. Retrieved from http://bit.ly/39aF7dT
2. Algorithm Watch: AI ethics guidelines global inventory. 2019. Retrieved from http://bit.ly/3cphq3F
3. World Medical Association: WMA statement on augmented intelligence in medical care. 2019. Retrieved from http://bit.ly/2uKrC5D
4. High-Level Expert Group on Artificial Intelligence: Ethics guidelines for trustworthy AI. 2019. Retrieved from http://bit.ly/3ahkqgv
5. Topol EJ: High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine 2019; 25(1): 44-56.
6. 科技部：人工智慧科研發展指引（108年9月版）。2019。Retrieved from http://bit.ly/39lzUAc