AI 時代學術倫理的風險與因應

吳清山
國立暨南國際大學教育政策與行政學系榮譽講座教授
臺北市立大學教育行政與評鑑研究所名譽教授

一、前言

AI（Artificial Intelligence，AI）的快速發展，改變了人類的生活、工作和溝通方式，也協助解決了組織營運、醫療診斷、交通管理、金融投資、環境保護及教育學習等實際問題，支持各行各業的廣泛應用，使其更具效率、準確性和安全性。Murphy 等人（2021）將 AI 稱為「第四次工業革命」，足見 AI 在人類社會中產生的變革性影響將會日益增強。

AI 具有驚人的深度學習、高速運算、數據處理、語言處理、文本生成和圖像辨別等能力，在某些功能上可能超越人類大腦，這與 AI 系統中的巨大記憶體、運算速度和資料的多層處理有關。未來，AI 在教育學習和學術研究中的應用亦將越來越廣泛，它為學習與研究提供了極大的助力，但也必然帶來巨大的衝擊。張芬芬（2023）指出，ChatGPT 與各式 AI 工具是必須善用的新工具，學校的教學、學習和學術研究已經開始受到它們的影響。

學術界運用 AI 工具從事各種探索性和開創性研究，將成為一個重要趨勢。然而，在學術研究中，如何避免誤用或濫用 AI 工具以致影響學術研究的正常發展，是一項涉及學術倫理和道德的問題。Trotta 等人（2023）提到：「隨著 AI 不斷發展並日益融入我們的日常生活，我們繼續研究其倫理和社會影響至關重要。」（頁 440）這同樣適用於學術研究，凸顯了 AI 學術倫理的重要性。

誠信是學術倫理的根基。The International Center for Academic Integrity（2021）將學術誠信定義為即使在逆境中也要堅守六項基本價值：誠實、信任、公平、尊重、責任和勇氣。這些價值觀產生了行為準則，使學術界能夠將理想轉化為行動。在 AI 工具應用愈來愈普遍的時代，如何持守對學術誠信的承諾，需要學術界的共同呼籲與努力。

隨著生成式 AI（Generative AI）浪潮的興起，它可能刺激研究者從事創造性的議題、數據模式建立、模擬情境、跨領域研究和個性化學習等多樣化研究，亦具有節省研究人力、時間和成本，體驗研究樂趣等各種益處；但在使用過程中，可能也存在一些潛在的倫理風險，值得加以正視。

因此，本文除了分析 AI 可能帶給學術倫理之潛在風險外，並提出因應的策略，以供參考。

二、AI 應用對學術倫理潛在的風險

AI 的快速發展涉及到倫理問題，已引起學術界、政府機關和國際組織的關切。UNESCO（2023）指出：

> AI 的快速變化也引發了深刻的倫理關切。這些關切源於 AI 系統可能內嵌的偏見、對氣候變化的助長、對人權的威脅等。與 AI 相關的這些風險已經開始加劇現有的不公平，進一步損害了已經處於邊緣化的群體。（頁 6）。

OECD（n.d.）的「AI 原則 1.1」（AI Principle 1.1）亦提到：

> 利害關係人應積極參與對值得信賴的 AI 的負責任管理，以追求對人類和地球有利的結果，例如增強人類能力和創造力，促進包容性不足的人群，減少經濟、社會、性別和其他不公平，以及保護自然環境，從而促進包容性成長、福祉、永續發展和環境永續性。

顯然 AI 應用是有其潛在的風險，必須加以正視。吳清山（2024）提到 AI 為教育發展帶來各種機會，但也帶來各種挑戰，包括內容的正確性及可信度、學生隱私及安全確保、學生過度依賴及認知風險、內容運用之倫理規範遵守、教育公平性及學習落差、教師角色及 AI 素養。其中，內容正確性、隱私及安全，過度依賴、倫理規範等均與學術倫理具有密切關係，而這些挑戰亦是學術倫理潛在風險的一部分。

AI 是一個很有用的學術研究工具，特別是生成式 AI，可用於回答特定問題、產生想法、總結文本、翻譯語言、生成全文等，對學術研究具有便利性。然而，我們只看到輸入和輸出，但卻不知其中運用方法的過程，加上生成式 AI 模型並沒有真正的對錯概念，因而隱藏著看不見的風險。Murphy 等人（2021）指出，AI 涉及與隱私、信任、問責制和責任以及偏見相關的一些常見道德問題。Kelly（2023）亦提到學術界使用生成式 AI 的幾個倫理觀點，包括偏見、公平和透明度、抄襲及誤傳等。

從以上說明中，可知 AI 的應用，的確存在各種不同的風險，在學術研究亦是如此。茲就 AI 應用對學術倫理潛在風險說明如下。

(一) 內容的正確性及可信度

生成式 AI 生成的內容雖然非常有效率，但其正確性卻常受到質疑。主要原因是使用者不了解其生成過程，且生成的內容並未標明資料或數據來源，難以判

斷其真實性和正確性。有時，生成的內容甚至會張冠李戴，導致失真的現象。

學習者或學術研究者一旦引用 AI 生成的不正確資料或數據，可能會造成學習偏差或影響學術研究品質，成為學習者或學術研究的一大風險。

(二) 內容的偏見或歧視

生成式 AI 生成的內容，是透過大量資料進行深度學習而來，倘若那些大量資料本身具有偏見或歧視，則輸出的內容必然也會有偏見或歧視；其次，設計 AI 系統模型者，本身有其既定的信念或價值觀，則輸出的內容也會帶有偏見或歧視；此外，使用者本身就有偏見或歧視，也只會擇其所需的內容。

Mukherjee 等人（2023）提到 AI 領域的所有重要概念，包括偏見、歧視、公平性和可解釋性等。這些概念也都屬於學術倫理關切的議題。而 Lucy 與 Bamman（2021），以及 UNESCO（2024）研究都發現，生成式 AI 會產生性別偏見或刻板印象。顯然，生成式 AI 造成性別、民族、種族、身障偏見或歧視是很常見（Decamp& Lindvall, 2023），甚至可能製造仇恨或對立，此將威脅到學術倫理。

(三) 個人隱私和安全

AI 系統仰賴大量數據運作、學習與生成內容，但這些數據可能包含敏感資訊。此外，AI 系統的運作過程缺乏透明度，且參與數據蒐集的使用者是否知情同意尚不明朗。一旦數據遭竊、被洩露或濫用，恐將嚴重侵害個人隱私和安全。

van Rijmenam（2023）指出，AI 可能成為侵犯隱私的工具。AI 系統需要大量（個人）數據，若落入邪惡人士之手，恐將被用於身分盜竊、網路霸凌等惡意目的。Balaban（2023）、Miller（2024）與 Solove（2024）等學者也持相同看法，認為 AI 可能侵害個人隱私和安全。

(四) 抄襲和不當使用

AI 系統具有強大的能力，能在短時間內生成論文或報告。然而，學生或研究者可能將這些生成的內容冒充為自己的作品，這種行為嚴重違反學術倫理。再者，生成式 AI 產生的內容通常不會註明來源，這可能涉及抄襲或剽竊問題，使得研究者在引用時難以區分內容的真實性，可能導致不當引用。

此外，AI 有可能操縱研究數據或結果，甚至刪除不符合預期的數據，這將

導致研究者得出錯誤的結論和建議。根據 Generative AI @ Harvard（2024）的報告，AI 具有傳播虛假或誤導性信息，或被用來欺騙或操縱個人的風險。另外，World Economic Forum（2024）發布的《全球風險報告》（Global Risks Report），提到誤導信息和虛假信息和 AI 技術的不利後果等風險，均與學術倫理密切相關。

三、AI 時代學術倫理因應的策略

當 AI 在學術研究中得到恰當運用時，它可以成為推動學術發展的一股重要動力；然而，若被濫用或誤用來從事不道德行為，則不僅帶來多種潛在風險，甚至可能危害學術的整體進展。Dolunay 等人（2024）提出了防止 AI 被不道德使用的一系列建議，這包括開設組織培訓課程以提升專業意識、建立專門針對 AI 的道德委員會，以及制定明確的 AI 使用規範等措施。茲就 AI 時代學術倫理的因應，提出下列策略，以供參考。

(一) 研訂 AI 學術倫理準則

隨著 AI 技術的快速發展，其在各個領域的應用也變得日益廣泛。在學術界，AI 已被廣泛應用於研究、教學和出版等方面。儘管現有的學術倫理準則提供了基本的道德和行為指導，但 AI 帶來的特定問題，例如數據隱私、智慧財產權、以及研究結果的可靠性和透明度，均需更具體和詳細的規範。

Gulumbe 等人（2024）指出，確保 AI 整合不會損害學術工作的誠信是迫切需要的，這要求有效地實施和執行倫理指導方針。由於 AI 學術倫理跨多個領域，因此需要教育學者、資訊科技專家、社會學者、倫理學者及法律學者共同研討並建立 AI 學術倫理應遵循的原則和方針，以提供必要的指導和規範。

(二) 建立 AI 學術倫理框架

AI 技術具有廣泛的應用前景，但也存在一定的潛在風險。例如，AI 可以被用於製造假新聞、進行深度偽造、或開發自主武器等。如果缺乏有效的倫理規範，AI 技術就有可能被濫用，造成危害。Trotta 等人（2023）指出 AI 倫理需要一個共同的原則和標準框架，以確保 AI 使用的問責制、公平性和透明度。此外，政府應制定政策和指南，確保負責任地使用 AI，同時促進創新和進步。

因此，建立一個 AI 學術倫理框架的主要功能是確保 AI 在學術研究中的使用符合道德規範和法律要求，促進研究的透明度和可靠性，並防止濫用技術導致的倫理問題。這個框架應該包含以下幾個主要內容：數據管理和隱私保護、透明度和可解釋性、責任歸屬、遵守法律和規範、公平性和反歧視、防止濫用和不當

行為，以及持續監督和評估，以確保 AI 增進學術研究和促進學術健全發展。

(三) 開設 AI 學術倫理課程

隨著 AI 技術的快速發展，其在學術研究中的應用也日益廣泛。在應用過程中，研究者難免會遇到各種挑戰和風險。因此，開設 AI 學術倫理課程是必要的，這可以幫助研究者了解 AI 技術的潛在倫理問題，培養對 AI 技術的倫理意識，並學會如何負責任地使用 AI。

為了使 AI 學術倫理課程更具實用性，建議其課程內容應包括：AI 學術倫理的基礎知識、AI 技術的潛在風險和倫理問題、如何設計和進行符合倫理的 AI 研究、AI 相關政策和法規，以及 AI 學術倫理的案例研究。這樣的課程設計應能顯著提升研究者的 AI 倫理素養。

(四) 發展生成式 AI 產生論文的使用規範

隨著生成式 AI 的能力增強，確保在使用這些工具時不侵犯學術誠信，已成為一個重要課題。因此，制定使用規範至關重要，以確保研究者在遵守規範的同時，保持負責任的態度，避免學術倫理問題的發生，從而提升學術研究的品質並促進其健康發展。

生成式 AI 可以作為輔助研究的工具，而其使用規範應旨在提供給研究者必要的審閱和回饋機制。此外，研究者產出的著作應具備原創性，使用者對於生成式 AI 產出的內容應具備辨別真實性的判斷力，且不能完全依賴或全面使用這些內容，以確保學術研究的誠信與品質。

四、結語

AI 本身是一種工具，其所造成的風險主要來自於人類的誤用或濫用，這常常帶來許多倫理問題，尤其在學術研究中尤為常見。學術研究的目的在於探索未知、擴展知識領域、解決實際問題，並促進社會與科技的進步，為人類社會創造無限的可能性。因此，堅守學術倫理是確保學術研究品質的關鍵。

隨著 AI 在學術研究中的應用日益廣泛，各種潛在風險不可避免地影響到學術研究的公信力與價值。在 AI 的時代，我們更應關注學術倫理是否需要隨著 AI 的發展進行調整，以豐富學術研究的內容並擴展其效能。

本文強調 AI 時代帶來的潛在風險，並呼籲學術界予以重視。為此，本文提

出了下列四項因應策略，包括研訂 AI 學術倫理準則、建立 AI 學術倫理框架、開設 AI 學術倫理課程、以及發展生成式 AI 產生論文的使用規範，以供參考。

參考文獻

■ 吳清山（2024）。**前瞻教育議題研究**。臺北：元照。

■ 張芬芬（2023）。活用AI讓教學展現新風貌。**聯合報**，**A10**。

■ Balaban, D. (2023). *Privacy and security issues of using AI for academic purposes*. Retrieved from https://www.forbes.com/sites/davidbalaban/2024/03/29/privacy-and-security-issues-of-using-ai-for-academic-purposes/

■ Decamp, M. & Lindvall, C. (2023). Mitigating bias in AI at the point of care: Promoting equity in AI in health care requires addressing biases at clinical implementation. *Science, 381*(6654), 150-152. doi: 10.1126/science.adh271

■ Dolunay, A. & Ahmet C. Temel, A. C. (2024). The relationship between personal and professional goals and emotional state in academia: A study on unethical use of artificial intelligence. *Front Psychol,* 15. doi: 10.3389/fpsyg.2024.1363174

■ Generative AI @ Harvard (2024). *Research with Generative AI: Resources for scholars and researchers*. Retrieved from https://harvard.edu/ai/research-resources/

■ Gulumbe, B.H., Audu, S.M. & Hashim, A.M. (2024). Balancing AI and academic integrity: What are the positions of academic publishers and universities? *AI & Soc*. doi.org/10.1007/s00146-024-01946-8

■ Kelly, N. (2023). *Artificial intelligence: Ethical considerations in academia*. Retrieved from https://blog.mdpi.com/2024/02/01/ethical-considerations-artificial-intelligence/

■ Lucy, L. & Bamman, D. (2021). *Gender and representation bias in GPT-3 generated stories*. Proceedings of the 3rd Workshop on Narrative Understanding.

■ Miller, K. (2024). *Privacy in an AI era: How do we protect our personal information? A new report analyzes the risks of AI and offers potential solutions*. Retrieved from https://hai.stanford.edu/news/privacy-ai-era-how-do-we-protect-our-p

ersonal-information

■ Mukherjee, A., Kulshrestha, J., Chakraborty, A. & Srijan Kumar, S. (eds.) (2023). *Ethics in artificial intelligence: Bias, fairness and beyond*. Berlin: Springer.

■ Murphy, K., Ruggiero, E. D., Upshur, R. Donald J. Willison, D. J., Malhotra, N., Cai1, J. C., Malhotra, N., Lui, V. & Gibson, J. (2021). Artificial intelligence for good health: A scoping review of the ethics literature. *BMC Med Ethics, 22*(14), 1-17. doi.org/10.1186/s12910-021-00577-8

■ OECD (n.d.). *Inclusive growth, sustainable development and well-being* (Principle 1.1) https://oecd.ai/en/dashboards/ai-principles/P5

■ Solove, D. J. (2024). *Artificial intelligence and privacy*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4713111

■ The International Center for Academic Integrity (2021). *The fundamental values of academic integrity*. Author.

■ Trotta, A., Ziosi, M. & Lomonaco, V. (2023). The future of ethics in AI: Challenges and opportunities. *AI & Society, 38*, 439-441. doi.org/10.1007/s00146-023-01644-x

■ UNESCO (2023). *UNESCO's recommendation on the ethics of artificial intelligence: Key facts*. Author.

■ UNESCO (2024). *Generative AI: UNESCO study reveals alarming evidence of regressive gender stereotypes*. Retrieved from https://www.unesco.org/en/articles/generative-ai-unesco-study-reveals-alarming-evidence-regressive-gender-stereotypes

■ van Rijmenam, M. (2023). *Privacy in the of AI: Risks, challenges and solutions*. Retrieved from https://www.thedigitalspeaker.com/privacy-age-ai-risks-challenges-solutions/

■ World Economic Forum (2024). *Global risks report*. Retrieved from https://www.weforum.org/agenda/2024/01/ai-disinformation-global-risks