從國際比較觀點探討生成式 AI 在中小學教學之倫理議題

張淑涵
國立臺中教育大學教育系博士生
阮孝齊
國立臺中教育大學教育系助理教授

一、前言

隨著生成式人工智慧（Generative Artificial Intelligence, GAI）技術迅速發展，並廣泛應用於教育領域，相關的倫理議題也逐漸成為全球關注的焦點。國際組織如 UNESCO、歐盟，以及我國教育部，紛紛針對 GAI 在中小學教學中的運用提出各類指引與政策，期望能在促進創新教學的同時，確保學生權益與學習環境的安全（Giannini, 2023；Holmes & Miao, 2023）。這些指引涵蓋的層面廣泛，從數據隱私、數位素養、到人工智慧如何影響師生互動與評量標準等，皆被視為不可忽略的重要課題。然而，現行的倫理原則往往偏向抽象與概念化，缺乏具體操作性，致使第一線教師在教學實踐時面臨困難，難以將原則有效落實於課堂中（Spector, 2016）。

因此，針對這一現象，本文採用比較研究的觀點，綜合分析國際間及台灣在 GAI 教育倫理方面的政策文件，進行系統性整理與歸納。透過比較國內外政策的異同與實踐挑戰，本文試圖提出適切且實務可行的建議，協助教師與教育決策者在推動 GAI 教學時能有具體依循。本文的重要性，不僅在於提供一份詳盡的國際比較資料，更在於引介具體且可落地的操作方針，協助教育現場與決策層面共同因應新興科技所帶來的倫理挑戰，進一步促進數位教學環境的優化與學生的全面發展。本文的主要目的有二，說明如下。

（一）探究聯合國教科文組織、歐盟與臺灣三方在 AI 教育應用的倫理指引中，分別強調哪些價值與實踐策略？

（二）從教師教學實踐的層面，討論臺灣現行政策如何回應上述倫理原則，尚有哪些挑戰與發展空間？

二、文獻回顧

本文首先從 Celik（2023）AIPACK（AI Pedagogical and Content Knowledge）量表中，對於倫理向度之概念進行說明，繼而說明聯合國教科文組織、歐盟以及我國數位教學指引中的數位教學倫理項目進行整理，最後進行比較。

（一）AIPACK 中的倫理向度

在教師如何運用生成式 AI 進行教學的討論中，倫理議題在教師整合科技中經常被忽略。除了少數幾項研究外，以往的 TPACK（Technological Pedagogical Content Knowledge）架構裡很少處理科技使用相關的倫理層面（Smakman et al., 2021；Yurdakul et al., 2012），生成式 AI 引發的倫理問題也尚未被充分考慮。

在討論 TPACK 架構中的測量面向時，學者 Yurdakul 等人（2012）指出倫理因素指的是在教學與學習環境中，無論是科技相關的倫理議題，還是教師專業倫理，均展現合法且合乎道德的行為，而教師應該關注科技使用中的存取權問題 、智慧財產權議題 ，並關注科技資訊的正確性，考量科技資訊安全及隱私議題等等。這些議題在生成式 AI 出現後更加受到重視，例如，引導學生在教學過程中尋找可靠的網路資訊來源，運用科技進行學生成就評量時展現倫理行為等（教育部學術倫理電子報，2024）。

學者 Celik（2023）針對教師 AIPACK 量表與 AI 倫理整合之研究成果，提出量表用以測量教師專業發展在 AI 面向上的成長。本文採用其所提出的四項教育 AI 倫理分析原則，分別為：透明度（transparency）、公平性（fairness）、包容性（inclusiveness）與問責性（accountability），說明如下。

1. 透明度（Transparency）

透明度意指 AI 系統的運作應具有可解釋性（explainability），使用者能理解其背後的決策邏輯與資訊來源（Floridi et al., 2018）。在教育應用中，透明度涉及 AI 推薦系統是否揭露其演算法依據、數據使用方式與模型限制，以及學生與教師是否具備足夠的知情權來選擇或拒絕使用該系統（Giannini, 2023；Spector, 2016）。

若 AI 系統作為評量或學習診斷工具，其是否清楚標示何種數據被使用、如何影響學習評估結果，即為透明度的實踐關鍵。教育系統若未能提供足夠的說明與使用警示，可能造成學生或教師過度依賴，並忽略 AI 判斷可能的偏誤與侷限。

2. 公平性（Fairness）

公平性關注 AI 技術是否會再製既有社會不平等結構，例如性別、種族或社經地位的偏見（Binns, 2018）。教育場域中，公平性特別關係到 AI 系統是否會因為訓練數據不均或演算法設計缺陷而產生不利特定學生群體的結果，比如某些學生可能因數據資料不完整或偏差而獲得較優待遇，進而擴大教育不平等（Parsons, 2021）。

舉例而言，若 AI 系統未能妥善處理語言或文化差異，可能導致非主流語言學生在學習推薦中處於不利地位（Holmes et al., 2019）。此外，公平性亦牽涉到資源分配，例如弱勢學校是否能平等取得優質 AI 教學工具與支援（Spector, 2016； Parsons, 2021）。

3. 包容性（Inclusiveness）

包容性強調 AI 技術在設計與應用時，應主動納入多元背景與能力的使用者，包括語言、文化、學習風格與特殊教育需求者（UNESCO, 2021）。在教育場域中，AI 若以單一標準對待所有學生，將可能排除部分學習風格差異大或語言能力不同的學生，造成新型態的數位排除。

例如，若 AI 輔助學習平台無法支援雙語學生或視覺障礙學生的需求，則形同將其排除在個別化教學之外。因此，包容性並非附加條件，而是確保教育機會平等與文化敏感性的核心原則。

Alevizos 等人（2023）指出，隨著 AI 在專案管理中的廣泛應用，AI 系統的「包容性」成為不可忽視的關鍵議題。所謂 AI 的包容性，指的是能公平對待所有使用者、避免因性別、種族、年齡等差異而產生偏見的系統。研究強調，現行 AI 系統常因訓練數據偏頗、開發者的無意識偏見，以及來自網路與用戶回饋的偏差資訊而複製甚至加劇既有的社會不平等。此外，包容性在專案管理中不僅能提升決策品質與任務分配的公平性，也能透過語言支援、會議安排與文化敏感度，打造更尊重多元的工作環境。為了強化 AI 的可解釋性與使用者信任，Alevizos 等人 (2023) 強調「可解釋人工智慧」（Explainable AI, XAI）與「決策透明度」的角色，並提出一套具體的評估指標與統一包容性分數（Unified Inclusiveness Score, UIS）來衡量 AI 在技術、互動、倫理與社會層面上的包容表現。研究亦呼籲未來應發展能即時過濾回饋偏見的互動式機制，並鼓勵跨領域合作，以落實 AI 在各產業中的公平與有效應用。

4. 問責性（Accountability）

問責性是 AI 倫理治理中最核心卻也最具爭議的原則，指當 AI 系統出現錯誤、偏誤或歧視結果時，應有明確的責任機制與干預可能性（Cath et al., 2018）。在教育場域，教師是否需為 AI 建議結果負責？學生若因 AI 評量系統遭受不當分類，應由誰來解釋與修正？這些問題正日益浮現。

Tomé et al.（2022）明確指出，人類應保有最後的決策權，教師應具備足夠判斷力與制度支持來監督 AI 在教學過程中的應用。若缺乏制度設計與角色界定，問責將流於空泛，反使 AI 成為模糊責任的遮蔽工具。

（二）聯合國教科文組織（UNESCO）的生成式 AI 倫理建議

聯合國教科文組織（UNESCO）於 2023 年發布的《生成式 AI 與未來教育》（Generative AI and the Future of Education）(Giannini, 2023）與 2022 年發布的《教育與研究中的生成式 AI 應用指引》（Guidance for Generative AI in Education and Research）（Holmes & Miao, 2023）中，提出了相關的倫理規範。

UNESCO 在 2023 年先後發布兩份關於生成式 AI 與教育的政策文件，展現其作為全球教育倫理領導者的治理觀點。Generative AI and the Future of Education 以教育轉型的全球視野出發，主張 AI 應當以促進人權、社會正義與文化多樣性為前提，批判 AI 產業化與數據壟斷的傾向，提醒教育系統不得讓生成式 AI 主導學習內容與價值（Giannini, 2023）。

緊接著的 Guidance for Generative AI in Education and Research 則以實務導向為主，針對教育部門與教師如何導入生成式 AI 提出十大指導原則，特別強調：不應強制使用生成式 AI，應保障學生的選擇權與知情權；教師需具備識別偏誤、理解演算法與解釋性結果的能力；各國政府應設立 AI 治理機制，確保技術開發符合教育公共性（Holmes & Miao, 2023）。

UNESCO 的政策強調「人類應保有最終決策權」（human-in-command）、「防止文化單一化與數據殖民」等概念，試圖超越純粹技術導向，從教育人權與文化正義出發重新定義 AI 應用。

（三）歐盟（2022）的生成式 AI 素養原則

歐盟強調教師應作為數位素養與民主守門人。歐盟執委會發布的《Guidelines for Teachers and Educators on Tackling Disinformation and Promoting Digital Literacy》雖未直接針對生成式 AI，但在數位科技與教育倫理的觀點上具備高度參考價值。該文件以教師專業發展為核心，提出「資訊判讀力、批判思考、數位公民意識」為當代教師不可或缺的數位素養構面（Tomé et al., 2022）。

其中強調教師在數位環境中應扮演下列角色：假訊息識別者（disinformation detector）；多元媒體教育推動者（media and information literacy promoter）；技術應用與倫理風險的自律者（reflective and ethical user of technologies）。雖文件未系統化提出 AI 倫理原則，但其強調資訊透明度、學生自主性與對技術過度依賴的反思，已隱含對「透明度」「問責性」等價值的重視。

（四）我國教育部數位教學指引中的倫理建議

教育部於 2024 年 1 月發布《中小學數位教學指引 3.0》，標誌著國內首次

於官方政策中系統性納入生成式 AI 使用說明。該指引定位為教師教學工具應用的輔助手冊，內容涵蓋 AI 平台的種類、使用建議與注意事項，包括：建議教師使用具備「內容標註功能」的平台，以利學生辨識 AI 生成內容、提醒教師注意平台的資安風險與資料使用規範、提供教師備課與設計評量的應用情境案例等（教育部 , 2024）。

儘管《中小學數位教學指引 3.0》提及「不應完全依賴生成式 AI」「應保有教師專業判斷」等語句，但整體取向仍偏向技術使用建議，對於 AI 系統的倫理風險、偏誤處理或文化包容性等面向，較少明確規範或引導。教師專業角色亦較少與倫理判斷結合，更多著墨於「操作能力」「平台選擇」「功能理解」等技術素養層面。

三、生成式 AI 倫理向度的比較

綜合分析上述三份政策文件可歸納出以下幾項初步觀察：首先，倫理原則強度差異：UNESCO 系統性建立 AI 倫理治理架構，歐盟以數位素養與民主價值為核心，臺灣則偏向工具操作與課堂應用，尚未建構系統性的倫理實踐架構。

其次，教師角色定位不同：UNESCO 與歐盟強調教師的倫理判斷力與價值引導者角色，臺灣則仍以「技術熟悉者」作為教師應具備的 AI 能力。

最後，落實策略與配套程度：UNESCO 提出治理機制與教師培訓建議，歐盟強調課程整合與公民素養，臺灣以個別平台應用為主，缺乏全國性制度設計。

表 1 生成式 AI 倫理向度的比較

倫理原則	UNESCO（2023）	歐盟（2022）	教育部（2024）
透明度（Transparency）	強調 AI 應具可解釋性；教師與學生須了解 AI 生成邏輯與資料來源；強調知情同意與選擇權。	教師須培養學生的批判性資訊判讀力與來源識別能力；間接強調資訊透明度。	提醒教師告知生成式 AI 使用狀況與標註內容。
問責性（Accountability）	明確主張人類應保有最終決策權；提出建立 AI 校園治理與風險審查機制。	鼓勵教師反思技術結果與引導學生負責任使用；未明列技術錯誤的制度處理機制。	未提及問責機制與責任歸屬，強調教師專業介入與決策角色模糊。
公平性（Fairness）	指出 AI 偏誤與歧視風險；建議審查模型是否造成學生族群差異的結果偏誤。	關注數位歧視與刻板印象教育；鼓勵教師引導學生反思資訊偏見。	從差異化教學技術操作層面，提醒教師注意生成式 AI 的演算法偏誤、教育不平等或公平性問題。

包容性（Inclusiveness）	提倡文化多樣性、語言權與數據主權；批判資料殖民現象。	重視多語資源與文化理解能力的建構；強調數位包容與反仇恨教育。	提及差異化學習與學生背景考量，但未聚焦於 AI 設計與文化包容的倫理思維。

資料來源：研究者自行整理

四、對我國中小學教學上運用生成式 AI 的建議

我國目前教學現場，因應生成式 AI 的興起，亦面臨新的教學與倫理挑戰。首先，學生對 AI 生成內容的過度依賴，導致其批判性思考與問題解決能力日漸削弱，進而影響學習品質與深度（張仁家、陳建州，2024）。教師亦難以判斷學生作業是否為 AI 生成，進而引發學術誠信與評分公正性問題（楊智傑，2025）。此外，由於生成式 AI 內容的正確性與資料來源經常無法驗證，導致誤用風險大增，亦可能產生性別、種族等偏見，加劇教育上的不公平現象（吳清山，2024；陳俊民、伍柏翰，2025）。

不僅如此，AI 的運作邏輯與人類理解世界的方式存在根本差異，例如 ChatGPT 等大型語言模型本質上僅透過語言機率預測產生內容，並無真正的理解與思考能力（楊智傑，2025）。這可能導致學生誤將其當作具備知識判斷力的學習來源，進一步弱化其認知基礎與資訊素養。同時，學生在未經引導下將敏感個資輸入 AI 工具，也可能造成資安風險與個人隱私外洩問題（謝昀澤、趙慶宏，2023）。

面對這些挑戰，學界與教育主管機關陸續提出應對策略，包括研訂 AI 倫理準則、設計 AI 素養課程、建立教學使用規範與偵測機制，並強調教師應優先培養自身 AI 素養，以善導學生負責任地運用 AI 工具（陳俊民、伍柏翰，2025；張仁家、陳建州，2024）。綜合以上討論，本文提出四項建議，提供我國教師在生成式 AI 的浪潮下，關注倫理議題的焦點，作為自我檢視以及規劃專業成長的可行方向。

（一）透明度與知情權

教師應該告知學生生成式 AI 的使用狀況，並標註 AI 生成的內容。然而，這還不夠，教師應該進一步解釋 AI 的運作原理和資料來源，確保學生了解 AI 生成邏輯，並強調知情同意與選擇權的重要性。

（二）問責機制與責任歸屬

目前的關注焦點大多在技術的使用上，而缺少反思的成分。建議透過強化

問責機制與責任歸屬的認知訓練，讓教師明確自己的專業介入與決策角色，並建立清晰的問責機制，確保在技術錯誤發生時有相應的處理制度。

（三）公平性與教育不平等

現有建議未提及演算法偏誤、教育不平等或公平性問題。教師應該關注 AI 可能帶來的偏見與歧視風險，並審查模型是否會造成學生族群差異的結果偏誤。這樣可以確保教育的公平性。

（四）包容性與文化多樣性

雖然目前的 GAI 使用多強調需要差異化學習與學生背景考量，但未聚焦於 AI 設計與文化包容的倫理思維。教師應該提倡文化多樣性，例如原住民族的資料來源確認問題，以及對於微歧視等的相關議題，以增進優質教育的實現。同時，也應該重視多語資源與文化理解能力的建構，並強調數位包容與反仇恨教育。

參考文獻

■ 吳清山（2024）。AI 時代學術倫理的風險與因應。**臺灣教育評論月刊**，13（10），12-18。

■ 張仁家、陳建州（2024）。高中教學現場運用 AI 之挑戰與應對。**臺灣教育評論月刊**，13（12），84-89。

■ 教育部（2024）。**教育部中小學數位教學指引 3.0**。取自 https://pads.moe.edu.tw/pads_front/index.php?action=download

■ 教育部學術倫理電子報（2024）。大學校園因應生成式 AI 之指引及教學建議。**教育部學術倫理電子報**。取自 https://ethics.moe.edu.tw/resource/epaper/html26/html

■ 陳俊民、伍柏翰（2025）。生成式 AI 工具的崛起——現代教育與設計的機遇與挑戰。**台灣教育**，52，29-37。

■ 楊智傑（2025）。生成式 AI 融入教學的風險和機會。**台灣教育**，52，59-66。

■ 謝昀澤、趙慶宏（2023）。生成式 AI 潛藏的道德風險與資安危機。**會計研究月刊**，450，82-87。

■ Alevizos, V., Georgousis, I., Simasiku, A., Karypidou, S., & Messinis, A. (2023). Evaluating the Inclusiveness of Artificial Intelligence Software in Enhancing Project Management Efficiency--A Review. *arXiv preprint arXiv:2311.11159*.

■ Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. In *Conference on Fairness, Accountability and Transparency* (149-159). PMLR.

■ Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial intelligence and the ‘good society’: the US, EU, and UK approach. *Science and Engineering Ethics*, *24*, 505-528.

■ Celik, I. (2023). Towards Intelligent-TPACK: An empirical study on teachers’ professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior*, *138*, 107468.

■ Floridi, L. et al. (2018). *AI4People’s Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*. Retrieved from https://www.eismd.eu/wp-content/uploads/2019/03/AI4People’s-Ethical-Framework-fora-Good-AI-Society.pdf

■ Giannini, S. (2023). *Reflections on Generative AI and the Future of Education*. https://unesdoc.unesco.org/ark:/48223/pf0000385877

■ Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. *Center for Curriculum Redesign*.

■ Holmes, W., & Miao, F. (2023). *Guidance for Generative AI in Education and Research*. *UNESCO Publishing*.

■ Parsons, T. D. (2021). Ethics and educational technologies. *Educational Technology Research and Development*, *69*(1), 335-338.

■ Spector, J. M. (2016). Thinking about educational technology and creativity. *Educational Technology*, *56*(6), 5-8.

■ Smakman, M., Vogt, P., & Konijn, E. A. (2021). Moral considerations on social

robots in education: A multi-stakeholder perspective. *Computers & Education*, *174*, 104317.

■ Tomé, V., Kılıç, A. M., Bargaoanu, A., Varanauskas, A., Hague, C., Sádaba, C., Hjorth, C., Frau-Meigs, D., Kyza, E., & Thinsz, G. (2022). *Guidelines for Teachers and Educators on Tackling Disinformation and Promoting Digital Literacy through Education and Training*. https://op.europa.eu/en/publication-detail/-/publication/a224c235-4843-11ed-92ed-01aa75ed71a1/language-en

■ UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. https://unesdoc.unesco.org/ark:/48223/pf0000380455

■ Yurdakul, I. K., Odabasi, H. F., Kilicer, K., Coklar, A. N., Birinci, G., & Kurt, A. A. (2012). The development, validity and reliability of TPACK-deep: A technological pedagogical content knowledge scale. *Computers & Education*, *58*(3), 964-977.