

國立臺灣大學生成式 AI 使用與研究誠信指引

115 年 6 月

本指引參酌 2010 年世界研究倫理大會研究誠信聲明、人工智慧基本法第 4 條及本校相關規範訂定，旨在引導本校師生及研究人員妥善使用生成式 AI，維護研究誠信、研究倫理及學術公信力。本校原則上不禁止將生成式 AI 作為學習或研究之輔助工具；惟其使用應符合法令、本校規範及各學術單位之專業要求。

壹、基本原則

生成式 AI 已廣泛用於資料檢索、整理、翻譯、文字與圖像生成及程式草擬等研究活動。其可提升效率，亦可能帶來錯誤、虛構、偏見、隱私外洩、機密洩漏、智慧財產權爭議及責任歸屬不清等風險。研究人員於使用生成式 AI 時，應遵循下列原則。

- 一、**以人為本與人類自主**：生成式 AI 僅為研究輔助工具，不得取代研究人員之學術判斷、原創思考與責任承擔。研究之構想、設計、分析、詮釋、結論及發表，均應由研究人員主導。
- 二、**誠實與透明**：研究人員應依研究情境、學門慣例及發表規範，適當揭露所使用 AI 工具之名稱、版本、使用時點、目的、方式及介入程度。
- 三、**核實與可追溯**：研究人員應查核資料來源與 AI 生成內容之真實性、正確性及適用性，並保存必要使用紀錄，以利說明、查核與追溯。
- 四、**隱私保護與資料治理**：研究人員應妥善保護個人資料、敏感資料、未公開資料及其他機密資訊，不得因使用生成式 AI 而違反資料治理、保密義務或相關法令。
- 五、**公平與不歧視**：研究人員應注意 AI 模型可能存在之偏見、遺漏與不公平，避免研究設計、分析、詮釋與建議因此失衡或失真。
- 六、**責任歸屬**：研究人員就其使用生成式 AI 之方式、輸入資料、生成內容、採納決定及最終研究成果，負最終責任。生成式 AI 不得列名為作者，亦不得作為卸責之理由。

貳、使用規範及注意事項

生成式 AI 之使用，應符合研究誠信、法令規範及本校相關規定。

- 一、**正確認識生成式 AI 之侷限**：生成式 AI 雖可提升整理、摘要、翻譯及文字草擬之效率，惟仍可能產生虛構、不完整、錯誤或同質化內容。學術研究之核

心仍在於問題意識、批判思考、方法能力與原創貢獻。教師及資深研究人員應加強學生、研究助理及新進研究人員對其侷限與風險之認識。

- 二、**明確界定其輔助角色及作者列名**：生成式 AI 不具法律人格，亦無法承擔學術、法律或倫理責任，故不得列名為論文、報告、研究計畫或其他學術成果之作者或共同作者。本校各單位所出版之學術期刊，應於徵稿須知或相關文件明定此一原則。
- 三、**適當揭露 AI 之使用**：於研究或論文寫作中使用生成式 AI 者，應於論文、研究報告、附錄、方法說明、致謝或依期刊要求，揭露所使用工具之名稱、版本、日期、使用目的、操作方式、介入程度，以及 AI 生成內容如何納入研究或文稿。必要時，得區分研究過程中之使用與文稿撰寫階段之使用。
- 四、**保存使用紀錄並確保可稽核性**：研究人員應盡可能保留與研究判斷有關之 AI 使用紀錄，包括重要提示詞、輸入內容、生成結果、後續人工修正及採納理由。各學術單位或研究機構得視需要建立標準化之 AI 使用紀錄格式，供所屬人員遵循。
- 五、**核實幻覺、偏見與不實資訊**：研究人員不得逕以 AI 生成內容作為可信資料。對於文獻、引註、數據、圖像、研究紀錄、實驗設計、分析結果及其他重要內容，均應逐一查核其存在、來源、正確性及適用性，並審慎檢視其是否對特定族群、議題或觀點存在偏差、遺漏或不公平對待。
- 六、**落實資料治理、隱私與機密保護**：使用生成式 AI 前，研究人員應先評估資料性質、風險與服務環境，避免將個人資料、敏感性個人資料、未公開研究資料、營業秘密、審查文件或其他依法或依契約應保密之資訊，任意輸入開放式或無法掌握資料流向之 AI 工具。於確有使用個人資料之必要時，除應符合法令規定之使用目的外，並應遵守資料最小化、去識別化(假名化)及必要安全維護等原則。
- 七、**研究倫理審查**：研究計畫如涉及以個人資料或其他受保護資料使用生成式 AI，應於研究倫理審查申請文件中說明其必要性、方法、資料處理流程及保護措施。
- 八、**審查保密義務**：本校師生及研究人員受邀擔任論文、學位論文、研究計畫或其他學術成果之審查人員時，應遵循委託單位規範，原則上不得將受審文件輸入開放式生成式 AI 工具。
- 九、**遵守智慧財產權與其他法律規範**：使用生成式 AI 從事研究或教學時，應注意是否涉及著作權、專利權、營業秘密及其他智慧財產權，並避免生成、利用

或散布違反公序良俗或法令禁止之內容。研究人員不得因 AI 可快速改寫、翻譯或重組內容，即忽視原始著作、資料或成果之權利保護。

參、可能違反研究誠信之情形

為說明生成式 AI 之使用與研究誠信之關聯，以下列舉若干可能涉及違反研究誠信之情形。下列例示僅供參考，具體認定仍應依個案事實、使用程度、揭露情形及相關規範綜合判斷。

- (一) 僅提供提示詞或其他有限輸入，對生成內容無實質學術貢獻，卻逕將 AI 產出內容作為自己之研究成果者，可能構成抄襲或代寫。
- (二) 將他人外文著作之全部或部分，藉由生成式 AI 翻譯為中文後作為自己之研究成果者，可能構成抄襲。
- (三) 將他人研究計畫申請書、未公開著作、審查資料或其他未公開成果輸入生成式 AI 後加以生成、改寫或重組，並據以作為自己之研究成果者，可能構成抄襲。
- (四) 於論文或研究報告中大量使用生成式 AI，且其對內容形成具有相當影響，卻未依規定揭露其使用情形，情節重大者，可能構成研究成果未註明出處或註明不當之抄襲。
- (五) 以生成式 AI 虛構不存在之研究資料、文獻、引註、數據、影像、實驗結果、觀測結果或研究紀錄者，可能構成造假。
- (六) 以生成式 AI 改寫、變更、扭曲或重組原始資料、數據、紀錄、影像、實驗結果或觀測結果，致其失真者，可能構成變造。
- (七) 使用生成式 AI 改寫或翻譯自己已發表之中文或外文著作，而未適當引註或揭露者，可能構成自我抄襲。
- (八) 擔任論文、學位論文或研究計畫審查人員時，未經同意將受審文件輸入生成式 AI，並以其生成內容作為審查意見之部分或全部者，可能構成其他違反學術倫理行為。

肆、附則

- 一、本校各單位得依其學門特性、研究型態與風險程度，另訂更具體之作業規範。
- 二、本指引未盡事項，依相關法令、本校研究倫理與學術倫理規範，以及各單位之規定辦理。