AI 與學術期刊:賦能與祛魅

張耀銘

[提　要]　1956 年達特茅斯會議標誌著人工智能的誕生,符號主義、連接主義與行為主義三大學派共同構建了 AI 的理論基石。歷經 70 年“兩起兩落”的技術浪潮,生成式人工智能(GAI)在 21 世紀實現了從“分析系統”到“生成系統”的範式躍遷,推動人工智能從“模仿人類”走向“協同人類”的“後學派時代”。生成式人工智能對學術生態的影響呈現多維變革:在科研範式維度,催生“第五科研範式”,AlphaFold 蛋白質預測、AI 輔助定理證明等案例,打破了傳統學科壁壘,實現數據驅動與模型驅動的融合;在知識生產維度,從主題構建、文獻綜述到文本修飾,AI 以高效的信息整合能力輔助作者重構創作模式;在編輯審稿維度,學術期刊編輯必須提高識別 AI 生成論文的能力,從編輯加工角色進化為“人機協作管理者”,通過一次次與 AI 的合作,實現熱點追蹤、選題策劃、稿件審讀、同行評審等工作效率的提升。然而,技術狂飆背後暗藏深層風險:訓練數據的版權模糊性、AI 生成物的著作權歸屬爭議、“AI 幻覺”導致的虛假信息傳播,均對現有學術規範與法律體系提出挑戰。如何在技術賦能中守護學術本真,在範式變革中堅守人文內核,不僅是學術期刊面臨的時代命題,更是整個知識生產領域需要共同書寫的答卷。

[關鍵詞]　生成式人工智能　三大學派　人機協作　AI 幻覺　版權侵權　倫理風險

[中圖分類號]　G234　[文獻標識碼]　A　[文章編號]　0874 - 1824 (2025) 04 - 0125 - 17

毋庸諱言,1956 年注定是在人類歷史上留下濃墨重彩一筆的特殊年份。這一年的夏天,美國新罕布什爾州漢諾威鎮的達特茅斯學院召開了“人工智能夏季研究項目”會議,先後有約翰・麥卡錫、馬文・明斯基、塞爾弗里奇、內森・羅切斯特、克勞德・香農、赫伯特・西蒙和艾倫・紐厄爾等四十多位數學、心理學、神經學、信息科學和計算機科學等方面的專家學者參與了這次暑期會議。在這次會議上麥卡錫非凡地提出,“讓機器達到這樣的行為,即與人類做同樣的行為”可以被稱為人工智能(Artificial Intelligence AI)。[①]1956 年也就成了人工智能元年。1969 年國際人工智能聯合會議組織在美國成立,標誌著人工智能學科正式確立。人工智能在之後的發展進程中,基本形成了符號主義、連接主義、行為主義三大學派。三大學派研究的側重點雖然不同,但都豐富了人工智能的理論體系。正如基辛格所言,人工智能有望在人類體驗的所有領域都帶來變革,“變革的核心將發生在哲學層面,即改變人類理解現實的方式以及我們在其中所扮演的角色。”[②]

近年來隨著生成式人工智能技術的廣泛應用,正推動知識生產的“範式革命”,重塑論文寫作

模式。學術期刊作為知識生產內容的重要載體,從內容生產到傳播模式,從審稿流程到學術誠信體系,均面臨系統性重構。在生成式人工智能應用場景下,人與 AI 的合作恐怕將成為常態,學術期刊編輯如何認識自身與 AI 的關係?如何通過人機協作,既提高工作效率又能保持科學精神和人文底線?學術期刊出版業態會被如何重新定義與重構?又會帶來哪些版權風險與倫理問題?這些問題都值得學術期刊從業者進行深入思考。本文僅做一些初步探討,懇請方家不吝賜教。

一、生成式人工智能的技術崛起

從歷史維度審視,學術期刊與 AI 技術的發展呈現出深刻的雙向驅動共生關係。這種關係在不同階段以不同形式展開,既體現為學術期刊對 AI 技術突破的記錄與傳播,也表現為 AI 技術對學術研究範式和出版模式的賦能與革新。20 世紀 50~60 年代是人工智能發展的重要起步階段,學術期刊在這一時期為早期 AI 理論的孕育提供了關鍵支持。早在 1950 年艾倫·圖靈在《心靈》(*Mind*)發表《計算機器與智能》論文,就"提出了如今人盡皆知的圖靈測試理論以及機器學習、遺傳算法、強化學習等多種概念"。這篇論文為人工智能研究奠定了重要的理論基礎。也是在 1950 年,克勞德·香農在《哲學雜誌》(*Philosophical Magazine*)發表《為下棋編程計算機》,首次系統地提出了用算法實現計算機下棋的理論框架,為後來人工智能在遊戲領域的應用奠定了基礎。人工智能的歷史與計算機歷史相伴,隨著馮·諾依曼結構計算機的誕生,人工智能悄然在大學實驗室裡嶄露頭角。其實在麥卡錫 1956 年正式提出人工智能概念之前,"香農早已完成他的三大通信定律,為計算機和信息技術打下基礎。明斯基已經造出第一台神經網絡計算機(他和同伴用 3,000 個真空管和一台 B-24 轟炸機上的自動指示裝置來模擬 40 個神經元組成的網絡),不久寫出了論文《神經網絡和腦模型問題》。這篇論文當時沒有太受重視,日後卻成為人工智能技術的鼻祖。而圖靈早在 1950 年就提出了如今人盡皆知的圖靈測試理論以及機器學習、遺傳算法、強化學習等多種概念。"[③] 1955 年美國西部計算機聯合大會在洛杉磯召開,塞弗里奇提供了一篇關於模式識別的論文,代表了企圖模擬神經系統的一派觀點;紐厄爾則探討了計算機下棋,試圖模擬心智系統的方法。討論會主持人、美國神經科學家沃爾特·皮茨總結時指出兩者殊途同歸,這一觀點也預示了人工智能後續幾十年關於"結構與功能"兩條路線的發展態勢。經過近 70 年的發展歷程,人工智能研究經歷了兩起兩落,終於在 21 世紀初迎來了第三次發展浪潮。

1. 人工智能的潮起潮落

(1)人工智能的第一次浪潮(1950~1969)。以英國數學家艾倫·圖靈 1950 年發表論文《計算機與智能》,提出"圖靈測試"為標誌,符號主義、連接主義和行為主義等人工智能研究的多個分支出現,掀起了人工智能研究的第一次浪潮。1958 年麥卡錫開發了 LISP 編程語言,成為一種通用高級計算機程序語言。1968 年愛德華·費根鮑姆使用 LISP 語言開發了第一個專家系統 DENDRAL,標誌著符號主義學派逐漸形成。此後符號主義盛行,LISP 編程語言更是長期壟斷人工智能領域的應用。1958 年弗蘭克·羅森布拉特在《心理學評論》(*Psychological Review*)發表論文《感知器:大腦中信息存儲與組織的概率模型》,首次系統性地提出了感知器模型,為人工神經網絡的發展奠定了基礎。1969 年馬文·明斯基和西摩·帕爾特在著作《感知器》中指出單層感知器無法解決非線性問題,其存在局限無法突破應用瓶頸。這一結論對連接主義學派的神經網絡研究產生了重大影響,進而導致了"第一次 AI 寒冬"的到來。存在局限無法突破應用瓶頸,導致連接主義學派的神經網絡研究進入低谷。儘管如此,這段時期仍被後來的研究者稱為"黃金時代"。

(2)人工智能的寒冬(1970~1979)。1965 年 12 月,蘭德公司顧問休伯特·德雷福斯發布《人工智能與煉金術》的研究報告,對蘭德公司主導的人工智能研究提出了嚴厲的批評:"包括國際象棋中的組合爆炸、啟發式方法在機器定理證明中的停滯、10 年來投入了 1,600 萬美元的機器翻譯面臨的上下文歧義問題、模式識別只能做到識別手寫的摩爾斯電碼和英文字母的水平。"[④]這使得人工智能研究的聲望在美國空前下降,很難再從美國國防部高級研究計劃局(DARPA)申請到經費資助。1973 年,英國科學研究理事會發布了劍橋大學知名物理學家詹姆斯·賴特希爾教授提交的一份關於英國 AI 研究現狀的調查報告。報告的結論是僅支持對神經生理學和心理學過程的計算機模擬,而放棄對機器人和語言處理的資助,這導致科學研究理事會終止了多數大學 AI 研究的支持。一連串的黑天鵝事件,導致人工智能研究跌入低谷經歷至闇時刻。不過寒冬中還是出現了一抹亮色,1970 年愛思唯爾創辦了《人工智能》(*Artificial Intelligence*)期刊,成為全球首個專注 AI 研究的學術陣地,系統收錄了符號主義學派的經典成果,並通過同行評審機制篩選出具有突破性的研究,推動 AI 領域的學術規範形成,促進研究者之間的批判性討論和思想碰撞,為後續的 AI 研究提供了可信賴的參考依據。

(3)人工智能的第二次浪潮(1980~1992)。20 世紀 80 年代至 90 年代初,被認為是人工智能的高速發展時期。第一個標誌是"專家系統"的研發,該系統將人類專家的知識和經驗以一定的形式存儲在計算機中,通過推理機制來解決特定領域的問題,讓人工智能從理論研究走向了醫療、氣象、地質等領域實際應用。第二個標誌是"反向傳播算法"的提出與應用,讓人工智能神經網絡研究再次引起了人們的廣泛關注。1986 年 7 月,傑弗里·辛頓與戴維·魯梅爾哈特等合作在《自然》雜誌發表了《反向傳導誤差的學習表徵》論文,提出了反向傳播算法。這種算法的核心思想是通過計算每一層的誤差,並將這些誤差逐層向後傳播,進而調整每一層的權重和偏置,使得整個網絡的輸出誤差最小化,這就創造性地解決了多層神經網絡無法高效訓練的理論難題。第三個標誌是"行為式機器人"概念的提出,1986 年羅德尼·布魯克斯批判了傳統的符號主義人工智能過於依賴複雜的推理系統,忽視了身體與環境的互動對智能行為的關鍵性貢獻。他強調機器人通過身體與環境的動態交互實現自主學習和進化,其核心在於將感知、行動與認知深度融合。這一理論被視為行為主義學派在人工智能領域的重要代表,為後來的具身智能研究奠定了堅實基礎,並對當時的機器人學領域產生了深遠影響。但從整體上看,人工智能的第二次浪潮"仍然籠罩著濃厚的學術研究和科學實驗色彩,雖然激發了大眾的熱情,但遠沒有達到與商業模式、大眾需求接軌並穩步發展的地步。"[⑤]

(4)人工智能的低谷(1993~2005)。人工智能第二次低谷的起點之所以劃在 1993 年,主要考慮五個關鍵節點:其一,日本第五代計算機研究開發計劃,因最終沒有突破關鍵性技術難題宣告失敗;其二,基於規則和符號推理的專家系統表現不佳,難以處理不確定性問題風光不在;其三,當時的硬件算力和數據規模無法支撐機器學習(尤其是深度學習)的發展,導致許多理論難以落地;其四,部分 AI 項目(如語音識別、機器翻譯)因商業化失敗被詬病,投資信心受挫;其五,"奇點理論"的提出。1993 年,美國計算機科學家兼著名科幻作家弗諾·文奇寫了《即將到來的技術奇點:後人類時代如何求生》,對人工智能潛在的風險發出警告。論文開頭就危言聳聽地寫道:"在未來 30 年間,我們將有技術手段來創造超人的智慧。不久後,人類的時代將結束。"[⑥]弗諾·文奇成為最早的人工智能威脅論的吹哨者。這一年馬賽克瀏覽器的流行,使得覆蓋互聯網的萬維網成為新的連接世界的平台,也引發了硅谷的電子商務革命。"在這種大的背景下,DARPA 的新任領導認為人工智

能並非‘下一個浪潮’,因此大幅度地削減了資助,AI 研究再次遭遇經費危機陷入低谷,並在漫長的寒冬中蟄伏起來。”[⑦]儘管 1997 年 IBM 的超級計算機“深藍”擊敗了當時的世界冠軍加里・卡斯帕羅夫,2001 年保羅・維奧拉和邁克爾・瓊斯提出的 Viola-Jones 算法(人臉檢測算法)廣泛應用於數碼相機和安防監控,但只不過是資本寒冬時期燃放的幾朵爆竹煙花。

2. 生成式人工智能技驚世人

2003 年約書亞・本吉奧在《機器學習研究雜誌》(*Journal of Machine Learning Research*)發表《神經概率語言模型》論文,其核心貢獻在於將分布式表示與概率建模深度融合,不僅顯著提升了語言模型的性能,更開創了 NLP 的深度學習時代,為計算機理解和處理人類語言提供了新的思路和方法,推動了機器翻譯和自然語言理解系統的重大轉變。2006 年傑弗里・辛頓提出具有劃時代意義的“深度神經網絡”框架和“逐層預訓練”方法,解決了深層神經網絡訓練困難的問題,引發了“深度學習”革命。2024 年諾貝爾物理學獎授予傑弗里・辛頓,以表彰他利用人工神經網絡進行機器學習的奠基性發現和發明。2014 年伊恩・古德費洛提出“生成對抗性網絡”概念,創新機器學習算法,在圖像生成等領域有了廣泛的應用。2017 年谷歌與多倫多大學研究團隊共同發表論文《注意力滿足一切》,提出了 Transformer 自然語言處理模型,突破傳統架構,完全基於注意力機制,大幅提升了並行計算能力和長序列處理效率。2021 年理希・博馬薩尼等學者共同發表的報告《基礎模型的機遇與風險》中將 Transformer 模型稱為“基礎模型”。可以說,“大基礎模型 Transformer 已成為這一輪人工智能迭代發展的範式之基。並且,已有諸多實驗證明,基礎模型規模越大,效果就越好。”[⑧]

生成式人工智能(Generative Artificial Intelligenc,GAI)是人工智能的重要分支,指一種基於算法、模型和規則生成文本、圖像、音頻、視頻、代碼等內容的技術。生成式人工智能與傳統 AI 的數據處理和分析功能不同,其核心技術包括大規模數據集、生成對抗網絡、大規模預訓練模型和人的存在這個價值尺度。“相較於過往各階段的人工智能,生成式人工智能的典型特徵是運作邏輯的轉變,即從分析式系統轉向為生成式系統,極大提升了技術的感知易用性和感知有用性,使人類能夠更便捷高效地應用先進技術。生成式人工智能通過學習和理解數據的分布,已經具備強大的生成能力、遷移能力和交互能力,展示出類人的感知與認知智能。”[⑨]生成式人工智能的發展已經突破了“科技”與“人文”的傳統邊界,例如技術邏輯的人文隱喻、多模態生成的跨學科屬性和人機何以價值對齊等問題,其對人類社會的滲透已深度觸及人文社會科學的核心議題。因此,生成式人工智能不僅僅是一個科學技術問題,也是一個人文社會科學問題。“倫理與技術同向而行,確保科技向善,人類需要發展對倫理問題的敏感性和負責任的應對能力。”[⑩]

2022 年 11 月,美國 OpenAI 公司發布一種大型語言模型 ChatGPT3.0,引起了學界和業界的高度關注,被媒體稱為“有史以來向公眾發布的最佳人工智能聊天機器人”。比爾・蓋茨甚至認為,它是一項革命性的技術,其出現的重要性不亞於互聯網和個人電腦的誕生。它的面世炒熱了人工智能生成內容(AIGC)這一概念,多家科技公司紛紛推出了 Bard、Ernie Bot、LLaMA 等同類產品。“ChatGPT 能夠自己讀取網絡上包括維基百科等在內的海量文本,從中模擬出語言生成模型。ChatGPT 在數學計算和數據儲存方面,遠遠超過了真人的大腦,正是因為有這兩個超強的能力,自然語言中的詞語可以真正被自動標注,實現高維向量化,形成複雜網絡關聯,人工神經網絡可以進行大規模運算得到最佳輸出。”[⑪]ChatGPT 能夠根據用戶需求生成符合要求的文本內容,根據用戶輸入的問題或疑難,提供個性化的答案和建議。同時,可以處理各種類型的自然語言任務,包括但

不限於對話交互、即時問答、文本生成、文字翻譯、論文摘要、撰寫論文和文學作品等,具有很高的語言自組織能力和內容的再創造能力。

2024 年 2 月,美國 OpenAI 公司推出的文生視頻模型 Sora,可以根據文字描述生成高質量、逼真的短視頻(最長約 60 秒)。Sora 擁有高度逼真的場景,能模擬物理世界,生成複雜的光影、運動和細節;擁有深入的語言理解能力,能準確解釋提示,生成表達豐富情感的角色,更好地理解並忠實反映用戶的文本指令;擁有多鏡頭生成能力,在單個生成視頻中創建多個鏡頭,同時保持角色和視覺風格的一致性;擁有從靜態圖像生成視頻的能力,可以準確地動畫化圖像內容,或擴展現有視頻;擁有物理世界模擬能力,如物體的移動和相互作用,展示了人工智能在理解真實世界場景並與之互動的能力。"Sora 作為生成式人工智能在視頻生成領域的劃時代產物,掀起了一場全新的智能生成革命。圍遶 Sora 產生的技術討論與技術想像層出不窮,在生成式人工智能的多維話語迷思中,Sora 的社會應用與未來前景愈發撲朔迷離。"⑫

2024 年 12 月,中國大語言模型 DeepSeek-V3 上線並同步開源,憑藉其卓越的性能和顯著的成本優勢,迅速成為全球矚目的焦點,被譽為"來自東方的神秘力量"。作為多模態智能系統,DeepSeek 支持文本、圖像、音頻、視頻、代碼等多種格式輸入,具備智能問答、內容生成、數據分析、任務管理和學習助手等核心功能,可應用於日常生活、家庭教育、職場工作等多個領域,能幫助用戶寫演講稿、制定旅遊攻略、解答學術問題、整理會議紀要等。DeepSeek 在生成式人工智能領域的最重要突破,體現在對中文信息的處理更加精準與流暢,鮮明的本土化特徵使其能夠輸出更符合中國文化背景的內容。從處理速度、分析深度和覆蓋廣度等多個方面,DeepSeek 正持續推動著生成式人工智能從輔助工具向知識生產主體的轉變,進而對人文社科學術研究範式的解構與重塑產生深遠影響。DeepSeek 採用開源策略,將模型權重、訓練框架及數據渠道全部開源,吸引了大量開發者和研究人員參與,形成活躍的社區,推動了技術的快速迭代和創新。

3. 生成式人工智能對三大學派理論的融合

生成式人工智能(GAI)作為當前人工智能領域的重要分支,其發展深受傳統人工智能三大範式——符號主義、連接主義和行為主義的影響。符號主義學派強調邏輯推理,認為人類認知和思維的基本單元是符號,認知過程是在符號表示上的運算。GAI 在一定程度上體現了符號主義的思想。如自然語言處理中的文本生成任務,模型會將輸入的文本理解為符號序列,通過對符號的處理和組合來生成新的文本。在生成代碼、摘要等任務中,模型同樣是基於對符號的操作來實現。但是 GAI 並非單純的符號處理。傳統符號主義面臨知識獲取瓶頸等問題,而 GAI 藉助深度學習等技術,能從海量數據中自動學習符號間的關係,無需人工手動編碼大量規則,這是對符號主義的發展與突破。

連接主義學派從神經生理學和認知科學出發,強調智能活動是由大量簡單單元通過複雜相互連接的人工神經網絡並行運行的結果。GAI 中的神經網絡架構,正是連接主義的具體應用。以圖像生成模型為例,模型由眾多神經元組成的網絡結構,通過對大量圖像數據的學習,調整神經元之間的連接權重,從而能夠生成逼真的圖像。再如 GPT-4 的思維鏈技術既需要神經網絡的特徵提取能力,又依賴符號化的推理步驟。生成式人工智能在訓練過程中採用的大規模並行計算方式,也符合連接主義並行運行的理念。它充分利用連接主義的優勢,通過不斷優化網絡結構和訓練算法,提升生成內容的質量和多樣性。

行為主義認為智能是系統與環境的交互行為,是對外界複雜環境的一種適應。GAI 通過引入

強化學習人類反饋機制,將行為主義的"試錯—反饋"範式融入生成過程。模型根據環境反饋的獎勵信號,不斷調整自身的生成策略,以更好地適應環境需求。在智能對話系統中,GAI 根據用戶的輸入(環境刺激)生成回復(行為反應),並通過用戶的反饋不斷優化回復策略,體現了從感知到行動的行為主義模式。

總之,生成式人工智能的發展正在消解傳統學派的理論邊界,它融合了三大學派的部分理念與技術,在不同領域展現出與各學派的千絲萬縷聯繫,推動了人工智能技術向更智能的方向發展。西方學術期刊對生成式人工智能的貢獻,遠不止於學術成果展示窗口的表層功能。學術期刊通過構建嚴謹的學術規範、促進跨學科對話、容納批判性聲音、引導資源流向,推動 AI 技術從"碎片化創新"走向"系統性突破"的生態系統。從某種意義上說,生成式 AI 的進化史,就是學術期刊上的理論爭鳴與實驗室裡的技術試錯相互滋養的歷史;Transformer 架構的論文曾因"看似簡單"遭遇拒稿,後來卻在學術期刊的不斷爭論中被重新審視;早期擴散模型的數學推導因"複雜晦澀"鮮受關注,卻在學術期刊的長期傳播中被工程師發現應用價值。當我們驚嘆於生成式人工智能的驚艷表現時,不應忘記那些刊載於紙本期刊頁面上的公式、圖表與爭論,這些看似枯燥的文字,正是學術期刊對 AI 發展最深刻的饋贈。所以,有學者認為 GAI 標誌著人工智能進入後學派時代。

二、人工智能對知識生產的賦能

學術研究與學術期刊是現代學術體系的核心組成部分與關鍵支柱。學術研究的本質是知識生產、問題回應和學術傳承與創新,而學術期刊則通過成果傳播、質量篩選、學術對話、知識存檔、規範引導履行著學術共同體公認的"知識守門人"的職能。二者深度依存、協同而行,共同支撐著知識的生產、傳播、驗證與積累。從這個意義出發,AI 技術對於學術研究的賦能抑或魅惑都會集中地表現在學術成果——論文之中,而論文要進入主流傳播領域,需要順利通過學術期刊的合規審查和價值判斷,所以,AI 的賦能或魅惑能否最終實現,實取決於學術期刊。因此,學術期刊能否正確引導學術研究利用 AI 賦能和實現對 AI 祛魅,就成為關鍵。要做到這一點,還在於學術界以及期刊界對 AI 如何賦能學術生產的看法和認識。

隨著人工智能(AI)從技術層面到社會層面的深度滲透,知識生產正在經歷前所未有的深刻變革。AI 對知識生產賦能的核心是從加速器到認知革命。首先,自然語言處理與海量數據分析能力使人類得以從信息迷霧中提煉洞見,實現知識創造指數級加速。其次,AI 以超越人類直覺的模式識別與複雜關聯分析能力,突破傳統認知邊界,揭示隱藏規律,催生跨學科融合的新知識圖譜。再次,AI 推動知識生產從"人類中心"向"人機協同"躍遷。人類提供批判思維、價值導向與意義追問,AI 則貢獻強大的計算能力與模式發現,共同編織更複雜深邃的知識網絡。AI 時代,正加速進入一個自然人、機器人、數字人"三人"共生的新時代。我們"需要注意其技術本身,但是更要關注其次生效應,更要關注其給語言、語言生活和語言治理帶來的新問題、新形態和新趨勢,關注'三人'之間的共生、共享、共治和共建。"[13]

1. 知識生產的"範式革命"

1962 年,美國科學哲學家托馬斯·庫恩在其出版的《科學革命的結構》中,第一次提出了"範式"理論並進行了系統性闡述。庫恩認為,"科學成就"既吸引了"一批堅定的擁護者",又為後來者"留下有待解決的種種問題",具有這兩個特徵就可以稱之為"範式"。庫恩強調,儘管"一個範式就是一個科學共同體的成員所共有的東西,而反過來,一個科學共同體由共有一個範式的人組成。"[14]

受範式理論的啟發,美國科學家、圖靈獎獲得者吉姆·格雷在 2007 年發表報告《科學方法的革命》,提出了人類科學研究經歷了四種範式的演進:第一種範式為伽利略時代之前的經驗觀察,第二種範式為牛頓之後的理論建構,第三種範式為計算仿真技術支持的仿真模擬,第四種範式為數據密集型的科學發現。而隨著人工智能的發展,有學者提出或將催生出"第五範式"——"在數據密集型範式的基礎上引入智能技術,強調人的決策機制與數據分析的融合,將數據科學和計算智能有效結合起來。"[15]

人工智能驅動的科學研究是以生成式為代表的人工智能技術與科學研究深度融合的產物。中國工程院院士李國傑研究員將智能化科研(AI4R)稱為第五科研範式。概括它的一系列特徵包括:"(1)人工智能(AI)全面融入科學、技術和工程研究,知識自動化,科研全過程的智能化;(2)人機融合,機器湧現的智能成為科研的組成部分,暗知識和機器猜想應運而生;(3)以複雜系統為主要研究對象,有效應對計算複雜性非常高的組合爆炸問題;(4)面向非確定性問題,概率統計模型在科研中發揮更大的作用;(5)跨學科合作成為主流科研方式,實現前 4 種科研範式的融合,特別是基於第一性原理的模型驅動和數據驅動的融合;(6)科研更加依靠以大模型為特徵的科研大平台,科學研究與工程實現密切結合等。"[16]近幾年,人工智能技術在國際生物、化學、材料、製藥等領域的科學研究中得到廣泛應用,催生了大量的 AI 模型、基礎軟件和應用。在生命科學領域,DeepMind 開發了一款 AlphaFold 程序,採用深度學習技術,破解了蛋白質結構預測這一長期困擾科學界的難題。在數學領域,2017 年以來科學家嘗試使用機器學習、ResNet、seq2seq 模型等技術求解偏微分方程,獲得了更快更準的結果。2021 年,DeepMind 開發了啟發數學家直覺靈感的機器學習框架,幫助數學家和 AI 研究人員在紐結理論方面發現新定理,並在證明 Kazhdan-Lusztig 多項式古老猜想上取得了進展。在物理學方面,2022 年 8 月,物理學家通過使用基於多年實驗數據的神經網絡找到了質子中存在隱性內含粲夸克的證據,這一發現可能會改寫量子色動力學的教科書。在材料科學領域,2016 年《自然》發布了美國哈弗福德學院和普渡大學的研究成果,科研人員利用機器學習算法,用"失敗"的實驗數據預測了新材料合成,這啟示機器學習等 AI 技術成為材料科學的重要研究方式。[17]我國 2017 年啟動"新一代人工智能重大科技項目",開始布局重點領域的相關研究。2022 年 10 月,中國科學技術大學建立數據驅動的 AI 化學家機器人"小來"。目前,"AI 驅動科研範式變革的主要特徵體現在嵌入科研全流程、推動科研設施升級、重構科研人員與儀器設備定位及角色分工、促進科研組織治理模式變革 4 個部分。"[18]中國科學院計算技術研究所就在蛋白質三維結構預測、分子動力學模擬和芯片全自動設計領域,取得三個成功案例。[19]生成式人工智能技術正以驚人的滲透力重塑科研全鏈條,帶來科研管理各環節深層次的變革。

生成式人工智能對哲學、歷史、文學、藝術等人文社科學科的衝擊較重,為學術研究和知識創新帶來前所未有的機遇和挑戰。對人文社科研究範式而言,生成式人工智能展現出了獨特的能力。在哲學領域,生成式人工智能通過分析經典著作、學術論文等海量哲學文本,學習哲學概念的語義網絡、邏輯關聯和論證模式,進而模擬人類的哲學思維過程,能夠獨立完成概念推演。"這些變化不僅體現在人工智能應用會激發出新的哲學問題,如人工智能的道德地位問題、智能駕駛的事故歸責問題、數據偏見與隱私保護問題等。更為重要的是,人工智能的出現和發展本身就會改變甚至顛覆傳統的哲學觀念和視野。在這個意義上,人工智能可以被視為一個'哲學事件'。因此,當我們在人工智能時代問出'人是什麼'的問題時,我們不是要回到人文主義的舊圖景,而是要在人工智能技術的衝擊下重新定義人的獨特性和不可替代性。這將成為人工智能時代的哲學母題。"[20]在歷

史學領域,“歷史學家使用人工智能協助標注史料、查詢筆記、總結文獻主題、史料分類等工作,都是在助力而不是阻礙學術研究。在具體研究實踐展開的環節,人工智能同樣能夠發揮效能,幫助歷史學家更充分利用史料,提高研讀文獻的效率。”[21]生成式人工智能憑藉大規模史料的信息提取與整合、歷史脈絡的關聯性分析與敘事重構、歷史文本的校勘與辨偽、跨地域跨時段的歷史比較研究等文本挖掘與分析能力,正逐漸成為重構歷史認知的重要工具,並還原更接近真相的歷史脈絡。在文學領域,DeepSeek 促進了文學與其他學科的交叉研究。研究者可以將文學與歷史學、社會學、政治學等領域的數據結合在一起,通過 DeepSeek 的分析工具探討更為複雜的人類行為和歷史進程,形成多維度的研究視角,豐富了文學研究的內涵。“DeepSeek 對傳統詩學的賦能,體現在:第一,創作能力加持;第二,重構學術進路;第三,跨媒介傳播。”[22]在藝術領域,“AI 技術早就部分參與、輔助並賦能人文研究了。比如,利用生成對抗網絡技術,藝術家可以生成山水畫、水墨畫和花卉等進行藝術創作,並將傳統中國畫轉換為逼真圖像;利用深度卷積生成對抗網絡技術,研究者可以識別圖像的損壞部分並填充缺失部分,修復文化遺產;利用語音生成技術,研究者可以探索古代語言和音樂的聲音,促進跨文化溝通與理解。”[23]

總之,“DeepSeek 等人工智能的異軍突起突破了技術圈層,直抵人文社科學術體系,驅動著傳統學術研究範式轉向人機協同的創新範式,這一過程引發了知識生產權力的深度遷移,催生了一種新型知識生產方式。構建兼具人文深度與前沿技術的學術新範式已經成為當前人文社科學術研究的新興理論生長點。”[24]

2. 人機協同寫論文

早在 2005 年,美國麻省理工學院的計算機科學與人工智能實驗室的傑里米·斯特里布林等三位研究生聯合編寫了一個叫做 SCIgen 的計算機程序,能夠自動生成英文科技“論文”,包含摘要、引言、文獻綜述、實驗結果、結論、圖片和參考文獻等。此後幾年用這個神器生成的論文,堂而皇之地出現在世界各地的科技學術會議,有的甚至通過了同行評審,被 CSSE 雜誌錄用。2010 年,法國約瑟夫·傅立葉大學的計算機科學家西里爾·拉貝,虛擬了一個叫做 IkeAntkare 的機器人作者,製造了 102 篇機器生成論文來測試“谷歌學術”是否收錄。結果這位機器人成功了,甚至還成為世界上第 21 位被引用次數最高的“科學家”。[25]15 年過去,生成式人工智能“在學術研究中的輔助創新方式是多維度的,貫穿於整個創作流程中,其可能完成的輔助性任務包括翻譯、簡潔呈現研究結果、文本生成、提煉摘要、上下文理解、數據分析等,其在推理、對話和總結方面的突出表現,可充分滿足人們在短時間內低成本獲取密集知識的需求”。[26]機器人輔助寫稿正在成為常態,人文社會科學研究也將面臨生成式人工智能的野蠻敲門。截至 2024 年 10 月,我國完成備案並上線為公眾提供服務的生成式人工智能大模型近 200 個,注冊用戶超 6 億。[27]目前國內流行的 AI 學術工具有很多種,如美國 OpenAI 公司的 ChatGPT、中國深度求索的 DeepSeek、抖音有限公司的豆包、騰訊公司的騰訊元寶、阿里巴巴公司的夸克 Quark 和通義千問、百度公司的文心一言、北京閱之暗面公司的 Kimi、北京智譜華章公司的智譜清言、上海知否知否公司的包閱等,總體呈現出“垂直化、專業化”趨勢。對於人類個體而言,生成式 GAI 就是一個巨大且持續擴充的知識庫,“作為個體的人遠不如匯集了集體訓練成果的機器系統儲存的知識多”。[28]

學術研究的根本目的在於知識生產,就是在總結人類理解宇宙、世界、社會、人生的實踐經驗的基礎上,反思已有的思想、觀點和命題,並賦予已有知識體系以新的思想內涵、時代內涵和文明內涵。隨著算法不斷迭代創新,預訓練模型不斷優化,多模態能力不斷升級,使得社會科學研究也更

加重視計算思維、協同思維、跨學科思維和數據驅動,生成式人工智能正悄然重塑論文寫作模式。

(1)主題構建。傳統論文主題的構建,主要依賴作者的學術經驗與知識儲備。以 CHatGPT、DeepSeek、文心一言、豆包等為代表的大語言模型,在接收到研究者輸入的初步研究方向與關鍵詞之後,能夠快速掃描海量文獻數據庫與核心期刊發文規律,通過對大量學術數據的深度挖掘和分析,可以洞察各學科領域的研究熱點及其變化趨勢,快速列舉出目前存在的關鍵問題和揭示潛在研究空白。研究者通過與 GAI 的多層次、深維度交互探討,可以進一步明確研究的核心問題,甚至實現個性化的主題定製,確保論文主題具有針對性與研究價值。不過,"GPT 生產的知識確實是沒有意向對象的,它形成知識文本的過程是依靠邏輯關係實現的,但人類的參與,卻在一定程度上使得人類與 GPT 一問一答的'雜合文本'有了某種'弱意向性',特別是 GPT 生產的雜合知識文本若經過人類的深度再加工,其意向性的特質就更加突出。只不過,GPT 介入知識生產後,人類也將不得不面對'雜合知識'急遽擴充的局面。"[29]因此,研究者必須深度融入個人學術見解,創造性重構框架,嚴格規範使用流程,才可能有效規避風險。

(2)文獻綜述。傳統文獻綜述,需要研究者手動篩選和閱讀圖書館、數據庫的大量文獻,不僅耗費大量時間與精力,而且容易遺漏重要文獻,導致綜述的完整性與準確性受到影響。生成式人工智能憑藉其強大的自然語言處理能力,能夠理解研究者輸入的自然語言查詢指令,快速在各類學術數據庫中進行精準檢索和篩選。不僅能迅速檢索出大量相關文獻,還能篩選出相關的高影響力研究主題、核心觀點、實驗結果等關鍵信息,並且按照主題、時間、研究方法以及與其他文獻的差異點等維度進行分類整理與歸納總結,極大地提高了文獻篩選的效率。

(3)論文寫作。生成式人工智能能夠依據研究者給定的論文主題、研究框架、關鍵詞、研究方向等信息,在短時間內生成邏輯連貫、結構完整的論文初稿。論文初稿雖然在專業深度和學術創新方面存在缺陷,但它為研究者提供了一個良好的寫作基礎。如果研究者已經完成部分論文內容寫作,但需要對某些觀點進行進一步拓展或深化時,生成式人工智能能夠根據已有的觀點,通過搜索相關文獻,為研究者提供更多支持性證據或不同的學術觀點,幫助研究者從多個角度對研究結果進行深入分析和討論,使論文的內容更加豐富、全面和深入。

(4)觀點提煉。生成式人工智能能夠快速提煉數千篇論文的核心觀點,構建領域研究脈絡;能夠橫向對比多篇論文的觀點,對比發現領域未解決問題;能夠突破學科壁壘,從聚類結果中提取交叉領域的共識與爭議。生成式人工智能對論文觀點的提煉,遵循"解構—提取—聚合—驗證"的閉環流程,通過結構化分析將碎片化文本轉化為系統化知識體系。這一過程不僅依賴文本處理技巧,更需對生成式人工智能的技術本質、應用場景及學術範式有深層理解,最終為學術研究提供清晰的理論參照。

(5)文本修飾。生成式人工智能對論文的文本修飾已形成多維度、系統性的賦能體系,具體體現在語法與用詞規範校準、句式多樣性與語言流暢度優化、文獻引用格式標準化、學術倫理合規提示等方面。不同的學科研究和學術期刊審稿,往往有特定的寫作風格和學術規範要求。生成式人工智能可以根據用戶指定的規範指南,對論文進行全面的格式化和風格調整。它能夠自動設置字體、字號、行距、頁邊距等文檔格式,統一各級標題的樣式;按照要求調整參考文獻的格式和引用方式,甚至模仿特定學者或學術流派的寫作風格,使論文在風格上更加契合目標期刊和研究領域的要求,增強論文的專業性和說服力。總之,生成式人工智能對論文的文本修飾已超越簡單的"糾錯改字",已形成從語言表層到邏輯內核的多層級賦能。

(6)生成結論。生成式人工智能具備強大的知識整合能力,通過分析不同研究成果之間的關聯和差異,能夠幫助研究者從更宏觀的視角審視自己的論文,從而使結論更加系統、客觀。通過識別文章中的核心觀點、重要數據,按照一定的邏輯順序進行歸納整理,形成更具深度和廣度的研究結論,提升論文的學術價值。生成式人工智能的快速突破,使學術研究主體正進化為人機交互的協同模式。“在這一模式中,研究者作為需求提出者和反饋提供者,為研究提供了豐富的語境、價值觀和行為特徵輸入。生成式人工智能不再停留於傳統人機交互模式中輔助者的角色,而是承擔了包括方案制定者在內的多種角色,憑藉其全域知識和預測能力,幫助研究者突破傳統思維局限,為研究者制定更加準確和高效的解決方案。”[30]

綜上所述,AI 技術對學術研究的賦能,體現在科研效率的指數級提升、跨學科研究的融合加速和數據驅動的範式革命。AI 技術對學術研究的魅惑,也會引發研究者認知理解錯覺、學術誠信危機和學術主體性消解等深層倫理挑戰。當 AI 既能撰寫論文又能生成評審意見,傳統的“作者—編輯—讀者”三者的關係面臨重構。過度依賴 AI,可能使研究者淪為“算法操作工”,喪失批判性思維能力。過度依賴 AI,可能使編輯逐漸失去職業核心的“主動性、創造性與人文性”,從“內容的把關者”降格為“技術的跟隨者”。

三、生成式人工智能正重塑學術期刊出版

隨著 ChatGPT、DeepSeek 等生成式人工智能工具的普及,學術不端行為呈現出新特點與新形態。據《自然》雜誌調查,2023 年提交至預印本平台的論文中,疑似 AI 生成的佔比達 12%。德國蒂賓根大學數據科學家德米特里·科巴克領導的團隊,分析了學術數據庫 PubMed 中 2010 年至 2024 年 6 月期間發表的 1,400 萬篇論文摘要。他們估計,2024 年上半年,至少有 10% 的生物醫學論文摘要(約 7.5 萬篇)使用了大語言模型(LLM)進行寫作。[31]生成式人工智能輔助寫作的出現,人們的反應各有不同,恐懼者有之,抵觸者有之,擁抱著更有之。2024 年 11 月復旦大學教務處發布了《復旦大學關於在本科畢業論文(設計)中使用 AI 工具的規定(試行)》。規定中明確提出了“六個禁止”:禁止在研究方案設計、創新性方法設計、論文結構設計等關鍵環節使用 AI 工具;禁止使用 AI 生成或修改原始數據、實驗結果和圖像;禁止直接使用 AI 生成的正文文本、致謝等內容;禁止使用 AI 進行語言潤色和翻譯;禁止評審專家使用 AI 工具進行評審;對於涉及保密內容的論文,禁止使用任何 AI 工具。[32]這一規定,明確劃定了人工智能工具在學術寫作中的使用邊界。此後國內多所高校和學術期刊紛紛出台相關辦法與措施,集體抵制 AI 生成論文,強化審查機制,呼籲守護學術尊嚴,拒絕技術捷徑。但採取的這些防禦措施,做出的各種規定,是否又過於簡單粗暴,潑髒水容易把孩子一起潑出去。生成式人工智能充滿了張力,許多人正在上面做著充滿想像力的探索,不妨“讓子彈再飛一會”,不要匆忙下結論。

(1)提高識別 AI 輔助論文的能力

作為學術期刊,當然要重視生成式人工智能輔助寫作論文中存在的版權與倫理風險,建立必要的使用標準和規制框架。作為學術期刊編輯,更要結合技術原理、專業知識、學術規範和實踐經驗提升識別 AI 生成論文的能力。一是從語言表達特徵分析,AI 生成的論文過度流暢且顯得機械刻板,缺乏人類寫作時的口語化表達和個性化措辭。由於對語義的理解存在局限性,AI 生成論文會出現詞彙搭配不合理或語義前後矛盾的情況。二是從內容邏輯與專業深度考量,AI 生成論文只是對已有知識的簡單拼湊與複述,缺乏專業作者對學科前沿問題的深入思考與獨特見解。此外,AI

生成論文在論證時會出現邏輯跳躍,論據與論點之間缺乏緊密聯繫,難以形成完整的邏輯鏈條。三是從格式與排版細節觀察,AI 生成論文存在標題格式不統一、數據與文字描述不符、圖表類型選擇不當、參考文獻不符合規範等。四是數據分析表面化且缺乏情感與人文關懷。AI 能生成合理的圖表和數據,但分析往往停留在描述層面。對比發現,人類作者即使處理簡單數據,也會嘗試挖掘潛在模式或提出解釋機制,而 AI 生成的分析常止步於數據表面特徵的陳述。五是結論部分泛泛而談且缺乏針對性。結論未能緊扣研究成果在實際應用中的可行性、局限性以及與其他相關技術的比較優勢,沒有深入挖掘和討論研究過程中發現的問題和潛在的改進方向,僅說了一些籠統空洞的大話,沒有提出針對性強的對策建議和具體的優化進路,缺乏實際指導意義。六是參考文獻引用不規範。首先是引用文獻與討論問題關聯性弱,雖然引用的是高被引作者和高被引論文,但與正文討論內容匹配度很低。其次是存在"幽靈引用"現象,引用的參考文獻實際出版物中並不存在。再次參考文獻引用存在異常片面特徵,有的重要基礎性文獻集體缺失,有的重要文獻引用比例過高,跨學科引用幾乎為零。七是有時會"一本正經地胡說八道"。生造晦澀術語、堆砌流行概念,將簡單思想包裹成高深理論。比如"ChatGPT 給出的綜述性文獻,雖然很多都不存在,但看起來卻能'無懈可擊';它輸出的代碼,雖然大多無法使用,卻總能給人'一本正經'的規範感。可以說,ChatGPT 在對社會、金融、教育等學科的專業性問題提出建議時,幾乎總顯得'真理在握',即便是應對古典學科的某些'冷門絕學'問題,它也敢大開腦洞地'瞎編'。"[33]

(2)人機協作將成為一種常態

在生成式人工智能的應用場景下,學術出版已不再局限於人類內容創作的單一維度,而是向人機協作模式延展;不再局限於傳統的期刊與圖書,還包括視頻、音頻、增強現實(AR)和虛擬現實(VR)等多種媒介形式。"學術編輯的定義可以擴展為:基於自身的知識積澱和業務經驗積累,合理選用人工智能工具,挖掘發現內容價值,依據學術出版規範對內容進行整理、加工、製作、發行,向讀者和用戶提供知識服務的行為者。"[34]傳統學術期刊編輯僅僅滿足於會策劃、會組稿、會審稿、會改稿、會校對,顯然已經無法適應生成式人工智能應用場景下的崗位要求。單純的內容加工者轉變為多模態內容的整合者和人機協作的管理者,將成為學術期刊編輯的核心競爭力。一個人機協作的編輯,應當是學術前沿的追蹤者,應當是問題導向的倡導者,應當是內容質量的把關者,應當是人文價值的發現者,應當是多模態內容的整合者,應當是內容創新的策劃者,應當是平台化傳播的推動者。"學術期刊的平台化生存,不僅關乎編輯在虛擬空間中的存在,也關乎編輯在現實空間裡的生存。因此學術期刊要努力實現從'數字化'到'數智化'、從'內容為王'到'創新為要'、從'單一傳播'向'複合傳播'、從'紙媒期刊'向'媒體平台'的轉型。"[35]從一定意義上講,"編輯這一行當的日常,在今後或許將逐漸演變成與 AI 的一次次合作。我們需要逐漸習慣使用 AI,並掌握與之共同完成創作的本領。"[36]人機協作可以讓 AI 技術在學術期刊選題、審稿、評審、編輯加工、出版發行等各個環節發揮獨特作用,與編輯的專業判斷形成互補,從而推動學術期刊向更加智能化與專業化的方向邁進。

第一,選題策劃。生成式人工智能能夠顯著增強熱點選題的追蹤效率。通過對海量學術文獻、社交媒體數據以及科研動態信息的實時監測與分析,生成式人工智能可以快速識別出當前學術界的研究熱點和潛在趨勢。如中國科學院文獻情報中心開發的"學術之眼"系統,旨在通過數據挖掘、知識圖譜和人工智能技術,幫助編輯用戶高效獲取學術資源、分析研究趨勢、發現潛在作者等,通過深度學習數百萬篇已發表論文的語料庫,能夠提供有價值的選題參考,極大提升了選題策劃的

科學性、時效性和匹配作者的準確性。

第二,稿件審讀。學術期刊的傳統工作流程,編輯要面對大量投稿,憑藉專業知識與經驗篩選出具有學術價值與創新性的優質稿件,審稿過程耗時費力且主觀性比較強。生成式人工智能輔助審稿主要體現在三個方面:一是論文格式規範。利用自然語言處理技術,能夠快速檢查論文的參考文獻格式、圖表排版格式等是否規範。二是學術不端檢測。通過學術文獻數據庫進行對比,能夠檢測稿件中潛在的學術不端風險。三是作者學術水平評估。基於學術數據庫統計,對作者的研究方向、過往發表成果、引用情況等進行分析,生成作者學術水平評估報告。編輯可據此瞭解作者在相關領域的研究能力和影響力,輔助判斷稿件的學術價值,提高審稿效率和準確性。

第三,同行評審。在同行評審環節,生成式人工智能可根據稿件的學科領域、研究方向、關鍵詞等信息,從龐大的專家數據庫中快速匹配出合適的評審專家。同時,它還能對專家過往評審記錄進行分析,評估專家的評審風格、評審效率以及評審質量,為編輯提供更全面的參考,大大縮短了專家邀請週期,提高了同行評審的效率與質量。不過,"生成式人工智能可有效提升學術評價效率,但不宜取代人類評價;可作為有效輔助工具,但必須以人類專家評價為主導;學術評價因技術發展而推陳出新,但評價的人文屬性不變。"[37]

第四,編輯加工。在編輯加工環節,生成式人工智能能在語言潤色、邏輯結構優化等方面提供有力支持。它可以檢查稿件中的語法錯誤、拼寫錯誤,對語句進行通順性調整,還能根據學術寫作規範,對論文的結構進行分析,提出優化建議,使論文邏輯更加清晰。編輯在此基礎上,對稿件的語言規範、邏輯結構、圖表格式、觀點論證等進行審核完善,大大縮短了編輯加工時間,提升了稿件質量。生成式人工智能已成為編輯加工不可替代的輔助工具,但仍需與人類編輯的專業判斷相結合——AI 負責"標準化、重複性"工作,人類則聚焦"創新性、人文性"內容的把控。

第五,內容校對。生成式人工智能在學術論文校對中的核心價值,首先體現在語言規範性校驗的高效性,可以自動檢測語法錯誤、優化句子結構,提升表達流暢度。其次格式與規範的標準化控制,按期刊要求自動調整參考文獻格式、檢測圖表標注、公式編號、段落縮進等格式的一致性。再次學術嚴謹性的輔助審查,可以識別邏輯斷層、數據表述一致性、預警潛在學術不端等。一些先進的校對軟件利用深度學習技術,能夠學習不同學科領域的語言風格和專業術語,從而實現更精準的校對。總之,"人機協同並不是指機器完全取代人類,而是指人與機器之間相互配合、互補的關係。在人機協同中,機器的優勢在於高速計算、大數據處理和精確性方面,而人類則具有創造性思維、靈活性和情感等優勢。因此,人機協同的真正價值在於充分發揮人與機器各自的優勢,從而實現更高效、更智能的工作和生活方式。"[38]

四、生成式人工智能帶來的版權與倫理風險

生成式人工智能近年來取得的突破性進展,為諸多行業帶來了革命性的影響。不過,硬幣通常都有兩面。在學術出版領域,隨著生成式人工智能應用場景不斷拓展,在給學者、編輯帶來便捷與新奇體驗的同時,一種"AI 附魅"現象悄然興起。人們將 AI 賦予了超越其技術本質的魔力光環,甚至將其神化。美國計算機科學家摩西・瓦迪在 2012 年發表觀點認為,"當前人工智能的發展已經很快,所有人類勞動力都將在未來 30 年內被廢棄。現在,機器正在與人類的大腦角力。機器人兼具大腦和肌肉。我們都正在面對'被我們的造物完全取代'的未來。"[39]中國學者趙汀陽 2022 年曾明確表示,"人工智能試圖改變智能的本質,這是要創造一種新的存在,所以是一個存在論級別的

革命”,而“模仿了人性和人類價值觀的人工智能就和人類一樣危險”,它們會試圖“自主建立遊戲規則”,“因此安全的人工智能必須限制在沒有反思能力的圖靈機水平上”。[40]近年來,新聞媒體更是對 ChatGPT、DeepSeekGPT 等生成式人工智能給於過度解讀與渲染,常冠以“顛覆性革命”或“行業終結者”、“突破人類極限”或“超越人類的文化理解”等附魅之詞,不僅影響公眾對 AI 的客觀認知,更可能對學術研究和學術期刊出版帶來版權與倫理風險。

1. 機器學習使用訓練數據侵權

機器學習需要海量數據進行訓練優化算法、提升性能,如 GPT—3 語言模型就需要挖掘 1750 億個參數加以支持。這些數據涵蓋文本、圖像、音頻、視頻等多種形式,來源廣泛,包括互聯網上的公開數據、用戶生成內容以及企業內部數據等。在數據收集和使用過程中,由於相關法律法規尚不完善,部分企業和開發者為追求效率和降低成本,未經授權擅自使用受版權保護的作品、侵犯個人隱私數據或違反數據使用協議,導致訓練數據侵權事件頻發。從法律層面看,不同國家和地區的法律規定與司法實踐存在很大差異。歐盟在《數字化單一市場版權指令》(2019)中創設了《文本與數據挖掘例外條款》,允許科研機構和文化遺產機構在未經著作權人授權的情況下使用作品進行 AI 訓練,但要求權利人可通過“選擇—退出”機制保留控制權。美國尚未通過專門立法確立 AI 訓練的合理使用地位,司法實踐傾向於嚴格解釋。在《湯森路透訴羅斯智能案》中,法院明確駁回了 AI 訓練適用合理使用的抗辯,認定未經許可使用 Westlaw 數據庫內容構成侵權。我國《著作權法》第十條規定:複製權是以印刷、複印、拓印、錄音、錄像、翻拍、數字化等方式將作品製作一份或者多份的權利。機器人在挖掘數據的過程中,無論是將這些數據讀入系統還是進行格式轉換和數據分析,均涉及受著作權人控制的複製行為。[41]2023 年 8 月國家互聯網信息辦公室發布《生成式人工智能服務管理暫行辦法》第七條要求,生成式人工智能服務提供者應當依法開展預訓練、優化訓練,增強訓練數據的真實性、準確性、客觀性和多樣性,使用具有合法來源的數據和基礎模型。涉及知識產權的,不得侵害他人依法享有的知識產權。涉及個人信息的,應當取得個人同意或者符合法律、行政法規規定的其他情形。這個《暫行辦法》僅提出原則性要求,具體如何界定侵權仍需深入探討。生成式人工智能使用多種來源的海量數據、知識作為訓練數據,但這種直接或間接“搬運”,掩蓋了原創內容引用的嚴謹性,會使作者與學術期刊編輯誤將模型輸出等同於對現象的真正理解,忽略添加注釋和參考文獻,結果導致涉嫌抄襲或學術不端。

2. 人工智能生成物是否具有可版權性

在人工智能技術蓬勃發展的當下,AI 生成的繪畫、論文、小說、音樂等作品不斷湧現,其是否具有可版權性成為法律界和學術界熱議的焦點,主要形成“工具說”與“無版權說”兩種對立觀點。“工具說”強調人類對工具的主導性。AI 生成物當以人為本,不能以機器為本。因為 AI 生成物的數據準備、特徵提取、算法規則、模型訓練、標籤應用、效果優化等都是由人工完成的。“比工具更重要的是設計工具、製造工具、使用工具的人。”[42]因此,AI 生成物的人類創作者可基於其貢獻獲得版權保護,AI 僅被視為高級創作工具。根據《伯爾尼公約》及多數國家版權法,作品要獲得版權保護,通常需滿足兩個條件:一是原創性,作品需體現作者的獨立創作,而非簡單複製。二是人類作者,傳統版權法默認作品需由人類創作。英國 1988 年發布《版權、外觀設計和專利法》,將計算機生成作品的著作權歸“創作作品進行必要安排的人”所有。“必要安排的人”被視為作者,可能包括編寫人工智能的程序員、人工智能系統或設備投資者、人工智能的最終使用者等主體。[43]2023 年美國版權局裁定美國藝術家卡什塔諾娃創作的科幻漫畫書 *Zarya of the Dawn* ,其中“作者創作的文字

和對人工智能生成的作品的選擇、協調和安排”可獲部分版權,但通過 Midjourney 生成的圖片不受著作權保護。“無版權說”強調“作者必須是自然人”原則。基於這一標準,作品中包含的人工智能成果,統統被排除在版權保護之外。美國版權局 2023 年 3 月發布的《版權登記指南:含有人工智能生成材料的作品》指出,在面臨涉及人工智能的成果時,需要區分:一種是人類主導、機器輔助的產物,這種客體具有可版權性;一種是主要由機器進行構思與落實的產物,這種客體不具有可版權性。“在確認人工智能生成物的可版權性時,人類作者與創造性,是其中兩個最核心的元素。為此,《版權指南》要求申請人在提交注冊人工智能作品時,必須解釋人類在其中的創造性貢獻,以方便識別版權類型。”[44]中國案例為 2023 年 8 月“李某與劉某侵害著作權糾紛案”,原告李某起訴稱其使用開源軟件 Stable Diffusion 生成涉案圖片並發布在小紅書平台,後發現被告劉某在百家號文章中使用該圖片且截去署名水印。北京互聯網法院經審理認為,涉案圖片生成過程中原告進行了較多智力投入,如設計人物呈現方式、布局構圖、選擇提示詞、重複調整參數等行為,體現了原告的獨特審美、個性化表達,具備“獨創性”要件,屬於美術作品,原告享有著作權,被告行為構成侵權。“該案將文生圖作品定性為人類智力成果,扭轉了社會評價中人工智能機械式生成和人類無法控制輸出內容的偏見,也蘊含了對於不同 AIGC 成果施以分級版權保護設計的內涵。”[45]

3. 生成式人工智能對人格權的侵害

人格權是民事主體享有的生命權、身體權、健康權、姓名權、名稱權、肖像權、名譽權、榮譽權、隱私權等權利,是民事主體維護其人格尊嚴和人格自由發展的重要權利基礎。而生成式人工智能的運行機制和應用模式,卻在多個維度對人格權構成了威脅。首先,直接侵害個人隱私和信息。未經授權生成高度逼真的個人肖像、聲音,直接盜用或扭曲他人身份。冒用他人姓名生成虛假言論或行為,導致身份混淆。泄露學者科研成果數據、學術交流隱私、未發表作品內容、個人聯繫方式,曝光讀者閱讀偏好、購買記錄、個人身份信息等,都嚴重侵害了他人的隱私權和個人信息權益。其次,人格尊嚴與名譽損害。生成虛假負面信息,貶損他人社會評價。編造關於學者的虛假學術成果、學術不端行為等,一旦這些虛假信息傳播開來,會嚴重損害學者的學術聲譽,對其長期積累的學術成果和人格尊嚴造成極大傷害。利用生成式人工智能生成與作者風格相似的低俗、違法作品,並將其歸為作者名下,會嚴重玷污作者的創作聲譽。再次,侵害肖像權。“利用生成式人工智能自動生成的圖像,比深度偽造造成的損害後果更加嚴重,其不僅會侵害個人的肖像權,甚至可能透過圖片、視頻扭曲個人的形象,特別是當生成的圖片涉嫌性騷擾、猥褻、非法同居等虛假信息後,將會造成對他人的名譽權、隱私權等權益的嚴重侵害。”[46]

4. 存在 AI 幻覺現象

生成式人工智能存在“生成與特定來源相關但無意義或不真實的內容的機器幻覺現象,可能導致危險的發生。例如對藥品說明書翻譯錯誤會產生嚴重後果”。[47]AI 幻覺林林總總,大致可以歸納為四種主要類型。一是語義層面的幻覺。模型在回答涉及具體數據、歷史事件或科學原理的問題時,可能編造不存在的信息;在推理過程中,出現邏輯矛盾、缺乏因果關係的表述。二是感知層面的幻覺。生成式人工智能在處理視覺、聽覺等感知任務時,對輸入信息產生與實際不符的錯誤理解或解讀,形成類似人類幻覺的認知誤判。三是生成式幻覺。生成式人工智能模型在生成內容時,由於數據中的噪聲、錯誤標注或不明確的信息,可能使模型在學習過程中產生混淆,從而在生成內容時出現不準確或虛假內容的情況。四是認知層面的偏見。生成式人工智能在多元行業和領域中大放異彩的同時,伴隨而來的偏見問題也變得日益嚴峻。“人設遊戲中的信任誘導陷阱、人機交往中

群性迷失陷阱、公共實踐中的合法歧視陷阱”以及“技術大眾化風險、媒介私密化風險、後真相迂迴的挑戰”等風險也日漸凸顯”。[48]英國科技分析師本尼迪克特・埃文斯“將 ChatGPT 描述為一個自信的扯淡的傢伙,可以寫出非常有說服力的廢話。其很大的危險在於 ChatGPT 是錯誤的或有偏見的,但聽起來卻像是正確的和權威的。”[49]中國 360 集團創始人周鴻禕認為,能否胡說八道,恰恰是智能的分水嶺,將來很多新的 GPT 大腦在某種程度上要保留這種幻覺的能力。[50]硬幣都有兩面,並不是所有的人都能夠找到正面。兩位的觀察不約而同地犀利、獨立,既可見對事物本質的深刻洞察,也能感受到字裡行間那份不迎合、不盲從的思考。這樣的判斷,不管是對是錯?足以引發更多人對問題的重新審視。

人工智能技術的狂飆突進和知識生產的人機協作,似乎不可阻擋,正極大挑戰我們的想像力,也極大衝擊現有高度學科化的知識體系和學術傳播秩序。學術期刊無疑再次置身於這場變革和挑戰的風口浪尖。學術期刊的“祛魅”,始於技術工具理性的反思。《自然》雜誌率先建立透明化機制,要求作者明確標注論文中 AI 參與的環節,將技術工具從幕後推向台前。德國《科學計量學》期刊則讓審稿人同時評審人類論文與 AI 生成論文,在實戰中提升專業群體的鑒別力。實踐證明破除 AI 附魅,需要制度約束與認知覺醒的雙重驅動。學術期刊的“祛魅”,終於價值重構:回歸學術共同體的精神內核。撕掉 AI 附魅的華麗外衣,我們會發現學術的本質從未改變——它是對知識的創造性生產,亦是思想傳承的橋梁,更承擔著不可推卸的社會責任。因此,學術期刊“應當保持審慎的認知,即不盲目地媚俗技術,也不因技術的強大而產生畏懼,而是應當在技術的膨脹中,堅守人的主體性,在與技術的良性互動中共同實現人機關係的平衡”。[51]

結　語

生成式人工智能對學術期刊的衝擊,本質上是一場關於“知識生產權力”的重新分配。當 ChatGPT、DeepSeek 能在數秒內完成文獻綜述,當 Sora 可將研究成果轉化為可視化視頻,學術期刊的角色正從“內容提供者”轉向人機協作的“質量把關者”與“價值整合者”——編輯需以學術洞察力識別 AI 發現和提出問題的真偽,以學術概括力分析 AI 總結、凝練問題的優劣,以學術想像力判斷 AI 重組和創新問題的良莠,以學術思辨力考察 AI 闡釋和論證問題的是非。在選題策劃中融合 AI 的熱點預測與人類的問題意識,在同行評審中結合算法匹配合適的審稿專家,在學術傳播中平衡多模態形式與學術深度。這場變革的深層啟示,在於技術理性與人文價值的辯證統一。基辛格曾預言人工智能將“改變人類理解現實的方式”,而這種改變的終極意義,不應是機器對人類的替代,而是通過技術賦能釋放人類的創造力。在生成式人工智能重構學術生態的進程中,我們既要擁抱“第五科研範式”帶來的效率革新,也要堅守學術研究的本質——對真理的探索、對思想的創新、對人文精神的傳承。唯有如此,才能在智能革命的浪潮中,構建一個技術與人文相互滋養的學術新生態,讓生成式人工智能成為推動知識進步的工具,而非消解學術價值的威脅。

斯蒂格勒認為,技術是用來彌補人類的原初性缺陷的,愛比米修斯在分配“屬性”時,忘記了給人留下一個屬性,所以人必須依靠技術來使得自身存在。[52]隨著人機對話不斷走向人機交往,人機融合程度勢必不斷加深。不過即使人工智能這樣的“技術神話”,也永遠無法複製完整的人性,永遠不能代替人類進行是非善惡的判斷。“人機交往作為一種理想化的未來人機形態,將人人主體交往的問題擴展到人與非人交往的、更為廣闊的層面,這或許能夠一定程度上減少人類之間的摩擦和內捲,引導人們關注機器和其他非人客體對於人類的影響。”[53]作為 AI 輔助寫作的擁護派,如何

在技術賦能中守護學術本真,在範式變革中堅守人文內核,不僅是學術期刊面臨的時代命題,更是整個知識生產領域需要共同書寫的答卷。

學術期刊與人工智能融合的趨勢已經無法阻擋,誰也無法關起門來營造自己的圍城。融合,或許不是主動邁出的步伐,而是時代浪潮下必須接納的現實。不過穿越視見看出遼闊,何嘗不是另一番景色?

---

①騰訊研究院、中國信息通信研究院等:《人工智能》,北京:中國人民大學出版社,2017 年,第 4 頁。

②基辛格:《人工智能時代與人類未來》,北京:中信出版集團股份有限公司,2023 年,第 16 頁。

③李彥宏:《智能革命——迎接人工智能時代的社會、經濟與文化變革》,北京:中信出版集團股份有限公司,2017 年,第 5 頁。

④陳自富:《煉金術與人工智能:休伯德・德雷福斯對人工智能發展的影響》,濟南:《科學與管理》,2015 年第 4 期。

⑤李開復、王詠剛:《人工智能》,北京:文化發展出版社,2017 年,第 46 頁。

⑥馬丁・福特:《機器人時代——技術、工作與經濟的未來》,王吉美、牛筱萌譯,北京:中信出版集團股份有限公司,2015 年,中文版序 XXIII。

⑦張耀銘、張路曦:《人工智能:人類命運的天使抑或魔鬼——兼論新技術與青年發展》,北京:《中國青年社會科學》,2019 年第 1 期。

⑧王文:《生成式人工智能時代人文學科的機遇與使命》,北京:《數字人文》,2024 年第 3 期。

⑨㉚馬費成、陳帥樸:《生成式人工智能賦能哲學社會科學研究》,武漢:《武漢大學學報》,2024 年第 6 期。

⑩付長珍:《智能與智慧:人機關係的倫理前瞻》,哈爾濱:《求是學刊》,2024 年第 6 期。

⑪陳保亞、陳樾:《人類語言習得的親知還原模式——從 ChatGPT 的言知還原模式說起》,北京:《北京大學學報》,2024 年第 2 期。

⑫張愛軍、郭鎮毓:《生成式人工智能的技術神話、祛魅與認知進路——以 Sora 模型為例》,合肥:《理論建設》,2025 年第 1 期。

⑬王春輝:《自然人、機器人、數字人"三人"共生時代的語言生活》,北京:《語言戰略研究》,2024 年第 3 期。

⑭托馬斯・庫恩:《科學革命的結構》,金吾倫、胡新和譯,北京:北京大學出版社,2012 年,第 147 頁。

⑮董波:《探索哲學社會科學研究的新範式》,北京:《中國社會科學報》,2024 年 5 月 6 日。

⑯⑲李國傑:《智能化科研(AI4R):第五科研範式》,北京:《中國科學院院刊》,2024 年第 1 期。

⑰王飛躍、繆青海:《人工智能驅動的科學研究新範式:從 AI4S 到智能科學》,北京:《中國科學院院刊》,2023 年第 4 期。

⑱余江、張越等:《人工智能驅動的科研新範式及學科應用研究》,北京:《中國科學院院刊》,2025 年第 2 期。

⑳魏犇群、王田等:《人工智能時代的哲學思考》,北京:《光明日報》,2025 年 5 月 12 日。

㉑王濤:《生成式人工智能之於歷史研究的機遇與挑戰》,北京:《史學理論研究》,2024 年第 3 期。

㉒胡曉明:《DeepSeek 與人機同題比較詩學》,長春:《社會科學戰線》,2025 年第 4 期。

㉓曾軍:《從數字人文到 AI 人文:人文研究範式的變革》,福州:《東南學術》,2024 年第 4 期。

㉔姜秀敏、張嘉印:《DeepSeek 介入人文社科學術體系催生的"智能增強學術範式"》,北京:《北京行政學院學報》,2025 年第 3 期。

㉕張耀銘:《人工智能驅動的人文社會科學研究轉型》,濟南:《濟南大學學報》,2019 年第 4 期。

㉖張萌、朱鴻軍:《知識暗流的合規實踐:ChatGPT 在學術出版中的應用與挑戰》,北京:《科技與出版》,2023 年第 5 期。

㉗方經綸:《工信部:我國生成式人工智能服務大模型的注冊用戶超 6 億》,人民網,2024-10-23,http://finance.people.com.cn/n1/2024/1023/c1004-40345610.html

㉘肖峰:《生成式人工智能介入知識生產的功用探析——藉助 ChatGPT 和"文心一言"探究數字勞動的體驗》,重慶:《重慶郵電大學學報》,2023 年第 4 期。

㉙姜華:《從辛棄疾到 GPT:人工智能對人類知識生產

格局的重塑及其效應》,南京:《南京社會科學》,2023 年第 4 期。
㉛劉霞:《生成式 AI 學術使用亟須關注》,國家自然科學基金委,2024-08-08, https://www.nsfc.gov.cn/publish/portal0/tab434/info93343.htm
㉜《復旦發文規範畢業論文 AI 使用》,光明網,2024-11-29, https://baijiahao.baidu.com/s?id=1817016770187946607&wfr=spider&for=pc
㉝曾建華:《人工智能與人文學術範式革命》,北京:《北京師範大學學報》,2023 年第 4 期。
㉞謝壽光、王譽梓:《生成式人工智能應用場景下的學術出版和學術編輯:挑戰與機遇》,北京:《科技與出版》,2025 年第 1 期。
㉟張耀銘:《數智時代學術期刊的平台化生存》,南京:《南京大學學報》,2024 年第 6 期。
㊱瓦當、李敬澤等:《AI 時代的文學書寫——何以純粹?如何現代?》,上海:《文學報》,2025 年 4 月 25 日。
㊲葉繼元、郭衛兵:《生成式人工智能參與學術評價的反思》,北京:《中國社會科學評價》, 2024 年第 1 期。
㊳劉偉、孫惟一:《人機協同前沿問題及未來發展趨勢》,吉林延吉:《延邊大學學報》,2025 年第 1 期。
㊴約翰・馬爾科夫:《與機器人共舞——人工智能時代的大未來》,郭雪譯,杭州:浙江人民出版社,2015 年版,第 86 頁。
㊵趙汀陽:《人工智能會是一個要命的問題嗎?》,廣州:《開放時代》,2022 年第 6 期。
㊶吳高:《人工智能時代文本與數據挖掘合理使用規則設計研究》,北京:《圖書情報工作》,2021 年第 22 期。
㊷劉漢俊:《"人工智能+傳播"如何創新》,北京:《中國出版》,2025 年第 1 期。
㊸叢立先、李泳霖:《聊天機器人生成內容的版權風險及其治理——以 ChatGPT 的應用場景為視角》,北京:《中國出版》,2023 年第 5 期。
㊹葉娟麗:《概念視角的版權重構與變遷——兼談人工智能生成物的可版權性》,澳門:《澳門理工學報》,2024 年第 1 期。
㊺姬艾佟:《人工智能生成內容版權保護與證明要素》,北京:《中國出版》,2025 年第 8 期。
㊻王利明:《生成式人工智能侵權的法律應對》,北京:《中國應用法學》,2023 年第 5 期。
㊼姜澧格、于海防:《生成式人工智能的法律規制——從技術與人機關係出發》,哈爾濱:《求是學刊》,2025 年第 2 期。
㊽羅曉東、趙宏瑞:《從應用陷阱到防治壁壘:人工智能風險的兩重性認識及機制應對》,武漢:《湖北社會科學》,2024 年第 7 期。
㊾胡泳:《當 ChatGPT 產生幻覺,一個"幻覺時代"要來臨了》, 2023-04-05, https://baijiahao.baidu.com/s?id=1762284989739915299&wfr=spider&for=pc。
㊿周鴻禕:《GPT 有四個不可解釋的現象》,圖靈社區中國信息界,2023-08-16, https://mp.weixin.qq.com/s?__biz=MzIzNzA2NTY2NA==&mid=2651231196&idx=1&sn=b401d15560dcc36020fda7a5c2f94c14&chksm=f33c7773c44bfe6569a316c63b10a4c7421d4cb7217a5a3f5eef8622146bfb08007f3016f7aa&scene=27。
(51)張愛軍、郭鎮毓:《生成式人工智能的技術神話、祛魅與認知進路——以 Sora 模型為例》,合肥:《理論建設》,2025 年第 1 期。
(52) Bernard Stiegler, *Technics and Time, The Fault of Epimetheus*, trans. by Richard Beardsworth and George Collins, California: Stanford University Press, 1998, p. 188.
(53)陶鋒、劉星辰:《從人機對話到人機交往——人工智能大語言模型的哲學反思》,長春:《社會科學戰線》,2024 年第 5 期。

作者簡介:張耀銘,《新華文摘》原總編輯、編審。北京 100706

[責任編輯 劉澤生]