AI時代的事實查核與媒體識讀

陳偉婷

台灣事實查核中心總編輯

壹、AI浪潮下的資訊生態位移

在生成式人工智慧（Generative AI）技術普及的當代，全球資訊生態正經歷一場前所未有的結構性轉移。AI生成文字、圖片、影音等內容的能力大幅提升，使資訊生產成本驟降，資訊流通速度空前加快。然而，在便利性提升的同時，也帶來嚴重的資訊過載與資訊生態混亂。

大眾獲取時事資訊的行為模式，已從傳統媒體時代的「主動搜尋」演進為基於社群平臺（如LINE、Facebook、YouTube、TikTok）演算法推播的「被動接收」。這種演進雖然加速了資訊流動，卻也增加民眾資訊識讀的風險。

尤其AI生成工具的盛行，當具備煽動性的虛假內容能夠大量生成並精準透過演算法投放到特定族群時，個體的理性判斷空間將受到壓縮。當公眾難以辨別真實與虛構，社會信任基礎便可能受到侵蝕，而事實查核工作也面臨前所未有的挑戰。

貳、雙引擎策略的協同機制

當代假訊息的運作已轉向規模化的系統性操作，其中最常見的運作模式，是透過「受眾捕獲」與「極化敘事」的雙引擎策略，進行長期的認知建構。

「受眾捕獲引擎」採取生活化包裝，大量頻道以年長者為目標族群，以養生健康、生活政策、投資理財等民生資訊作為切入點，每日大量產製影片，透過「年齡限定」、「比

例恐嚇」與「損失警告」等手法，刺激受眾的焦慮心理。例如，針對65歲以上族群散播「高鐵臺鐵隱藏優惠」或「政府發放萬元紅包」等未經證實的政策福利，藉此提高觀看率與建立高黏著度的訂閱依賴。

在這種頻道與影片的批量化生產模式下，假訊息得以在短時間內形成覆蓋式傳播，其擴散速度與量能已非傳統、高度依賴人工的事實查核機制所能即時阻斷。

其次是「極化敘事」策略。部分頻道慣於製作誇張甚至荒誕的故事，如「阿拉伯王室來臺就醫愛上臺灣人」等誇大的敘事，讓受眾長期沉浸於此類內容環境後，原本看似無害的民生頻道便順勢植入地緣政治的恐慌或對立，轉變為散播焦慮與傳達特定政治立場的管道之一。而支撐此一龐大體系的技術核心，正是AI生成內容所帶來的低成本、高效率與規模化內容生產能力，使虛假資訊得以突破過往的人力限制，形成資訊工業化的新型態。

參、AI深偽技術操弄的系統特徵

支撐上述雙引擎策略持續運作的技術核心，正是生成式AI與深偽技術（DeepFake）的快速發展。當內容造假的技術門檻持續降低，而辨識真偽的難度不斷提高時，虛假內容能以更低成本、更高效率地大量生產，進一步加劇事實查核投入與產出之間的失衡，引發社會對資訊真實性的普遍懷疑與真假難辨的集體焦慮。

當前AI影音頻道呈現顯著的系統化操作特徵，包含AI身分複用。眾多活躍頻道共用相同的AI虛擬主播面孔，卻以不同名稱經營。技術上，操作手段已從早期利用戴口罩來掩飾嘴型對齊瑕疵的粗糙策略，逐漸進化為聲音、語調乃至手勢肢體動作皆能完美協作的高度仿真模型。這種系統化操作同時具備顯著的集體特徵，表現為帳號間同步休眠、統一的紅黃色警告排版，以及主題在養生、色情與政治敘事間的無過渡切換。

此外，AI深偽技術也廣泛被運用於醫療詐騙與名人冒用案件。常見手法包括盜用醫師公開演講畫面，變造醫師聲音與影像的AI醫療詐騙影片。例如，冒用臺大內科醫師影像製作的「一週減重 10公斤」廣告，實則導向一頁式詐騙網站[1]；或挪用臺大外科部主任

1 台灣事實查核中心（2024年12月2日）網傳影片「台大內科教授張上淳在網路販售減重產品」？。取自https://tfc-taiwan.org.tw/fact-check-reports/ntu-professor-zhang-shangchun-weight-loss-product-scam/

的演講畫面，以深偽變造唇形與聲音合成的「2025年肺病新藥」虛假推廣廣告[2]。這類影片不僅危害公眾健康與社會信任，更衍生造成被仿冒者困擾，使專家因擔憂影像被變造而減少公開發聲，導致真實的資訊管道日益限縮。

假訊息亦常鎖定高爭議議題。例如曾有訊息利用兩岸政治敏感事件，編造「臺灣器官將流向中國黑市產業鏈」等說法，導致民眾大規模撤簽器官捐贈同意書。此類案例顯示，假訊息不僅影響個人判斷，更可能對公共政策與社會信任造成實質衝擊。

肆、媒體在假訊息產業鏈中的「特殊角色」

傳播學的「第三人效應」指出，多數人傾向認為「自己受假訊息影響不大，他人較容易受騙」，這種對自身判斷力的過度自信，反而削弱了個體對假訊息的警覺，此現象在流量至上的當代尤為顯著。

在此背景下，部分媒體為追求點閱率出現內容農場現象，使得假訊息得以藉此透過跨境素材的挪用進行誤導。以2024年山陀兒颱風期間的報導為例，媒體未經核實即轉載LINE爆料群組中民眾合成之影像，報導所謂「空中漂浮內褲」奇景。然而透過「以圖反搜」技術，即可證實該影像原始發生地為海南島而非臺灣。又如將印度山洪暴發影片指稱為2025年花蓮馬太鞍堰塞湖潰壩影像，此類敘事結合地方政治敏感性，易激化社會對立。若能透過數位查核工具釐清影像原始脈絡，或能有效降低資訊干擾，從源頭破解認知誤導。

此種資訊干擾不限於影像的挪用，更包含境外論述的轉譯與再製。以2023年「美國毀滅臺灣計畫」假訊息為例，該訊息源於美國華府廣播節目主持人Garland Nixon在X平臺的諷刺性貼文，內容以戲謔方式影射美國對臺政策，經由中介者（如政治人物或意見領袖）的翻譯與加工，最終被臺灣媒體與政論節目在未經查證訊息的情況下轉載與報導，引發輿論震盪。

2 台灣事實查核中心（2025年11月5日）。網傳影片「台大名醫研發肺病神藥，7天逆轉肺纖維化」？。取自https://tfc-taiwan.org.tw/fact-check-reports/ai-deepfake-uses-ntu-doctor-name-selling-ineffective-lung-treatment/

「認為他人較易受騙」的心理偏誤，同樣顯現在年齡層的分佈上。儘管社會主流習慣將長輩視為假訊息的主要受害者，但警政署警政通報統計[3]指出，年輕族群（以青壯年18至49歲居多）反而是受騙頻率較高的群體。以2025年數據為例，該年齡層被害人計14萬6,757人，占整體受害者比例高達 73.48%，惟或因單次財損較小而未獲重視。各年齡被害人中，「未滿18歲」及「18至23歲」以「假網路拍賣（購物）」占比最高，24歲以上各年齡層被害人，以「投資詐欺」占比最高。

上述案例驗證了「第三人效應」使個體陷入認知盲區的危險性，更顯現出當前事實查核體系的結構性困境。目前臺灣事實查核中心即便獲得Meta或Google的支持，有限的人力仍難以抗衡AI自動化的產能。當社會公共信任空間在技術異化與媒體逐利中持續萎縮，民眾落實自主查核，便成為防堵假訊息擴散的關鍵力量。

伍、重建社會信任：數位素養與自主查核的雙軌實踐

當前，搜尋引擎導入AI生成摘要作為首要回覆雖提供了便利性，但系統普遍加註「AI可能出錯」的免責聲明，意味著資訊查證的責任已更直接地回歸到使用者身上。在資訊生態質變的時代，面對AI生成內容與詐騙手段的日益多元，重建信任不能僅依賴滯後性的法規限制，除了有賴第三方組織持續進行事實查核，更須深化數位公民的媒體與資訊素養，透過多管並進的策略補強社會資訊防線。

數位公民素養教育除了以圖反搜、AI偵測工具等技術應用，更應轉化為具備深度與韌性的社會心理防衛機制，實踐對抗假訊息的「忍、想、查、問、戳」五字訣：面對激進或可疑訊息時首先「忍」住情緒波動，截斷傳播動機；理性推「想」邏輯矛盾，抓住影像中的視覺破綻，如不自然的水波紋或斑塊；利用工具「查」證資訊的真實性；進而主動詢「問」專業領域人士、醫生或家人朋友，學會識別來源的可信度，化解資訊焦慮；最後確認事實後，「戳」破錯誤資訊，主動傳遞真相。

最重要的是，真正有效的資訊防禦能力，應綜合判斷，而非仰賴特定偵測工具。唯有當大眾展現主動思辨力、銳化資訊判讀力，我們才能在生成式AI的數位幻覺中，守住事實與公共信任的最後防線。

3　內政部警政署統計室（2026年4月1日）。警政統計通報（115年第14週）。取自https://www.npa.gov.tw/ch/index

本文為「知識經濟時代之圖書館服務系列專題演講（三十三）：虛實之間 — GenAI時代圖書館的資訊辨偽與數位敘事」（115.05.06）之演講紀錄，由林芸秀及盧怡伶協助整理（過程有使用AI工具），並經主講者過目授權同意刊登。