

研究論文

眼見不一定為實：生成式人工智慧 深度偽造圖片的真實性認知研究

白龍^{ab}、董飛飛^c、馮迪拉^d

^a東莞理工學院文學與傳媒學院，東莞市，中國大陸

^b南京師範大學新聞與傳播學院，南京市，中國大陸

^c馬來西亞國立大學人文與社會科學與傳播學院，萬宜新鎮，雪蘭莪州，馬來西亞

^d印尼阿赫瑪達蘭大學文化與傳播學院，日惹市，印度尼西亞

摘要

生成式人工智慧 (artificial intelligence, AI) 深度偽造技術創造出的智能欺騙，衝擊人類「有圖有真相」、「目知眼見」的真實性認知傳統。研究考察 229 位受訪者對深度偽造圖片的真實認知表現，發現識別 AI 深度偽造圖片的準確率僅有 39.3%。性別類型和 AI 技能經歷的不同，並不會導致甄別水準有顯著差異，但受訪者觀看文化接近性不同的視覺

白龍，東莞理工學院文學與傳媒學院副教授、南京師範大學新聞與傳播學院媒介發展與危機管理研究所研究員。研究興趣：新媒體與社會治理、智能傳播、新聞業務、媒介哲學。電郵：bailong@dgut.edu.cn

董飛飛，馬來西亞國立大學人文與社會科學與傳播學院博士候選人。研究興趣：健康傳播、資訊行為、社會資本、風險傳播、跨文化傳播。電郵：p129446@siswa.ukm.edu.my

馮迪拉，印尼阿赫瑪達蘭大學文化與傳播學院傳播科學系副教授。研究興趣：社交媒體、新聞業務、政治傳播。電郵：dani.fadillah@comm.uad.ac.id

論文投稿日期：2024 年 8 月 14 日。論文接受日期：2025 年 4 月 1 日。

《傳播與社會學刊》，(總)第 75 期 (2026)

內容在辨識水準上有顯著差異。研究認為人工智慧生成內容 (artificial intelligence generated content, AIGC) 深度偽造技術已經超越了恐怖谷理論 (uncanny valley)¹ 的閾值，虛假線索資訊丟失導致受訪者陷入決策的隨機性，對合成圖像真實性認知的不確定性增強。如未來不能正視真實性危機，人類社會的真實認知將可能被顛覆，「資訊災變」和「現實冷漠」的社會風險就有可能到來。

關鍵詞：生成式人工智慧、深度偽造資訊、智能欺騙、虛假資訊

Research Article

Seeing Is Not Always Believing: A Cognitive Study on the Authenticity of Deeply Falsified Images by Generative Artificial Intelligence

Long BAI^{ab}, Feifei DONG^c, Dira FENG^d

^a School of Literature and Media, Dongguan University of Technology, Dongguan, Mainland China

^b School of Journalism and Communication, Nanjing Normal University, Nanjing, Mainland China

^c Faculty of Social Sciences and Humanities, The National University of Malaysia, Bangi, Selangor, Malaysia

^d Faculty of Culture and Communication, Ahmad Dahlan University, Indonesia, Kota Yogyakarta, Indonesia

Abstract

Intelligent deception created by deep-fake technology in generative AI impacts the human tradition of “pictures do not lie” and “seeing is believing” perception of authenticity. The study evaluates the ability of 229 participants to discern the authenticity of deeply faked images and reveals that accuracy in

Long BAI (Associate Professor, Researcher). School of Literature and Media, Dongguan University of Technology, Institute of Media Development and Crisis Management, School of Journalism and Communication, Nanjing Normal University. Research interests: new media and social governance, intelligent communication, journalism practice, media philosophy. Email: bailong@dgut.edu.cn

Feifei DONG (Ph.D. Candidate). Faculty of Social Sciences and Humanities, The National University of Malaysia. Research interests: health communication, information behavior, social capital, risk communication, cross-cultural communication. Email: p129446@iswa.ukm.edu.my

Dira FENG (Associate Professor). Department of Communication Sciences, Faculty of Culture and Communication, Ahmad Dahlan University, Indonesia. Research interests: social media, journalism practice, political communication. Email: dani.fadillah@comm.uad.ac.id

Article History: Received on 14 August 2024. Accepted on 1 April 2025.

recognizing AI-generated deepfakes stood at only 39.3%. Differences in gender and experience with AI skills did not lead to significant differences in screening levels, but respondents viewing visual content with varied cultural proximity differed significantly in recognition levels. The study concludes that generative AI's deep forgery technology has exceeded the threshold of the "uncanny valley," where the loss of information about false cues leads to randomness in decision-making and increased uncertainty about the perceived authenticity of synthesized images. If this authenticity crisis is not addressed in the future, the perception of truth in human society may be subverted, and the social risks of "information catastrophe" and "reality indifference" may become actualized.

Keyword: Generative artificial intelligence, deep-fake information, intelligent deception, false information

前言

法國攝影師貝亞德 (Hippolyte Bayard) 1840 年捕捉了名為「溺水者的自畫像」的影像，這張照片因為成為歷史上首張公認的偽造照片而「聲名鵲起」。自此以後人類社會圖像操縱偽造技藝亦日臻「完美」。詩人波德萊爾 (Charles Baudelaire) 曾對攝影這一新興媒介嗤之以鼻，認為它是繪畫的勁敵，為那些「才疏學淺」、「懶惰成性」的畫家提供庇護所。如今生成式人工智慧 (Artificial Intelligence, AI) 技術已嶄露頭角取代了原有的圖像偽造技術，成為人類原創性產物的顛覆者。德國藝術家埃爾達格森 (Boris Eldagsen) 曾憑藉 DALL E2 軟體創作的圖片《偽記憶症：電工》(*Pseudomnesia: The Education*) 榮獲索尼世界攝影獎的最高榮譽，然而他最終卻拒絕領獎，此舉引發廣泛爭議。以 Midjourney、DALL-E、Stable Diffusion 等為代表的生成式 AI 軟體，所生成的深度偽造圖片已經達到了肉眼難以分辨真偽的地步。這就導致圖片在視覺文化時代的資訊傳播中佔重越來越大，大量的 AI 合成虛假照片充斥在社交媒體上，使人們對資訊的權威性和真實性產生實質性懷疑，並威脅新聞的可信度。鮑德里亞 (Jean Baudrillard; 1994, p. 47) 在其著作《擬像與仿真》(*Simulation and Simulacra*) 中曾對超真實狀態進行過描述。他認為模擬物的模擬過程始於真實的複製品，隨後逐漸演變為不真實的模擬，即生成複製品的副本。最終這種圖像 (副本之副本) 變得與原始圖像截然不同，從而形成了超真實的狀態。

在此種情境之下現實與表像之間的界限逐漸被消融，致使我們難以區分超現實與人類存在的本真狀態。當前 AI 技術的蓬勃發展引領了數字擬像新時代的降臨，生成式 AI 技術正逐步將人類社會推入到一種名為「數字超真實」的新型現實狀態。在這種由 AI 所構建的模擬環境，真實已不再是先驗的存在或一種現成之物，而是經過精心構造與打磨出來的「人工真實」，其在虛幻的模擬場景中得以呈現 (段偉文, 2020: 108)。生成式 AI 技術具備創造超高逼真度圖像的能力，無疑會加劇我們與現實世界的疏離，並推動人類越發深陷於精心構建的虛擬實境之中。深度偽造圖片的視覺框架，可以被理解為是一種符號化的認知系統，視覺內容構建起有別於話語修辭的另一種修辭範式，對閱聽者發揮強大的勸服功能 (劉濤, 2016)。

隨著視覺數字欺騙技術的疊代升級，人類對真實世界的認知將面臨前所未有的挑戰，深陷難以區分何者為真實，何者為人為製造的境地。由演算法編織的模擬或超現實世界裏，正如波斯特 (Mark Poster; 1988) 所深刻指出的，意義與現實的界定將失去參照物，失去依託與根源，它們將游離於表徵邏輯的既定框架之外，以獨立而自治的姿態運作 (Baudrillard, 2001, p. 7)。AI 生成的視覺誤導資訊，具有煽動恐慌、散播謠言以及操控公眾輿論的巨大風險，並由此可能產生災難性的連鎖反應。以 2023 年 5 月社交媒體上廣泛流傳的五角大樓爆炸偽造圖片為例，該圖片曾一度導致美國股市出現劇烈波動，顯示出不容忽視的破壞力。

生成式 AI 製造的視覺誤導資訊，已成為當今社會治理中不容小覷的重要議題。一方面，深度偽造技術的 AI 照片製造出極其逼真的智能假像，使得事實核查的難度急劇上升；另一方面，這些圖像刻意撩撥公眾的敏感神經，迎合獵奇與新聞觀點的需求，容易引發受眾強烈的情感與觀點碰撞，進而加劇視覺誤導資訊的病毒式擴散。可見深入洞察人類對於生成式 AI 圖片真實性的感知現狀，對於調整和優化現行的 AI 監管政策，以及提出有針對性的虛假圖片治理策略，具有至關重要的長遠意義。

文獻回顧與研究問題

識別 AI 視覺虛假資訊

任何運用原始或操控手法，故意誤導或欺詐，製造與現實大相徑庭的視覺形象，可被稱之為視覺虛假資訊。這種視覺虛假資訊是由不同程度的模態豐富程度和操作複雜性所決定的 (Weikmann & Lecheler, 2023)。視覺誤導資訊蓄意操控視覺感知，通過篡改圖像激發觀眾「熟悉的」錯覺，進而歪曲人們的記憶 (Nash et al., 2009)。這種篡改不僅提高了圖像的可信度，更增加了人們作出錯誤決策的風險。Kasra 等人 (2018) 的研究揭示了一個令人深思的反差現象，儘管人們在面對人工技術操縱的圖像時自認為是鑒別真偽的行家裏手，並事先對偽造圖像

進行充分討論，但他們實際上依然難以準確地鑒別出圖像的真實性。一項針對人類辨識人臉「深度偽造」圖像能力的調查研究表明，儘管參與者在主觀上普遍顯示出較高的自信度，但實證數據卻顯示出他們在識別這些圖像時準確率相對較低 (Bray et al., 2023)。類似的研究利用真實人臉與 AI 合成人臉的圖片進行對照分析，結果顯示，合成引擎所生成的面部圖像在逼真度和可信度方面均達到了令人難以置信的水準，與真實人臉相比其難以辨識的程度顯著提高，使得調查對象在辨識過程面臨更大的視覺挑戰，難以準確區分出哪些是 AI 合成的人臉 (Nightingale & Farid, 2022)。所以人們往往會高估自己對偽造圖片的檢測能力，採用了一種「眼見為實」的啟發式方法進行深度偽造檢測 (Köbis et al., 2021)，對自己水準有限、較弱的檢測能力過度自信。

研究生成式 AI 深度偽造圖片，是建立在全面挑戰圖像檢驗的新技術背景下展開的。相較於傳統的虛假圖片研究，生成式 AI 深度偽造圖片利用生成對抗網絡 (generative adversarial networks, GANs) 營造極其逼真的「虛假」現實，獨特的對抗性訓練技術優勢令合成圖像核查「失靈」。生成式 AI 深度偽造圖片細節的高還原度和細膩度，是人工偽造虛假圖片所不具備的，生成的深度贗品不再有明顯的漏洞或缺陷，開啟了人和機器都無法辨識的視覺欺騙時代。與拼接或複製傳統偽造圖片的圖像編輯專業門檻不同，生成式 AI 深度偽造圖片釋放了大量的自動化視覺欺騙，允許用戶以最簡單的對話指令創建深度偽造品。「視覺恐慌」不再是偶發事件，大量的視覺欺騙成為常態，生成式 AI 深度偽造圖片背後的「全民造假」和「技術濫用」更值得關注。

AI 圖片真實性認知研究

Caramancion (2021) 所進行的研究數據分析揭示，相較於純文本的新聞內容，視覺假新聞中的視覺元素更容易激起人們的懷疑，但是純文本新聞內容在欺詐性方面可能更勝一籌。相較於傳統的純文本和圖文結合的虛假資訊，AI 深度偽造的圖片憑藉更高的感知可信度、豐富的來源生動性以及逼真的客觀事實呈現，反而更易誤導人們做出錯誤

的資訊真實性判斷 (Lee et al., 2021)。Hameleers 等人 (2020) 的實驗亦為該主張提供了佐證，即受眾對於多模態虛假資訊的信任度，無論它的來源如何，均超越了對純文本虛假資訊的信賴，多模態虛假資訊在受眾心中佔據了更為可靠與可信的地位。這是因為用戶對於視覺上的誤導資訊與語言文本資訊的根本性區別有著深刻的認識，根據雙編碼理論的闡釋，視覺資訊在處理過程中完全獨立於語言文本資訊，圖片資訊通過非語言處理途徑被人類大腦所識別和理解 (Clark & Paivio, 1991)。從 Peirce (1955, p. 100) 的解釋可知，圖片通過其索引性創造了一種現實感，在拍攝的物體與觀察者之間圖像構築了一座物理的橋樑，在視覺的語境中，虛假資訊的核心支撐在於其索引性質。這種索引性確保了視覺傳達的真實性與直接性，它是對非仲介環境的物理事件與事實進行的精準刻劃，從而形成了與物理現實內在證據緊密相連的「索引」 (Hameleers et al., 2020)。這將對受眾對於真實性的判斷產生深遠影響，使他們忽略所有圖片背後人為操作和篡改的可能性。圖片的真實性感知離不開細節呈現的上下文背景線索的關聯，「細節問題」是識別 AI 合成圖像的關鍵。儘管存在某些合成圖片的缺陷或者漏洞，可以作為區分真實和虛假照片的線索，但這些缺陷並不總是出現在 AI 生成的圖像中 (Lu et al., 2023)。

因此受眾對圖片真實性的理解往往並不十分精確。鑑於人類工作記憶的持續時長與容量相對有限，從認知負荷理論的視角來看，動態圖像或影像可能會產生巨大的外在認知負荷。相比之下，靜態圖片並不會增加額外的認知負擔。為避免認知超載並最大限度地提高資訊認知效率，人們通常傾向於選擇靜態圖片來獲取大量資訊。圖片的可信度取決於受眾的感知與解讀，受到圖片資料來源、資訊、媒介以及具體文本背景的可信度等多重因素的綜合作用 (Pettersson, 1997)。人們之所會普遍持有「圖片不會撒謊」的堅定信念，其根源在於現實性啟示法所賦予的經驗影響，這種影響在無形中塑造著人們的認知 (Sundar, 2008)。逼真的圖片內容與現實世界具有更高的契合度，從而賦予其更高的可信度。因此當某事物呈現出逼真的外觀時，它更可能被認為是值得信賴的。憑藉圖片優勢效應，圖片在喚起記憶和識別資訊方面明

顯優於文字。人類對於事件中蘊含的象徵性資訊的記憶表現出極高的敏感性 (Paivio & Csapo, 1973)。默認真實理論指出，在人際交往中我們通常傾向於信任他人，認為他們是誠實的，除非有明確的證據表明他們不誠實。同樣我們在接觸到 AI 深度偽造圖片時也有類似預先感受，深度偽造圖片難以觸發人們的懷疑機制。在如今資訊超載的環境中，受訪者很難區分真實資訊與深度偽造圖片，只有當出現明確的懷疑跡象時，才會改變他們對資訊可信度的評估 (Clare & Levine, 2019)，默認真實的存在使得人們辨識 AI 圖片時難以給出統一相似的可信度。實際上，對於虛假視覺內容的信任或受圖像誤導，同樣是一個涉及深度說服的認知過程。

研究問題的提出

GANs 製造人們對 AI 生成超現實圖像的信任，使得肉眼區分真實和偽造內容造成巨大障礙，人們在識別經過操縱的圖像時往往表現不佳，從而對圖片可信度評估產生影響。雖然已有相當一部分研究者著手於探究人類如何辨識 AI 生成的視覺誤導資訊，但關於調查對象個體在性別、技能水準以及文化接近性方面的相關因素，是如何影響識別虛假視覺資訊準確率的研究仍然相對有限。尤其是目前深度偽造生成的 AI 圖片，以其高度逼真的視覺假象特徵影響受眾對真實性的判斷。然而當前對於受訪者識別 AI 圖片視覺內容的認知研究仍然較為稀缺，且鮮有研究關注到個體在多大程度上能分辨出真實與 AI 圖片，不同性別、AI 技能經歷的差異，以及文化接近性的遠近是否會對 AI 圖片分辨產生影響，圖片可信度的相關研究有待進一步拓展。

性別在辨識 AI 深度偽造圖片的能力有可能出現分化。一方面有可能是受社會分工、技術使用習慣等因素影響，用戶在識別深度偽造內容方面存在性別差異。另一方面，生理因素差別導致兩性對圖片感知的注意力不盡相同 (Jiang et al., 2006)，視覺注意力模式也可能因性別而異 (Nightingale et al., 2017)。男性對深度偽造技術的認識之所以更深入，是因為女性普遍面臨性別上的「數字鴻溝」，並沒有更多的時間、

資源、技能知識來接觸技術 (Umbach et al., 2024)。有研究表明，雖然女性在檢測虛假資訊上擁有自信，但女性自信水準與檢測虛假資訊能力之間存在負相關關係，實際檢測能力卻不如男性 (Arin et al., 2023)。結合已有研究，本研究要解釋不同性別在辨識 AI 深度偽造圖片時是否存在差異，故提出以下假設：

H1a：男性受訪者和女性受訪者辨識 AI 深度偽造圖片上無差異；

H1b：男性受訪者和女性受訪者辨識 AI 深度偽造圖片上有差異。

AI 技能經歷同樣有可能對正確識別虛假圖片有影響。多數時候欺騙的產生源自個體經驗不足，經驗對於甄別欺騙至關重要，AI 使用經驗在辨別 AI 虛假圖片的作用就像是「數字免疫力」，經過訓練的使用者可以動用技術經驗揭示合成圖片的漏洞和缺陷。已有研究顯示 AI 使用經驗對於辨別合成圖片確實有顯著優勢，具有 AIGC 背景 (即玩過 AIGC 的用戶) 的參與者在區分真偽圖片上表現更好 (Lu et al., 2023)。另一項研究也證明具有數字操作訓練和招聘編輯或攝影經驗的個人，比未經訓練的、沒有 AI 使用經驗的人更能識別出偽造的圖片 (Moshel et al., 2022)。因此，人們的數字媒體素養和社交媒體使用習慣是評估圖片可信度的重要因素 (Shen et al., 2019)。基於此，不同的技能經歷是否會導致辨識 AI 深度偽造圖片存在差異？這裏具有 AI 技能經歷主要是針對有 AI 作圖經驗，即「AI 文生圖」的使用經歷。結合現有發現，本研究提出以下假設：

H2a：具有 AI 技能經歷的受訪者更容易辨識出 AI 深度偽造圖片；

H2b：不具有 AI 技能經歷的受訪者更難辨識出 AI 深度偽造圖片。

文化接近性是影響個體對虛假圖片的辨別標準和信任程度的重要因素。這是因為文化背景會影響觀看圖片注意力的分配方式 (Boduroglu et al., 2009)，受到特定的文化視角影響，處於差異化的文化背景的個體表露出不同的資訊加工偏好。有實驗對比不同文化背景的參與者對偽造圖片的感知後發現，文化的接近性造成一種建設性的錯誤記憶，產生幻想記憶或者是「錯誤的」熟悉感 (Wang et al., 2021)。有學者發現在不同文化背景下，細節感知的差異對識別欺騙行為有顯著影響，有

文化經驗的個人更能夠從中察覺到不同於文化範式細節的異常信號 (Taylor et al., 2017)。文化上的接近性令個人的文化記憶和經歷被調動起來 (Howells, 2019)，人們更有可能有足夠的線索資訊來識別圖像。因此，本研究將文化接近性視為有可能是導致辨識 AI 深度偽造圖片存在差異的重要原因，提出以下假設：

- H3a：受訪者更容易辨識文化接近的 AI 深度偽造圖片；
H3b：受訪者更難以辨識文化不接近的 AI 深度偽造圖片。

研究設計及研究結果

研究通過採用問卷調查方式，獲取受訪者對於生成式 AI 深度偽造圖片真實性的認知表現。本研究於 2023 年 12 月在南京一所傳媒院校進行，為確保數據的一致性和可靠性研究選擇了在課間時段進行集中調查。問卷發放前進行了 20 人的小範圍前測，結合受訪效果對問卷的措辭和題目表述進行調整，並結合了認知訪談了解目標對象對 AI 深度偽造圖片的直觀評價。採用克隆巴赫 α 係數 (Cronbach's alpha) 評估後發現，克隆巴赫 α 係數為 .925，反映問卷的內部一致性較好。前為避免外部環境、時間等其他因素對問卷調查過程的影響，調查對象集中於課間同一環境進行訪問。受訪者事先並不知曉該研究問卷圖片的真偽性，圖片辨認順序採用隨機呈現的圖片刺激方式進行，問卷填寫後將圖片真實性情況均告知受訪者本人，同樣亦告知研究的真實目的。經過嚴格的篩選，共招募了 294 名本科大學生作為研究對象，年齡分佈主要集中在 18 至 25 歲之間，鑑於調研所在的學校是以傳媒藝術類專業的學生為主，女性學生的比例相對較高，佔到了有效樣本的 87.8%，而男性學生則佔 12.2%。調研對象遴選的要求首先是互聯網的深度使用者，需要擁有生成式 AI 平台使用經歷，並接觸過 AI 深度偽造圖片。在數據篩選過程中，研究剔除了未完成的問卷以及明顯不符合邏輯的問卷答案數據，最終獲得了 229 份問卷數據。問卷中的圖片素材經人工挑選，包括 18 張由 AI 生成的虛假圖片和 18 張真實拍攝的圖片，共計 36 張。AI 生成的偽造圖片均來源於中國外知名的新聞媒體或權威網站，並已

《傳播與社會學刊》，(總)第 75 期 (2026)

得到證實為人工智能生成的圖片。圖一為用於調查問卷的圖片樣本。真實圖片則隨機抽取自美國《時代週刊》和新華社的年度照片集，以降低圖片選擇偏好。這些實驗圖片材料來自中立來源的媒體圖片，避免有可能導致個人情感反應或與自己周邊環境相關的照片納入其中。受訪者在回答問題前並不知曉圖片的真實性。真實和虛假圖片均以隨機排列的方式展示，他們被要求在每張圖片上給出「是」或「否」的回答，以確定該圖片是否屬於深度偽造圖像。

圖一 教宗方濟各身穿巴黎世家羽絨服的 AI 偽造圖



註：原圖詳見 Huang (2023)

經數據統計分析後，我們觀察到受訪者在辨識真實圖片時的精確度為 58.6%，然而，在識別由 AI 深度偽造的圖片時，其準確率大幅下降至 39.3%，對比詳細資訊見表一。值得注意的是，儘管在調查開始之前大部分受訪者對自己區分真實圖像與深度偽造圖像的能力抱有充分的自信，但不足四成的正確率卻揭示了一個巨大的反差，即大多數受訪者實際上已難以準確判斷 AI 深度偽造圖像的真實性。這也同時反映

出圖片的臨場感更為突出，啟蒙時代以來科學理性為根基的真實觀，正在被應用「生成對抗網絡」演算法模型的深度偽造技術迅速瓦解。圖片的憑證效果曾被施瓦爾茨神話為是真實世界的「宣誓證人」與「沉默的目擊者」(徐笛、梁鶴，2023)，擁有生成式AI技術的加持，圖片的視覺欺騙能力進一步增強。當距離真實世界最近和「在場感」呈現能力最強的AI深度偽造圖片出現後，個人視覺感官在面對這種「無懈可擊」的視覺欺騙變得毫無招架之力。

表一 真實與AI圖片辨別正確率對比表

圖片類型	命中個案數(張)	識別準確率
真實圖片	2,414	58.6%
AI圖片	1,620	39.3%

由於性別樣本比例佔比相差較大，為確定性別因素是否會對識別AI圖片真實性存在影響，研究對個案進行了加權處理，利用卡方檢驗來判斷樣本性別比例是否與預期存在顯著差異。通過交叉表逐一檢視，結果顯示 p 值均小於.05，意味著拒絕原假設，即不同性別在識別生成式AI圖片時，其甄別能力並不存在顯著差異。表二展示的是偽造教宗方濟各的AI生成圖片的卡方檢驗。這一結果揭示了性別因素在辨別AI圖像真偽的能力上並未顯示出明顯的影響，無論是男性還是女性，均未能有效地從真實的視覺資料中區分出AI生成的虛假圖像。研究進一步證明了海德(Janet Hyde; 2005)的性別相似性假說在識別AI圖片真偽性上是成立的，即男女之間的相似性大於差異性，兩性在動用認知資源來判斷資訊真偽性的差異消失。

表二 不同性別的卡方檢驗表

卡方檢驗	值	自由度	漸進顯著性(雙側)
皮爾遜卡方	.27 ^a	2	.87
似然比	.29	2	.87
線性關聯	.03	1	.85
有效個案數	43		

註：^a1 個單元格 (16.7%) 的期望計數小於 5。最小期望計數為 3.84

針對生成式AI軟體使用經驗是否對辨識真偽能力產生影響的問題，研究測量題項是「您是否曾經用過生成式AI軟體(如Midjourney、Dall-E、Stable Diffusion、文心一言等)生成圖片」，對使用過與未使用過此類軟體的受訪者分辨能力分別進行了卡方檢驗。研究結果如下表三所示，發現皮爾遜卡方值為3.757，顯著性 p 值=.053>.05，即拒絕原假設，即認為兩類受訪者在分辨能力上並無顯著差異。由此可見，即便受訪者熟悉AI軟體操作或具備一定的AI技術基礎，他們在鑒別極其生動、逼真的合成圖像時是無力的。視覺形態資訊可以憑藉最直觀的畫面展示影響人們的態度和情感(劉國強、張思雨，2023)，深度偽造的「現場感」以多模態資訊說服受眾，使之信以為真。僅憑個人直觀的視覺感受，他們難以準確判斷出那些生成品質越來越高、缺陷越來越少的AI深度偽造圖像。

表三 不同AI技能經歷的卡方檢驗表

卡方檢驗	值	自由度	漸進顯著性 (雙側)	精確顯著性 (雙側)	精確顯著性 (單側)
皮爾遜卡方	.11 ^a	1	.74		
連續性修正 ^b	.08	1	.77		
似然比	.11	1	.74		
費希爾精確檢驗				.73	.39
線性關聯	.11	1	.74		
有效個案數	4,133				

註：^a0個單元格(0.0%)的期望計數小於5。最小期望計數為197.55，^b僅針對 2×2 表進行計算

調查對象同樣被安排觀察兩組具有顯著差異的AI圖像——一組源自國內，另一組則來自海外。經卡方檢驗後發現，皮爾遜卡方值為99.016，顯著性 p 值<.001，這意味著參與者在區分不同地域來源的AI圖像真實性方面確實展現出顯著差異。值得注意的是，當面對與自身文化背景更為接近的國內場景的AI圖像時，參與者的識別準確率達到了49%，這一成績顯著優於他們在面對異國場景的圖像表現。這一發現進一步證實，文化接近性在圖像識別過程中扮演著重要角色。尤其是當圖像來源於遠離參與者文化背景的異國環境時，識別難度顯著增加，導致準確率大幅下降至33%。

表四 不同地區的卡方檢驗表

卡方檢驗	值	自由度	漸進顯著性 (雙側)	精確顯著性 (雙側)	精確顯著性 (單側)
皮爾遜卡方	99.01 ^a	1	***		
連續性修正 ^b	98.35	1	***		
似然比	98.09	1	***		
費希爾精確檢驗				***	***
線性關聯	98.99	1	***		
有效個案數	4,132				

註：^a 0個單元格 (0.0%) 的期望計數小於5。最小期望計數為542.61，^b 僅針對 2×2 表進行計算：「*** $p < 0.001$ 」

結論與討論

通過問卷調查數據的綜合分析，本研究深入探討了人們對生成式AI偽造圖片真實性的認知表現。主要結論如下：

第一，在無文本資訊提示下，受訪者在辨識AI圖片真偽方面的 ability普遍較為薄弱，作出無線索的隨機決策。多數受訪者無法獨立驗證被操縱圖像的真實性，尤其是那些高度逼真的圖片，其真實性難以被觀看者所察覺，幾乎無法發現被篡改或編輯的痕跡。從現實生活的經驗出發，這些極具欺騙性的AI圖片很容易誤導人們，使其誤認為它們是真實的。此外，不帶文本資訊的生成式AI圖片出現的低準確率，進一步印證了 Powell 等人 (2015) 的觀點，即單獨的圖像資訊相較於單獨的文本資訊，具有更強的框架效應，能夠在沒有文本輔助的情況下影響人們的觀點和預期行為。在大多數情境中，我們往往對欠缺核心事實支撐的圖片缺乏深入的審視。視覺感官直接傳達給我們資訊，使個人自然而然地作出判斷，這種直覺式的處理方式無疑降低了我們的認知負擔。然而當圖片的資訊來源存疑或關鍵事實數據缺失時，人們常常表現出一種認知上的惰性，往往都是「認知吝嗇鬼」，傾向於採用更為簡單、省力的思維路徑來解決問題。在日常生活中人們習慣於依賴直覺來做決策，但直覺並非總是準確無誤。很多時候我們的判斷與預期結果大相徑庭。這種依賴於啟發式心理捷徑的決策方式，依賴於個

人的直覺和本能的判斷雖然快速，但往往缺乏嚴謹的邏輯推理，因此容易出錯。另一方面，從線索效用理論看，人們在多線索情景之下選擇有效線索進行決策判斷 (Cox, 1962)。但深度偽造的視覺資訊提供的有效線索有限，受訪者並不知曉圖片的真實來源，圖片也沒有附加提示等相關資訊，沒有給受訪者識別深度偽造視覺內容帶來足夠充足的外在線索，導致辨識精準度大為下降。

第二，性別類型對圖像真實性認知的影響並無顯著區別。儘管有一些研究揭示了公眾對不同性別在技術能力上的認知差異，認為技術主要是由男性主導並熟練掌握，且社會和文化規範也在一定程度上助長了男性在技術領域的更多獲得機會 (Mansoori et al., 2023)。但本研究數據揭示了一個「男女平等」的認知趨勢，便是在判斷 AI 圖片真偽的任務中，男性並未展現出任何明顯的優勢。事實上，男性和女性在感知 AI 圖片真實性方面普遍表現出類似的趨勢，性別之間的差異並不顯著。深入剖析「選擇性模型」，我們不難發現女性在處理資訊時更傾向於採取綜合的方式，她們會全面考慮資訊的主觀和客觀屬性，對微妙的細節和線索給予足夠的重視。相對而言，男性則傾向於採取選擇性的資訊處理策略，他們更可能依賴啟發式處理，從而可能忽略了那些微妙的線索 (Darley & Smith, 1995)。在視覺空間轉換的能力上，男性展現出明顯的優勢，而女性則更擅長於對視覺空間資訊的專注存儲。當涉及對靜態圖像細節的描繪和場景呈現時，女性相較於男性展現出更高的傾向性和敏銳度 (Lawton & Hatcher, 2005)。如果從對視覺資訊的易感性角度考慮，女性通常能更快地捕捉到 AI 深度偽造中的虛假痕跡與偽造線索。儘管女性在細節資訊的處理上具備優勢，但研究結果卻顯示她們在識別 AI 生成圖片的真實性方面並未表現出更高的精確度。這表明，在 AI 圖片鑒別的任務中，任何性別所特有的優勢均難以體現。

第三，AI 技能經歷並不會對辨識 AI 深度偽造圖片的能力產生顯著影響。這種由生成式 AI 所生成的深度偽造圖片，可以被視為一種認知上的威脅，因為它破壞了照片作為可靠證據的價值基礎 (Fallis, 2021)。深度偽造的圖片對於普羅大眾而言是極難精確辨識的，所以依賴經驗、習性與常規進行決策似乎是最為便捷的途徑。雙重加工理論強調

了資訊處理的兩個獨立認知加工系統，即經驗系統和理性系統 (Osman, 2004)。Petty 和 Cacioppo (1984) 提出的「推敲可能模型」(elaboration likelihood model, ELM) 詳盡闡釋了個體態度轉變的兩種核心路徑，即中心與邊緣認知路徑。當面對與個人緊密相關的對象或數據時，個體傾向於採取中心認知路徑，投入更多時間與精力進行全面而系統的思考。當對象或數據與個體的關聯度較低，或者個體缺乏足夠的動機時，他們往往會選擇付出較少的努力，並更加依賴易於處理的週邊線索進行認知。但在缺少上下文的經驗依賴性的前提下，受訪者並不能從經驗系統的路徑尋求資訊處理，同樣在理性系統決策過程也沒有辦法強調推理邏輯，有經驗的受訪者也只能尋求隨機決策，作出無線索的直觀判斷。因此，經驗在資訊處理加工的作用是失效，甚至產生副作用。這是因為若個體曾利用類似軟體創造深度偽造圖片，其內心會形成一種「認知預演」，對虛假圖片的逼真性保持警惕，甚至產生「免疫」機制，無法依賴過往經驗相似性做出判斷。研究表明，當涉及鑑別由生成式AI所創造的深度偽造圖片時，個人的AI技術使用經驗變得不再可靠。即便是那些熟悉並熟練掌握AI軟體技術的受訪者，其辨識AI深度偽造圖片的能力並未因此展現出更高的水準。

第四，受訪者觀看不同文化接近性的AI深度偽造圖片時展現出顯著差異。調查對象在國內文化背景的圖像中更容易察覺出AI的偽造痕跡，而對於國外文化背景的圖像，其辨識能力則相對較低。這一現象揭示了文化親近性在影響人們對AI生成圖片真偽辨識中的重要作用。依據弗洛伊德 (Sigmund Freud) 的潛意識理論，個體的行為多數受到內在潛意識的驅動，而決策過程則深受其所處的社會文化背景的影響。從這個角度看，文化的影響實際上構成了一種隱形的社會壓力，使得決策者在無意識中利用自身的文化傳統進行認知判斷並採取相應的行動 (郭鵬，2016：323)，從某種意義上說文化影響了人的決策。而文化也有差別，對文化的分類實質上是一種認知過程，它涉及將眾多具有不同特徵的實例歸類並打上相同的標籤，以「簡化」或「壓縮」這個極其複雜多變的世界，並讓我們能夠迅速作出反應和進行有效的溝通 (Brasselet & Arleo, 2018)。文化背景在塑造個體的知覺、解讀與回應方面扮演著至關重要的角色。個體對文化的感知通常以自我為核心，向

外逐層遞減，這取決於個體對文化的熟悉程度。在日常生活中，人們習慣於根據文化背景的親疏進行分類，並在處理與自己文化背景相似的事物時，展現出更高的決策效率。文化環境的相近性與相似性對我們做出精準判斷具有顯著影響。當面對物理距離較近或文化背景相似的事物時，人們更容易產生共鳴，並迅速識別出異常之處。然而，當個人接觸到與自己文化背景差異較大的事物時，他們可能會缺乏必要的文化敏感性，難以利用自身的文化背景進行準確辨識，從而可能導致決策失誤。

本次研究亦存在如下局限和可繼續拓展的方向：一是AI深度偽造圖片態度認知探索不夠深入。雖然深度偽造圖片的真實性認知屬於態度感知的範疇，但鑑於實驗條件和研究目的限制，本研究只局限於合成圖片的真偽認知層面，後續AI深度偽造圖片所引發的行為和情感同樣也值得關注。二是個人特徵影響認知的變數維度覆蓋面不足。對性別、技能經歷、文化接近性三個維度的考察，只能是對生成式AI真實性認知的部分解釋，有更多元的個人特徵可以進一步探究。三是樣本總體豐富性存在局限。為方便抽樣調查，研究對象主要來自青年群體，並且女性受訪者居多，這忽略了不同學歷、年齡、職業等層面對AI真實性認知的影響，未來可採用更大樣本或更廣泛的目標群體開展研究，以證明以上結論的普適性。

深度偽造圖片的來源對於大多數人而言，往往顯得無關緊要或撲朔迷離。圖片本身所蘊含的資訊量，往往不足以激發受眾採取中心說服路徑的動機。當脫離具體語境的視覺資訊處理能力受限時，邊緣路徑往往成為主導認知的主要通道。在這種情況下，人們的決策過程更傾向於依賴啟發式線索和認知捷徑來進行資訊加工。雙重過程理論指出我們大腦的推理決策過程，源於兩種截然不同的思維方式：一種是自動且無意識的過程，另一種則是明確且有意識的過程 (Shleifer, 2012)。當辨識策略轉向依賴自動且無意識的邊緣路徑時，這必將導致可信度的波動。AI所生成的深度偽造圖片，通過一種視覺上的超現實感官，讓「眼見為實」的原則受到挑戰，進而塑造出一種看似毫無修飾、直接且未經篡改的現實視角。這種視覺內容的現實表現，其多模態性以及豐富的直接線索，可以被視作是一種啟發式的指引 (Hameleers et al., 2024)。

深度偽造圖片資訊提供了更為根本且省力的啟發式線索，從而能夠規避錯誤的消息參數處理，進而提升資訊可信度的接納度。這些生動逼真的視覺資訊源自不同管道，卻增加了可信度的不確定性。

科拉維塔效應彰顯了視覺資訊的非凡優越性，在眾多感官刺激的交織中，視覺處理能力顯然超越其他感官，它佔據主導地位而且處理速度更快 (Colavita, 1974)。相較於口頭傳達的資訊，人們通常更偏愛信任視聽內容，因為後者往往能夠以更為逼真的方式展現出現實世界的原貌。所以圖片理應被期待成為描繪世界真實面貌的藍圖，攝影圖片作為其源頭，是對現實世界的可靠描繪，深深烙印著客觀性的印記，是一種不會引起人們對現實存在產生懷疑的媒介。但是數字成像技術的興起開始扭曲了現實的本質，數字化處理創造了「摻雜」的發明，除了電腦本身，沒有其他來源，混雜的圖像素材僅僅是符號化的表像，缺乏真實性和可信度 (Batchen, 1994)。特別是生成式AI的深度偽造技術，已經超越了「恐怖谷」的閾值，進一步削弱了「圖像真實性」的固有防線。在埃爾達格森 (Boris Eldagsen) 的洞察中，AI生成的圖片是一種「啟示性攝影」，這不僅僅是簡單的複製或剽竊。相反，人類的智慧被巧妙地委托給了演算法，這些演算法存儲了所有的疊代過程，掌握了我們集體無意識的深厚知識。因此AI實際上是在踐行柏拉圖的理念論，即知識和真理的本質並非源自物質世界，而是源於抽象的、理想的形式。與此同時，我們也需要構建一個「紅綠燈體系」，即人類社會亟須對圖片進行明確標註，區分為真實、經過操縱或生成的三種不同類別，以確保資訊的透明度和準確性 (Williams, 2023)。AI所催生的虛擬與現實的無縫融合正逐漸成為社會發展的新常態。科學技術的介入令視覺媒介構建多元、多重的流動性「真實」，「看」世界的角度趨向於高度統一，但深度偽造圖片以高度明晰的「可見性」遮蔽了建構和組織世界的本質 (胡翼青、王沐之, 2022)，AI技術描摹出世界想像的豐富可見度，卻意味著對真實世界的不可見性增強了。若我們未能採取果斷行動，以確保構建一個能夠明確區分真偽的資訊生態系統，我們可能會面臨一場如奧瓦迪亞 (Aviv Ovadya) 所描述的「資訊災變」。這一災變象徵著思想市場的毀滅性失敗，人們在不斷接觸錯誤資訊的過程中，逐漸失去了對所見所聞的信任，最終滑向「現實冷漠」的深淵，

對真實世界的感知和關懷日漸淡漠 (Ovadya, 2018)。此時沒有人相信任何真相，也沒有人相信任何謊言。

註釋

- 1 日本機器人專家森昌弘提出了恐怖谷效應 (uncanny valley)，說明當人類接觸到某些類似人形特徵的機器人外觀之時，通常會由於物體外觀或是行為過於複雜而難以區分，導致認知失調後引發不適感甚至是恐懼、反感的情緒。參見 Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19(2), 98–100.

披露聲明

本文作者未報告潛在的利益衝突。

Disclosure Statement

No potential conflict of interest was reported by the authors.

研究經費資助

本文受國家社會科學基金重大招標專案「我國青少年網絡輿情的大數據預警體系與引導機制研究 (20 & ZD012)」及 2023 年度江蘇省高校哲學社會科學研究一般專案「智能傳播時代的媒介哲學思想創新與技術應用倫理風險研究 (2023SJYB0637)」項目資金資助。

Funding

This article was funded by National Social Science Fund Major Projects program under the grant “Big Data Early Warning System and Guidance Mechanism for Online Public Opinion Among Chinese Adolescents (20 & ZD012)” and The 2023 Annual Jiangsu Provincial Philosophy and Social Sciences Research General Project for Higher Education Institutions for the

study titled “Innovation in Media Philosophy and Ethical Risk Assessment of Technology Applications in the Era of Intelligent Communication (2023SJYB0637).”

ORCID

白龍 (Long BAI) <https://orcid.org/0000-0001-9879-5195>

董飛飛 (Feifei DONG) <https://orcid.org/0009-0001-4211-4650>

馮迪拉 (Dira FENG) <https://orcid.org/0000-0003-4387-4501>

參考文獻

中文部分 (Chinese Section)

胡翼青、王沐之 (2022)。〈作為媒介性的可見性：對可見性問題的本體論探討〉。《新聞記者》，第4期，頁8–19。

Hu, Y., & Wang, M. (2022). Visibility as mediation: An ontological exploration of the issue of visibility. *Shanghai Journalism Review*, 4, 8–19.

段偉文 (2020)。《資訊文明的倫理基礎》。上海人民出版社。

Duan, W. (2020). *The ethical foundation of information civilization*. Shanghai People's Publishing House.

徐笛、梁鶴 (2023)。〈循跡網絡：深度造假與新聞真實體制〉。《全球傳媒學刊》，第10期，頁153–169。

Xu, D., & Liang, H. (2023). Networking of tracing: Deep fake and the regime of journalistic truth. *Global Journal of Media Studies*, 10, 153–169.

郭鵬 (2016)。《數據、模型與決策》。西北工業大學出版社。

Guo, P. (2016). *Data, models and decision-making*. Northwestern Polytechnical University Press.

劉國強、張思雨 (2023)。〈演算法失誤會影響用戶的使用意願嗎？——基於事實核查演算法的實驗研究〉。《新聞記者》，第10期，頁16–27。

Liu, G., & Zhang, S. (2023). Does algorithm error affect users' intention to use? An experimental study based on fact-checking algorithms. *Shanghai Journalism Review*, 10, 16–27.

《傳播與社會學刊》，(總)第75期(2026)

劉濤(2016)。〈西方數據新聞中的中國：一個視覺修辭分析框架〉。《新聞與傳播研究》，第23期，頁5–28。

Liu, T. (2016). China in Western data journalism: Searching for an analytic framework of visual frame based on visual rhetoric. *Journalism & Communication*, 23, 5–28.

英文部分 (English Section)

- Arin, K. P., Mazrekaj, D., & Thum, M. (2023). Ability of detecting and willingness to share fake news. *Scientific Reports*, 13(1), 7298.
- Batchen, G. (1994). Phantasm: Digital imaging and the death of photography. *Aperture*, 136, 46–51.
- Baudrillard, J. (1994). *Simulacra and simulation*. University of Michigan Press.
- Baudrillard, J. (2001). *Jean Baudrillard: Selected writings*. Stanford University Press.
- Boduroglu, A., Shah, P., & Nisbett, R. E. (2009). Cultural differences in allocation of attention in visual information processing. *Journal of Cross-cultural Psychology*, 40(3), 349–360.
- Brasselet, R., & Arleo, A. (2018). Category structure and categorical perception jointly explained by similarity-based information theory. *Entropy*, 20(7), 527.
- Bray, S. D., Johnson, S. D., & Kleinberg, B. (2023). Testing human ability to detect ‘deepfake’ images of human faces. *Journal of Cybersecurity*, 9(1), tyad011.
- Caramancion, K. M. (2021). Textual vs. visual fake news: A deception showdown. In *2021 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM)* (pp. 31–35). IEEE.
- Clare, D. D., & Levine, T. R. (2019). Documenting the truth-default: The low frequency of spontaneous unprompted veracity assessments in deception detection. *Human Communication Research*, 45(3), 286–308.
- Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, 3(3), 149–210.
- Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics*, 16(2), 409–412.
- Cox, D. F. (1962). The measurement of information value: A study in consumer decision-making. *Emerging Concepts in Marketing*, 413, 21.
- Darley, W. K., & Smith, R. E. (1995). Gender differences in information processing strategies: An empirical test of the selectivity model in advertising response. *Journal of Advertising*, 24(1), 41–56.
- Fallis, D. (2021). The epistemic threat of deepfakes. *Philosophy & Technology*, 34(4), 623–643.
- Hameleers, M., Powell, T. E., Van der Meer, T. G. L. A., & Bos, L. (2020). A picture paints a thousand lies? The effects and mechanisms of multimodal

- disinformation and rebuttals disseminated via social media. *Political Communication*, 37(2), 281–301.
- Hameleers, M., Van Der Meer, T. G. L. A., & Dobber, T. (2024). Distorting the truth versus blatant lies: The effects of different degrees of deception in domestic and foreign political deepfakes. *Computers in Human Behavior*, 152, 1–13.
- Howells, K. (2019). *Exploring iconic images created by the Ministry of Information and their relation to cultural memory in Britain* [Unpublished doctoral dissertation]. King's College London.
- Huang, K. (2023, April 8). *Why Pope Francis is the star of AI-generated photos*. The New York Times. <https://www.nytimes.com/2023/04/08/technology/ai-photos-pope-francis.html>.
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, 60(6), 581.
- Jiang, Y., Costello, P., Fang, F., Huang, M., & He, S. (2006). A gender- and sexual orientation-dependent spatial attentional effect of invisible images. *Proceedings of the National Academy of Sciences, USA*, 103(45), 17048–17052.
- Kasra, M., Shen, C., & O'Brien, J. F. (2018). *Seeing is believing: How people fail to identify fake images on the web* [Conference presentation abstract]. The 2018 CHI conference on Human Factors in Computing Systems (CHI'18) (pp. 1–6). Montreal, Canada. <https://doi.org/10.1145/3170427.3188604>.
- Köbis, N. C., Doležalová, B., & Soraperra, I. (2021). Fooled twice: People cannot detect deepfakes but think they can. *Iscience*, 24(11).
- Lawton, C. A., & Hatcher, D. W. (2005). Gender differences in integration of images in visuospatial memory. *Sex Roles*, 53(9–10), 717–725.
- Lee, C., Yang, T., Inchoco, G., Jones, G. M., & Satyanarayan, A. (2021). Viral visualizations: How coronavirus skeptics use orthodox data practices to promote unorthodox science online. *Proceedings of the 2021 CHI conference on human factors in computing systems*. Yokohama, Japan (pp. 1–18). <https://doi.org/10.1145/3411764.3445211>.
- Lu, Z., Huang, D., Bai, L., Qu, J., Liu, X., & Ouyang, W. (2023). Seeing is not always believing: A quantitative study on human perception of ai-generated images. *arXiv*. <https://doi.org/10.48550/arXiv.2304.13023>.
- Lu, Z., Huang, D., Bai, L., Qu, J., Wu, C., Liu, X., & Ouyang, W. (2023). Seeing is not always believing: Benchmarking human and model perception of AI-generated images. *Advances in Neural Information Processing Systems*, 36, 25435–25447.
- Mansoori, A., Tahat, K., Tahat, D., Habes, M., Salloum, S. A., Mesbah, H., & Elareshi, M. (2023). Gender as a moderating variable in online misinformation acceptance during COVID-19. *Heliyon*, 9(9).

《傳播與社會學刊》，(總)第 75 期 (2026)

- Moshel, M. L., Robinson, A. K., Carlson, T. A., & Grootswagers, T. (2022). Are you for real? Decoding realistic AI-generated faces from neural activity. *Vision Research*, 199, 108079.
- Nash, R. A., Wade, K. A., & Brewer, R. J. (2009). Why do doctored images distort memory? *Consciousness and Cognition*, 18(3), 773–780.
- Nightingale, S. J., & Farid, H. (2022). AI-synthesized faces are indistinguishable from real faces and more trustworthy. *Proceedings of the National Academy of Sciences, USA*, 119(8), e2120481119.
- Nightingale, S. J., Wade, K. A., & Watson, D. G. (2017). Can people identify original and manipulated photos of real-world scenes? *Cognitive Research: Principles and Implications*, 2(1), 30.
- Poster, M. (1988). Introduction. In J. Baudrillard, *Selected writings* (p. 7). Stanford University Press.
- Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review*, 11(6), 988–1010.
- Ovadya, A. (2018, February 22). *What's worse than fake news? The distortion of reality itself*. The Washington Post. <https://www.washingtonpost.com/news/the-worldpost/wp/2018/02/22/digital-reality/>.
- Paivio, A., & Csapo, K. (1973). Picture superiority in free recall: Imagery or dual coding? *Cognitive Psychology*, 5(2), 176–206.
- Peirce, C. S. (1955). Logic as semiotic: The theory of signs. In J. Buchler (Ed.), *Philosophical writings of Peirce* (pp. 98–119). Dover Publications.
- Pettersson, R. (1997, August 9–12). *The credibility of pictures*. 13th Nordic Conference for Research on Mass Communications, Jyväskylä, Finland.
- Petty, R. E., & Cacioppo, J. T. (1984). Source factors and the elaboration likelihood model of persuasion. *Advances in Consumer Research*, 11(1), 668–672.
- Powell, T. E., Boomgaarden, H. G., De Swert, K., & De Vreese, C. H. (2015). A clearer picture: The contribution of visuals and text to framing effects. *Journal of Communication*, 65(6), 997–1017.
- Shen, C., Kasra, M., Pan, W., Bassett, G. A., Malloch, Y., & O'Brien, J. F. (2019). Fake images: The effects of source, intermediary, and digital media literacy on contextual assessment of image credibility online. *New Media & Society*, 21(2), 438–463.
- Shleifer, A. (2012). Psychologists at the gate: A review of Daniel Kahneman's *Thinking, fast and slow*. *Journal of Economic Literature*, 50(4), 1080–1091.
- Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 73–100). The MIT Press.
- Taylor, P. J., Larner, S., Conchie, S. M., & Menacere, T. (2017). Culture moderates changes in linguistic self-presentation and detail provision when deceiving others. *Royal Society Open Science*, 4(6), 170128.

- Umbach, R., Henry, N., Beard, G. F., & Berryessa, C. M. (2024). Non-consensual synthetic intimate imagery: Prevalence, attitudes, and knowledge in 10 countries. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, USA. <https://doi.org/10.1145/3613904.3642382>.
- Wang, J., Otgaar, H., Santtila, P., Shen, X., & Zhou, C. (2021). How culture shapes constructive false memory. *Journal of Applied Research in Memory and Cognition*, 10(1), 24–32.
- Weikmann, T., & Lecheler, S. (2023). Visual disinformation in a digital age: A literature synthesis and research agenda. *New Media & Society*, 25(12), 3696–3713.
- Williams, Z. (2023, April 18). 'AI isn't a threat'—Boris Eldagsen, whose fake photo duped the Sony judges, hits back. The Guardian. <https://www.theguardian.com/artanddesign/2023/apr/18/ai-threat-boris-eldagsen-fake-photo-duped-sony-judges-hits-back>.

本文引用格式

白龍、董飛飛、馮迪拉 (2026)。〈眼見不一定為實：生成式人工智慧深度偽造圖片的真實性認知研究〉。《傳播與社會學刊》，第 75 期，頁 119–143。

Citation of This Article

Bai, L., Dong, F., & Feng, D. (2026). Seeing is not always believing: A cognitive study on the authenticity of deeply falsified images by generative artificial intelligence. *Communication and Society*, 75, 119–143.