

專輯論文

社交媒體環境下公眾對生成式AI信任的形成機制研究

高芳芳^a、徐欣妍^a

^a浙江大學傳媒與國際文化學院，杭州市，中國大陸

摘要

信任是影響生成式AI社會應用與推廣的關鍵因素。本研究基於問卷調查($N = 787$)，採用結構方程模型(structural equation modeling, SEM)與模糊集定性比較分析(fuzzy-set qualitative comparative analysis, fsQCA)相結合的混合方法，考察了社交媒體環境下公眾對生成式AI信任的形成機制。SEM的結果表明，社交媒體接觸能通過強化AI知識、增強AI效能感及降低AI隱私風險感知三條路徑來促進公眾對生成式AI的技術信任與系統信任。但社交媒體接觸僅能通過降低社會風險感知來提高技術信任，卻無法經由這一路徑對系統信任產生顯著影響。fsQCA分析進一步揭示了上述因素對生成式AI信任的組態效應。系統信任的形成組態包括效能驅動—風險規避型、知識驅動—風險規避型

高芳芳，浙江大學傳媒與國際文化學院副教授。研究興趣：科學傳播、數字媒介研究。電郵：ffgao@zju.edu.cn

徐欣妍，浙江大學傳媒與國際文化學院碩士研究生。研究興趣：智能傳播、健康傳播。電郵：22423015@zju.edu.cn

論文投稿日期：2025年1月22日。論文接受日期：2026年7月2日。

《傳播與社會學刊》，（總）第78期（2026）

與社交媒體賦能型三類。技術信任的形成組態更為多元，在上述三類組態的基礎上還涵蓋媒介驅動－風險規避型。本研究拓展了對生成式AI信任的概念框架，釐清了技術信任與系統信任的差異化形成機制，並為政策制定者和技術開發者提供了實踐建議。

關鍵詞：生成式AI信任、社交媒體接觸、AI知識、AI效能感、AI風險感知

Special Issue Article

Exploring Public Trust in Generative AI and Its Formation Mechanism on Social Media

Fangfang GAO^a, Xinyan XU^a

^a College of Media and International Culture, Zhejiang University, Hangzhou, Mainland China

Abstract

Trust is a key factor influencing the social adoption and diffusion of generative AI. Drawing on a survey ($N = 787$), this study employed a mixed-methods approach that combines structural equation modeling (SEM) and fuzzy-set qualitative comparative analysis (fsQCA) to examine the mechanisms underlying public trust in generative AI in social media contexts. The SEM results indicate that social media exposure can promote both technical trust and system trust in generative AI by enhancing AI knowledge, strengthening AI efficacy, and reducing perceived privacy risks. However, social media exposure can only improve technical trust by reducing perceived social risks, while this pathway has no significant effect on system trust. The fsQCA analysis further reveals the configurational effects of these factors on trust in generative AI. The formation of system trust in generative AI is characterized by three

Fangfang GAO (Associate Professor). College of Media and International Culture, Zhejiang University. Research interests: science communication, digital media research. Email: ffgao@zju.edu.cn

Xinyan XU (Master's Student). College of Media and International Culture, Zhejiang University. Research interests: intelligent communication, health communication. Email: 22423015@zju.edu.cn

Article History: Received on 22 January 2025. Accepted on 2 July 2026.

Communication and Society, 78 (2026)

configurations, including efficacy-driven/risk-averse, knowledge-driven/risk-averse, and social media-empowered. The configurations for technical trust are more diverse, additionally incorporating a media-driven/risk-averse configuration beyond the three configurations identified for system trust. This study expands the conceptual framework of trust in generative AI, clarifies the different formation mechanisms of technical trust and system trust, and offers practical suggestions for policymakers and AI developers.

Keywords: trust in generative AI, social media exposure, AI knowledge, AI efficacy, AI risk perception

引言

生成式人工智能 (generative artificial intelligence, 以下簡稱「生成式AI」) 技術發展迅猛。ChatGPT等大語言模型的出現標誌著人工智能進入全新的發展階段。與早期基於規則或統計學習的AI相比, 生成式AI呈現出根本性差異。首先, 在技術特性上, 早期AI主要執行特定的認知任務, 而生成式AI作為深度學習的自然延伸, 超越了早期AI的能力範疇, 能夠生成文本、圖像、音頻等內容 (Kumar et al., 2025; Mariani et al., 2023)。其次, 在交互模式上, 生成式AI支持自然語言的雙向交互, 但其湧現能力和概率性輸出也帶來新的不確定性。最後, 在應用領域和社會影響上, 早期AI局限於專業領域應用, 而生成式AI已全面滲透到人們的日常生活中, 其生成內容的真實性、潛在偏見、責任歸屬等問題對社會的影響更為直接和廣泛 (Dwivedi et al., 2023; Kumar et al., 2025)。這些差異對AI信任的形成機制提出新的挑戰。

當前, 中國公眾對生成式AI呈現期望與憂慮並存的矛盾心態 (喻國明等, 2024)。一方面, 生成式AI作為新質生產力在信息、教育、醫療等領域展現出巨大潛力; 另一方面, AI幻覺、數據隱私、算法偏見、虛假信息等伴生風險亦引發廣泛擔憂。在此背景下, 信任成為影響生成式AI社會應用的關鍵因素 (張樂、李森林, 2023)。然而, 現有AI信任研究多聚焦於傳統AI系統, 對生成式AI的信任探討仍顯匱乏 (Glikson & Woolley, 2020)。且多數研究側重於識別單一影響因素, 忽視了各因素間的交互作用與組合效應 (Henrique & Santos, 2024)。此外, 本土語境的缺失也削弱了相關研究結論在中國實踐中的適用性與指導價值。

為回應上述研究缺口, 本研究試圖整合媒介接觸、內在能力、外在風險這三個關鍵維度, 系統考察中國語境下公眾對生成式AI信任的形成機制。該整合性分析框架的核心邏輯在於, 首先, 生成式AI的技術複雜性使個體能力成為信任的前提條件。與傳統技術不同, 生成式AI的有效使用需掌握提示詞工程 (prompts) 等新技能。技術門檻的存在使得效能感和知識水平等可能影響公眾對生成式AI的理解、使用乃至信任建構。其次, 生成式AI的潛在風險已從技術與隱私層面擴展至深

《傳播與社會學刊》，（總）第78期（2026）

度偽造、信息操縱等社會倫理層面，這要求公眾進行更為複雜的風險與信任評估。最後，在當前的媒介化社會中，社交媒體不僅是公眾接觸和了解生成式AI的重要渠道，更通過信息傳播、社會學習和觀點交流等影響著公眾的能力建設與風險認知，進而作用於信任形成。基於此，本研究將系統檢驗社交媒體接觸如何通過作用於個體的內在能力與外在風險評估來影響公眾對生成式AI的信任，以及媒介因素與個體因素如何共同形塑對生成式AI的信任。

為深入探究這一複雜機制，本研究立足科學傳播，將公眾對生成式AI的信任視為其對新興科技信任的一種形態，並區分技術信任（對AI技術本身的信任）和系統信任（對AI治理體系的信任）兩個維度，以期把握此種信任所蘊含的技術創新和社會治理雙重意涵。與此同時，公眾對生成式AI信任的形成兼具線性因果與多重併發因果的特徵，單一方法難以完整刻劃。一方面，社交媒體接觸、AI知識、AI效能感、感知隱私風險與社會風險等因素對信任的獨立效應與中介路徑尚待釐清，這類淨效應與中介機制的探究宜採用SEM加以檢驗。另一方面，現實中高信任的形成往往並非源於某一因素的線性驅動，而是多個前因條件併發聯動的結果。不同個體可能經由不同的條件組合到達相同的信任狀態。這種殊途同歸的等效性與條件間的組態效應難以通過SEM捕捉（Ragin, 2008），而fsQCA恰能識別多條通往高信任的充分性路徑及條件要素間的替代與互補關係（Fiss, 2011）。更重要的是，兩種方法的並置不僅形成互補，更提供結論上的對照與深化，從而更立體地揭示信任形成的複雜邏輯。基於上述混合研究框架，本研究旨在闡明中國語境下公眾對生成式AI信任的形成機制，為提升公眾科學素養與優化AI信任環境提供理論依據與實踐啟示。

文獻綜述

對生成式AI的信任

信任作為信任者(trustor)對被信任者(trustee)行為的主觀預期(Coleman, 1990)，可以指向個人、技術、機構或抽象系統(唐長樂、張

社交媒體環境下公眾對生成式AI信任的形成機制研究

曉娟, 2025)。在科學傳播中, 信任不僅構成連接科學與公眾的橋梁, 更是確保技術在社會中被有效接受和理解的核心要素 (Fiske & Dupree, 2014)。當前, 公眾對生成式AI的信任已成為其社會應用與推廣的關鍵所在 (Araujo et al., 2023)。

值得注意的是, 生成式AI信任呈現出有別於傳統AI信任的獨特性。與聚焦於特定任務執行的傳統AI不同, 生成式AI具有內容創造能力、交互自主性和輸出不確定性等特徵 (Dwivedi et al., 2023), 這為其可靠性評估帶來新挑戰。既有的AI信任研究多聚焦技術維度, 關注技術的可靠性、功能性和實用性 (Mcknight et al., 2011), 以及公眾對技術不確定性的接受意願 (齊玥等, 2024)。然而, 科學信任並不全然遵循人際信任模式 (Hoff & Bashir, 2015)。生成式AI的創造性輸出和類人交互特性模糊了人機邊界, 其黑箱特徵和潛在的偏見、幻覺等問題進一步增加了信任評估的複雜性 (Bodó, 2021)。對生成式AI的信任不能僅局限於技術層面, 還需納入對技術背後專家系統、規則體系等的考量 (趙婷婷等, 2024)。這種系統信任既反映了對技術開發和設計者的信任, 也反映了對技術監管體系的信賴 (鄭紅娥、張艷敏, 2008)。前人研究表明, 公眾對生成式AI技術的信任與其對AI技術公司的組織架構、資源調度、開發者專業素養及制度規範的信任密切相關 (陳慧敏等, 2024)。當公眾對新技術的開發者、設計者及技術監管機構的信任程度較高時, 會更傾向於信任新技術並降低風險顧慮 (Araujo et al., 2023; Sztompka, 1999)。基於此, 本研究借鑒前人所建立的生成式AI複合信任框架 (陳慧敏等, 2024), 從技術信任和系統信任兩個維度來考察公眾對生成式AI的信任, 並假設了兩者之間的相關性:

H1: 公眾對生成式AI的技術信任與其對生成式AI的系統信任正相關。

既有研究表明, AI信任受社會文化背景、使用情境、技術性能與個體特徵等因素影響 (詹海寶, 2024; Marcus & Davis, 2019; Montag et al., 2024; Oksanen et al., 2020)。基於此, 本研究聚焦社交媒體接觸、個體內在能力與外在風險感知三個核心維度, 系統考察其對生成式AI信任的影響路徑與作用機制。具體而言, 作為當前公眾了解新興科技

《傳播與社會學刊》，(總)第78期(2026)

的關鍵信息場域，社交媒體不僅影響公眾對新興技術的認知框架，也塑造著其對新技術的信任態度(Brossard & Scheufele, 2013)。個體的內在能力，如知識和效能感，構成了信任建構的積極驅動力。外在風險感知，如社會風險和隱私風險感知，則是信任的潛在抑制因素。這三個維度在媒介化環境中動態交織。社交媒體的信息環境不僅可能直接影響信任，也可能通過公眾的內在能力與外在風險評估產生間接影響(Scheufele & Lewenstein, 2005)。因此，本研究試圖整合這三個維度，深入探討其對生成式AI信任的聯合作用機制，並進一步檢驗其對系統信任與技術信任兩個維度的差異化影響路徑。

AI效能感、AI知識與對生成式AI的信任

科學能力(scientific competence)是公眾對科學信任的重要基礎(Zhu, 2019)。其不僅包括科學知識，還涉及個體在特定情境中調動和運用資源以滿足複雜需求的能力，如科學態度、效能感等(OECD, 2005)。通常來說，個體的科學能力越強，其對科學的信任度也越高(Bauer et al., 2007)。在生成式AI語境下，這一邏輯同樣適用。本研究將AI知識與AI效能感視作個體內在科學能力的重要組成部分，兩者共同奠定了公眾信任生成式AI的能力基礎。

知識被視為公眾信任科學的重要條件。在公眾理解科學(public understanding of science, PUS)的框架下，知識不僅是公眾理解和應用科學的基礎，也是增強公眾對科學信任的推力(英國皇家學會，2004)。既有研究發現，科學知識水平，特別是主觀科學知識水平正向影響人們的科學信任(游淳惠、金兼斌，2020；楊恆、金兼斌，2022)。生成式AI作為一項涉及多學科交叉的複雜前沿技術，其設計原理、計算過程和實現路徑具有一定的黑箱性。只有當公眾掌握了一定的生成式AI知識時，才能更理性地認識其運行機制和發展規律、客觀評估其風險，進而產生信任(張樂、李森林，2023)。本研究據此提出以下假設：

H2a：公眾的AI知識水平與其對生成式AI的技術信任正相關。

H2b：公眾的AI知識水平與其對生成式AI的系統信任正相關。

社交媒體環境下公眾對生成式AI信任的形成機制研究

除了AI知識外，AI效能感也是個體內在科學能力的重要組成部分。AI效能感(AI efficacy)是個體對自身在理解、使用生成式AI方面能力和信心的主觀評估(Liu et al., 2024)。前人研究發現，AI效能感越高的個體，往往更傾向於信任生成式AI(詹海寶，2024)。因此，本研究提出以下假設：

H3a：公眾的AI效能感與其對生成式AI的技術信任正相關。

H3b：公眾的AI效能感與其對生成式AI的系統信任正相關。

AI風險感知與對生成式AI的信任

風險感知被視為信任的重要前因。在存在風險和不確定性的情況下，信任體現的正是對風險和不確定性的認知與接納能力(Mcknight et al., 2011)。公眾對生成式AI的風險感知主要聚焦於該技術在私人領域和公共領域可能帶來的潛在威脅與負面影響(Schwesig et al., 2023)。

在私人領域，公眾對生成式AI的風險感知多聚焦於隱私安全問題。生成式AI存在的不可預測性、算法透明度不足、道德準則模糊、數據需求及潛在濫用可能性等問題，正在加劇用戶對隱私風險的擔憂(Zarifis et al., 2021)。生成式AI需要大量數據進行訓練和優化，因此可能導致平台過度收集、推斷甚至濫用用戶的個人信息，進而侵犯隱私、降低信任(Mazurek & Małagocka, 2019)。因此，平台能否僅收集必要數據並確保其安全成為公眾評估AI隱私風險的核心關切(Zhou, 2011)。用戶的控制權缺失及相關法律與監管的滯後可能進一步放大用戶的隱私風險感知，讓其對生成式AI的安全性、可靠性及相關運行和監管機制的可信賴度產生懷疑。因此，本研究提出以下假設：

H4a：公眾對生成式AI隱私風險的感知與其對生成式AI的技術信任負相關。

H4b：公眾對生成式AI隱私風險的感知與其對生成式AI的系統信任負相關。

在公共領域，生成式AI可能引發虛假信息大規模傳播、數字鴻溝進一步擴大、算法偏見與歧視問題加劇及就業市場結構性變革等社會

《傳播與社會學刊》，（總）第78期（2026）

風險 (Gerlich, 2023)。公眾對生成式AI社會風險的感知可能降低其對該技術在透明度、可解釋性及安全性等方面的正面評價，從而削弱對該技術的整體信任 (張樂、李森林, 2023)。同時，公眾對生成式AI潛在社會風險的感知會削弱其對該技術開發者、提供平台及監管系統的信任 (宋艷等, 2022)。本研究據此假設：

H5a：公眾對生成式AI社會風險的感知與其對生成式AI的技術信任負相關。

H5b：公眾對生成式AI社會風險的感知與其對生成式AI的系統信任負相關。

社交媒體接觸與對生成式AI的信任

媒介不僅是傳遞信息的渠道，更是形塑公眾科學認知的重要力量 (Nisbet et al., 2003)。隨著人們日益依賴社交媒體獲取科學相關信息，社交媒體接觸已成為塑造公眾科學信任的關鍵機制 (Brossard & Scheufele, 2013)。針對生成式AI，一方面，社交媒體通過科普內容、專家解讀和案例分析，幫助公眾更好地理解該技術，從而增強信任；另一方面，AI開發者、平台及監管部門等通過社交媒體與公眾直接互動，及時釋疑，進一步提升信任 (Nisbet et al., 2002)。隨著生成式AI嵌入各社交媒體平台，用戶在日常使用過程中與這些技術的持續互動也有利於對生成式AI的信任形成 (Gefen et al., 2003)。據此，本研究提出以下假設：

H6a：公眾的社交媒體接觸與其對生成式AI的技術信任正相關。

H6b：公眾的社交媒體接觸與其對生成式AI的系統信任正相關。

前人研究發現，公眾通過社交媒體進行的非正式學習能顯著提升其科學知識水平，繼而增強科學信任 (游淳惠、金兼斌, 2020)。社交媒體可通過傳播科學信息推動技術普及，或通過成功的技術應用案例和實踐經驗分享增強公眾對新技術的使用信心與效能感 (Chen & Bryer, 2012)。研究據此提出以下假設：

H7a：公眾的社交媒體接觸與其AI知識水平正相關。

H7b：公眾的社交媒體接觸與其AI效能感正相關。

社交媒體環境下公眾對生成式AI信任的形成機制研究

在風險感知方面，風險的社會放大理論 (social amplification of risk framework) 認為，媒介作為一種社會機制能夠放大或縮小公眾對風險的感知與反應 (Kasperson et al., 1988)。當前公眾對生成式AI社會影響的擔憂在很大程度上受媒介敘事影響 (Gilardi et al., 2024)。既有研究發現相較於美國的社交媒體，中國社交媒體上有關生成式AI風險的公共話語較少關注隱私風險，而更多強調生成式AI的積極影響，這種傾向與專業媒體機構高度一致 (Zhang et al., 2024)。當公眾通過社交媒體接觸到更多關於生成式AI的正面評價時，其對生成式AI的風險感知會顯著降低 (劉鳴箏、王碩，2024)。據此，本研究提出以下假設：

H8a：公眾的社交媒體接觸與其對生成式AI社會風險的感知負相關。

H8b：公眾的社交媒體接觸與其對生成式AI隱私風險的感知負相關。

組態視角下的信任形成

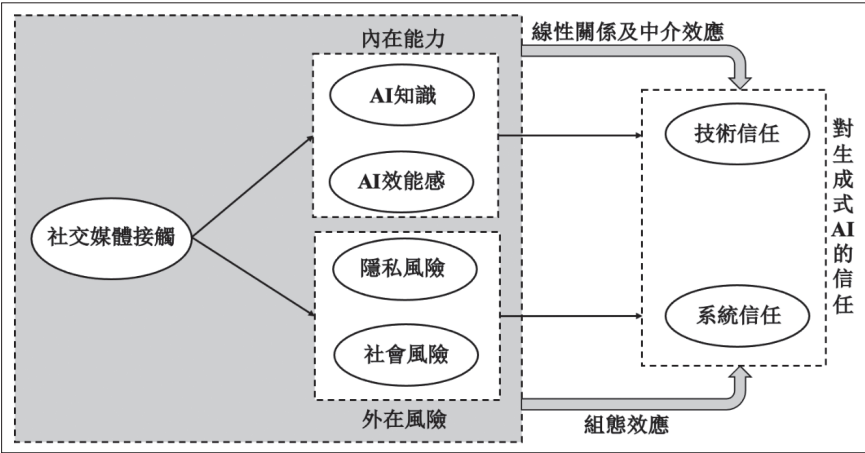
H1到H8聚焦各前因變量對生成式AI信任的獨立淨效應。然而，信任的形成本質上是多因併發、複雜交織的過程。人類對機器的信任受環境因素、系統特徵與個體差異的交互影響，任何單一類別的因素都不足以解釋信任的形成 (Hoff & Bashir, 2015)。本研究所涉及的社交媒體接觸、AI知識、AI效能感、感知隱私風險與感知社會風險分別對應媒介因素、個體的內在能力與外在風險等不同維度，構成影響信任形成的多維條件空間。這些條件之間可能存在替代與互補關係 (Fiss, 2011; Ragin, 2008)。這種交互組合的影響方式凸顯了從組態視角考察公眾對生成式AI信任的必要性。因此，本研究提出以下研究命題 (proposition)：

研究命題：社交媒體接觸、AI知識、AI效能感、感知隱私風險與感知社會風險的組態效應顯著影響公眾對生成式AI的技術信任與系統信任，且存在多條等效的差異化路徑驅動高信任。

《傳播與社會學刊》，（總）第 78 期（2026）

綜上所述，本研究關注的前因變量包括社交媒體接觸、AI 知識、AI 效能感、感知隱私風險與感知社會風險，結果變量包括公眾對生成式 AI 的技術信任與系統信任。研究旨在檢驗上述前因變量對兩類信任的淨效應與組態效應，以期更全面地揭示公眾對生成式 AI 信任的形成機制。研究框架如圖一所示。

圖一 研究框架



研究方法

數據收集

本研究採用問卷調查法，於 2024 年 11 月至 12 月通過自有超過三百萬全國樣本庫的「見數」平台發放問卷。基於該平台的「數據市集」，研究通過設置信用分和採納率、限制同一 IP 作答次數及最高答題次數等方式確保數據質量。在剔除作答時間過短、連續選擇相同選項及未通過注意力測試的問卷後，最終回收有效問卷 787 份。被調查者的基本信息如表一所示，性別分佈較為均衡，年齡上集中於中青年範圍，受教育程度較高，是當前社會中關注和使用生成式 AI 的主力人群，因此該樣本符合研究的基本需求。

社交媒體環境下公眾對生成式AI信任的形成機制研究

表一 研究樣本的基本信息 (N = 787)

變量	類型	人數	百分比 (%)
性別	男	395	50.19%
	女	392	49.81%
年齡	18–25 歲	280	35.58%
	26–30 歲	107	13.60%
	31–40 歲	193	24.52%
	41–50 歲	126	16.01%
	51 歲及以上	81	10.29%
學歷	本科以下學歷	132	16.77%
	大學本科	424	53.88%
	碩士研究生及以上	231	29.35%

變量測量

本研究的因變量為生成式AI信任，包括系統信任 and 技術信任兩個維度。借鑒前人的研究 (Pathak & Bansal, 2024)，基於7級李克特量表 (1 = 完全不同意，7 = 完全同意)，系統信任採用兩個題項測量，包括：(1) 我相信生成式AI的監管機構；(2) 我相信生成式AI技術/設備的開發者和提供平台 (Cronbach’s $\alpha = .71$, $M = 5.27$, $SD = .94$)；技術信任採用四個題項測量，包括：(1) 我相信生成式AI的能力；(2) 我相信生成式AI的可靠性；(3) 我相信生成式AI會按照其承諾的那樣運行；(4) 我覺得使用和依靠生成式AI是安全的 (Cronbach’s $\alpha = .81$, $M = 5.34$, $SD = .86$)。

本研究的自變量為社交媒體接觸。基於7級李克特量表 (1 = 從未，7 = 非常頻繁)，研究採用四個題項測量了被調查者在過去12個月中通過社交網絡 (如微博、微信、小紅書等)、視頻平台 (如抖音、快手等) 被動瀏覽、主動獲取生成式AI相關信息的頻率 (Cronbach’s $\alpha = .81$, $M = 5.15$, $SD = 1.04$)。

本研究的中介變量包括AI效能感、AI知識、AI社會風險和AI隱私風險感知。所有測量均基於7級李克特量表進行 (1 = 完全不同意，7 = 完全同意)。其中，對AI效能感的測量，參考了前人研究 (Wang & Chuang, 2024)，包括以下四個題項：(1) 如果接受了必要的培訓，我有信心學習和掌握簡單的生成式AI使用；(2) 我能利用生成式AI技術來

《傳播與社會學刊》，（總）第78期（2026）

實現想做的事；(3) 我能愉快、順暢地與生成式AI交互；(4) 我能理解和掌握生成式AI相關的術語 (Cronbach's $\alpha = .81$, $M = 5.56$, $SD = .79$)。

對AI知識的測量改編自前人的研究 (朱依娜、邱紫薇, 2023; Zhan & Fu, 2024)，包括四個題項：(1) 我知道有關生成式AI的知識；(2) 我掌握有關生成式AI的基本常識；(3) 我能夠辨別出使用了生成式AI技術的產品和服務；(4) 我是朋友圈子裏的AI「專家」之一 (Cronbach's $\alpha = .87$, $M = 4.74$, $SD = 1.52$)。

對AI社會風險感知的測量改編自前人的研究 (Gerlich, 2023)，採用五個題項：(1) 生成式AI可能導致失業；(2) 生成式AI可能被用於惡意目的；(3) 生成式AI可能導致錯誤或虛假信息傳播；(4) 生成式AI使偏見和歧視長期存在；(5) 生成式AI存在意想不到的未知風險 (Cronbach's $\alpha = .92$, $M = 4.34$, $SD = 1.46$)。

對AI隱私風險感知的測量改編自前人的研究 (Zhou, 2011)，採用三個題項：(1) 生成式AI技術/產品可能洩露我的隱私；(2) 生成式AI技術/產品可能無法妥善保護我的個人信息；(3) 生成式AI技術/產品可能收集超出業務功能實際需要的信息 (Cronbach's $\alpha = .88$, $M = 3.05$, $SD = 1.15$)。

此外，問卷還收集了被調查者的性別、年齡、受教育程度和AI使用經驗作為控制變量。

數據分析方法

本研究採用SEM與fsQCA開展數據分析。SEM與fsQCA的組合在社會科學研究中已有較為成熟的應用實踐，能夠提供不同視角的分析結論與交叉驗證 (Pappas & Woodside, 2021)。SEM聚焦於各前因變量對結果變量的獨立淨效應及其顯著性，適用於檢驗本研究的H1至H8。fsQCA則以集合論與布爾代數為基本邏輯，旨在捕捉不同前因變量的協同作用及在特定條件下如何共同作用於結果變量 (Ragin, 2008)，適用於檢驗研究命題。

SEM與fsQCA的方法組合在既有研究中主要分為遞進式與並列式。遞進式先基於SEM識別關鍵變量，再基於fsQCA分析關鍵變量的

社交媒體環境下公眾對生成式AI信任的形成機制研究

綜合作用路徑；並列式則在兩種方法中納入相同的前因變量與結果變量(杜運周、賈良定，2017)。本研究採用並列式，即在SEM檢驗淨效應與中介路徑後，使用fsQCA納入全部前因變量，進一步挖掘社交媒體接觸、AI知識、AI效能感、感知隱私風險與社會風險等要素間的協同作用與聯動匹配對技術信任與系統信任的影響，並識別驅動高信任的不同前因條件組合，以此作為對SEM結果的視角補充。由此，本研究能夠更全面地揭示公眾對生成式AI信任形成的複雜性與多樣性。

研究發現

SEM分析

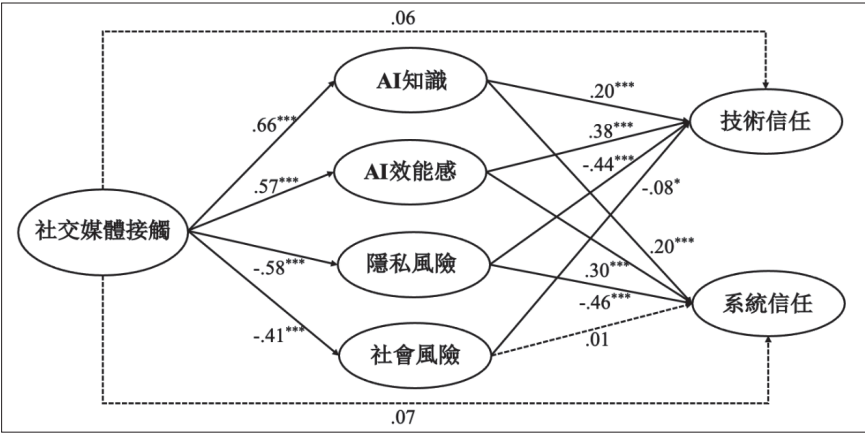
在進行SEM分析前，首先對各變量進行信效度檢測。在信度上，各變量的Cronbach’s α 值均大於.70。驗證性因子分析的結果顯示，各測量項的標準化因子載荷(standardized factor loading)值均大於.60，各變量的組合信度(composite reliability, CR)值均大於.70，平均提取方差(average variance extracted, AVE)值均大於等於.50，說明各變量內部測量項具有高度一致性。所有變量AVE值的平方根均大於該變量與其他變量間的相關係數，說明區分效度較好。此外，各變量兩兩相關均顯著，相關係數均未超過.80這一閾值(見表二)，說明變量間不存在多重共線性，可進行後續數據分析。

表二 收斂與區別效度及主要變量相關分析(N = 787)

	AVE	CR	1	2	3	4	5	6	7
1. 社交媒體接觸	.50	.80	.71						
2. AI效能感	.51	.80	.47**	.71					
3. AI知識	.68	.90	.54**	.55**	.82				
4. 社會風險	.69	.91	-.31**	-.21**	-.33**	.83			
5. 隱私風險	.70	.88	-.45**	-.45**	-.57**	.34**	.84		
6. 系統信任	.54	.70	.46**	.51**	.53**	-.29**	-.63**	.73	
7. 技術信任	.51	.80	.53**	.64**	.62**	-.36**	-.68**	.71**	.71

註：對角線粗體字為各變量AVE值的平方根，下三角為變量間皮爾遜相關係數。 ** $p < .01$ 。

圖二 SEM 結果（標準化係數）



註：* $p < .05$ ，** $p < .01$ ，*** $p < .001$ 。

接下來，基於最大似然 (maximum likelihood) 估計法，在控制年齡、性別、受教育程度和AI使用經驗的前提下，研究採用SEM檢測模型擬合度。結果顯示，模型擬合指標良好 ($\chi^2/df = 2.54$, CFI = .96, IFI = .96, TLI = .95, RMSEA = .04)。

從圖二的標準化路徑係數可以看出，媒介影響層面，社交媒體接觸正向影響AI知識 ($\beta = .66, p < .001$) 和AI效能感 ($\beta = .57, p < .001$)，同時負向影響隱私風險感知 ($\beta = -.58, p < .001$) 和社會風險感知 ($\beta = -.41, p < .001$)；但社交媒體接觸對系統信任 ($\beta = .07, p = .26$) 和技術信任 ($\beta = .06, p = .22$) 的直接影響不顯著。內在能力層面，AI知識正向影響技術信任 ($\beta = .20, p < .001$) 和系統信任 ($\beta = .20, p < .001$)；AI效能感也正向影響技術信任 ($\beta = .38, p < .001$) 和系統信任 ($\beta = .30, p < .001$)；且相較於AI知識，AI效能感對技術信任和系統信任的影響效力更大。外在風險層面，隱私風險感知負向影響技術信任 ($\beta = -.44, p < .001$) 和系統信任 ($\beta = -.46, p < .001$)，社會風險感知負向影響技術信任 ($\beta = -.08, p < .05$)，但對系統信任 ($\beta = .01, p = .93$) 影響不顯著；相較於社會風險感知，隱私風險感知對技術信任和系統信任的影響效力更強。此外，SEM分析結果顯示系統信任與技術信任之間的協方差估計值為.76 ($p < .001$)。因此，除了H5b、H6a和H6b外，其他所有研究假設均成立。

社交媒體環境下公眾對生成式AI信任的形成機制研究

為檢驗社交媒體接觸對生成式AI信任的間接影響，本研究在控制性別、受教育程度、年齡和AI使用經驗後，採用Bootstrap法對中介效應進行分析。如表三所示，針對技術信任，社交媒體接觸能夠通過提升AI知識水平 ($\beta = .08, 95\% \text{ CI} = [.03, .14]$) 和AI效能感 ($\beta = .10, 95\% \text{ CI} = [.07, .15]$)，或通過降低隱私風險感知 ($\beta = .13, 95\% \text{ CI} = [.09, .18]$) 和社會風險感知 ($\beta = .02, 95\% \text{ CI} = [.01, .03]$) 來予以提升；針對系統信任，社交媒體接觸能夠通過提升AI知識水平 ($\beta = .03, 95\% \text{ CI} = [.03, .23]$) 和AI效能感 ($\beta = .13, 95\% \text{ CI} = [.07, .20]$) 或通過降低隱私風險感知 ($\beta = .22, 95\% \text{ CI} = [.14, .31]$) 來予以提升。但社交媒體接觸通過社會風險感知對系統信任的間接影響不顯著 ($\beta = .00, 95\% \text{ CI} = [-.02, .02]$)。

表三 中介效應檢驗結果

中介效應檢驗	效應值	95%CI
技術信任		
社交媒體接觸→AI知識→技術信任	.08**	[.03, .14]
社交媒體接觸→AI效能感→技術信任	.10***	[.07, .15]
社交媒體接觸→隱私風險→技術信任	.13***	[.09, .18]
社交媒體接觸→社會風險→技術信任	.02**	[.01, .03]
系統信任		
社交媒體接觸→AI知識→系統信任	.03*	[.03, .23]
社交媒體接觸→AI效能感→系統信任	.13***	[.07, .20]
社交媒體接觸→隱私風險→系統信任	.22***	[.14, .31]
社交媒體接觸→社會風險→系統信任	.00	[-.02, .02]

註：* $p < .05$ ，** $p < .01$ ，*** $p < .001$ 。

模糊集定性分析

本研究利用fsQCA軟件中的calibrate函數進行數據校準。由於主要變量均基於7級李克特量表進行測量，需將其轉換為0到1之間的數據集，因此採用定點校準，將各變量錨點設定為5%（完全不隸屬）、50%（交叉點）和95%（完全隸屬）。完成校準後，對各前因變量（包括其非集）進行必要性分析。如表四所示，針對系統信任與技術信任這兩個結果變量，所有前因變量的一致性係數均小於.90，說明在本研究中不存在產生技術信任或系統信任的單一必要條件。

《傳播與社會學刊》，（總）第 78 期（2026）

表四 技術信任和系統信任的必要條件分析結果

技術信任			系統信任		
前因變量	一致性	覆蓋率	前因變量	一致性	覆蓋率
社交媒體接觸	.80	.77	社交媒體接觸	.78	.73
AI 效能感	.82	.81	AI 效能感	.80	.77
AI 知識	.81	.81	AI 知識	.79	.77
社會風險	.58	.60	社會風險	.61	.62
隱私風險	.51	.52	隱私風險	.54	.53
~ 社交媒體接觸	.49	.53	~ 社交媒體接觸	.52	.54
~ AI 效能感	.51	.53	~ AI 效能感	.54	.55
~ AI 知識	.51	.53	~ AI 知識	.54	.54
~ 社會風險	.73	.72	~ 社會風險	.71	.68
~ 隱私風險	.81	.83	~ 隱私風險	.81	.81

註：根據 fsQCA 軟件設定，變量前加上「~」符號表示該前因變量不存在，亦即布爾運算中「非」的關係。

接下來進行充分性分析。構建真值表時，設置案例閾值為 3，一致性閾值為 .80，一致性閾值 (proportional reduction in inconsistency) 閾值為 .75 (Park et al., 2017)，分析結果如表五和表六所示。單一解和總體解的一致性均高於 .80，總體覆蓋率高於 .70，達到較好的解釋水平。將具有相同核心條件的組態進行歸類發現，系統信任的組態路徑包括三類（見表五）。在效能驅動－風險規避型 (S1a, S1b) 中，AI 效能感為核心條件，隱私風險感知為核心缺失條件；在知識驅動－風險規避型 (S2) 中，AI 知識為核心條件，社交媒體接觸為輔助條件，隱私風險與社會風險感知為核心缺失條件；在社交媒體賦能型中，社交媒體接觸、AI 知識和 AI 效能感均為核心條件。技術信任的組態路徑包括四類（見表六）。除了與系統信任相似的效能驅動－風險規避型 (S1a, S1b)、知識驅動－風險規避型 (S2a, S2b) 及社交媒體賦能型 (S3) 外，還包括了以社交媒體接觸為核心條件、隱私風險和社會風險感知為核心缺失條件的媒介驅動－風險規避型 (S4)。

最後，由於 fsQCA 的分析結果對研究者設定的關鍵參數，如校準錨點、一致性閾值、頻數閾值等較為敏感 (Fiss, 2011)，因此有必要通過調整參數設定進行穩健性檢驗。本研究通過將一致性閾值從 .80 調整至 .90 和 .95，將案例數從 3 調整到 5 後分別再次進行組態分析，結果基本一致。此外，研究還將校準閾值從原始的 (.95, .50, .05) 調整為 (.80,

社交媒體環境下公眾對生成式AI信任的形成機制研究

.50, .20)，以及採用七級李克特量表常用的 (6, 4, 2) 的校準方式作為額外的穩健性檢驗 (Pappas & Woodside, 2021)。結果顯示核心組態未發生顯著變化，說明結論穩健。

表五 系統信任的組態路徑

組態條件	效能驅動－ 風險規避型		知識驅動－ 風險規避型	社交媒體 賦能型
	S1a	S1b	S2	S3
社交媒體接觸		•	•	●
AI知識			●	●
AI效能感	●	●		●
隱私風險	⊗	⊗	⊗	
社會風險	⊗		⊗	
原始覆蓋率 (raw coverage)	.56	.62	.52	.63
淨覆蓋率 (unique coverage)	.05	.01	.03	.04
一致性 (consistency)	.91	.92	.91	.89
解的覆蓋率 (solution coverage)			.75	
解的一致性 (solution consistency)			.86	

註：●和•分別表示核心條件和輔助條件存在；⊗和⊙分別表示核心條件和輔助條件缺席；
空格表示該條件可能存在亦可缺席。

表六 技術信任的組態路徑

組態條件	效能驅動－ 風險規避型		知識驅動－ 風險規避型		社交媒體 賦能型	媒介驅動－ 風險規避型
	S1a	S1b	S2a	S2b	S3	S4
社交媒體接觸		•		•	●	●
AI知識			●	●	●	
AI效能感	●	●	•		●	
隱私風險	⊗			⊗		⊗
社會風險	⊙	⊗	⊗			⊗
原始覆蓋率 (raw coverage)	.71	.56	.57	.62	.64	.56
淨覆蓋率 (unique coverage)	.05	.01	.01	.01	.02	.01
一致性 (consistency)	.93	.93	.93	.92	.93	.92
解的覆蓋率 (solution coverage)				.83		
解的一致性 (solution consistency)				.86		

註：●和•分別表示核心條件和輔助條件存在；⊗和⊙分別表示核心條件和輔助條件缺席；
空格表示該條件可能存在亦可缺席。

討論與結論

生成式AI的興起不僅帶來了技術範式的革新，更引發了人機信任的重構。本研究通過問卷調查（ $N = 787$ ），運用SEM與fsQCA的混合方法，系統考察了社交媒體環境下公眾對生成式AI信任的形成機制。研究發現，公眾在面對生成式AI時，能夠對技術本身的可靠性與制度體系的保障力進行區分性評估，進而形成差異化的信任判斷。SEM揭示了社交媒體接觸、AI知識、AI效能感、AI隱私風險感知與社會風險感知對技術信任與系統信任的直接與間接影響；fsQCA則進一步展現了這些因素的多種等效組合方式。二者相互印證、彼此補充，共同表明公眾對生成式AI的信任是在媒介接觸、內在能力建構與外在風險評估的動態交織中經由多元路徑形成的。深入理解這一形成機制對於在技術創新的浪潮中構建可持續的人機信任關係具有重要的理論價值與實踐意義。

研究的主要發現

第一，針對信任本身，研究發現公眾對生成式AI的信任可細分為系統信任與技術信任，兩者高度相關，同時保持相對獨立。這一發現不僅驗證了信任的多維性，更揭示了生成式AI既是獨立的技術工具，又因其開發者、監管者及使用者的多方參與而具備社會屬性。這使得公眾對生成式AI的信任不再局限於技術層面，而是呈現出技術信任與系統信任交織的複合結構。兩者之間的正相關關係在一定程度上契合信任轉移理論（Stewart, 2003; 2006），且在中國的社會文化語境下頗具討論價值。信任轉移理論認為，個體會將對某一對象（如個人、組織或系統）的信任延伸到與之相關的其他對象上（Stewart, 2003）。因此，在生成式AI情境下，公眾對技術本身的信任可能延伸到技術背後的開發者、提供平台與監管體系身上。反之，公眾對生成式AI技術背後系統的信任也可能延伸至技術（Saffarizadeh et al., 2024; Söllner et al., 2016）。但囿於橫截面數據的限制，本研究無法驗證兩種信任之間的雙向因果關係，未來研究可通過縱向設計予以驗證。

社交媒體環境下公眾對生成式AI信任的形成機制研究

前人研究發現，在集體主義文化影響下，人們更傾向於信任權威，認為由權威機構管理和背書的新技術更可靠 (Hofstede, 2001)。在中國這一深受集體主義文化影響且對制度高度信賴的社會中，公眾可能通過「誰在開發」和「誰在監管」來判斷技術是否可信。這既凸顯了系統信任在中國語境下對技術發展與擴散的戰略價值，也揭示了提升公眾對生成式AI的信任需要技術優化與制度建設並舉。

第二，在內在能力維度上，AI知識和AI效能感構成了AI信任的關鍵驅動力。SEM結果表明，AI知識和AI效能感均顯著正向影響系統信任 and 技術信任。這一發現呼應了PUS的理論傳統，驗證了公眾的知識水平和效能感在新技術接受與信任中的核心作用 (詹海寶, 2024)。路徑係數顯示，AI效能感對技術信任和系統信任的影響力均強於AI知識。這表明在生成式AI語境下，效能感比知識更能促進信任。這可能是因為生成式AI的黑箱特性和概率性輸出引發了公眾普遍的不確定感與控制焦慮。而當用戶確信自己能順暢駕馭生成式AI時，這種主觀控制感能有效緩解技術不透明性帶來的心理失衡，成為信任建構的支撐。同時，效能感不僅反映個體的信心和能力，更激發其主動探索和實踐的行為傾向。個體在邊用邊學 (learning by doing) 的過程中能強化掌控感，這種基於經驗的信心可能比知識更能轉化為穩固的信任基礎。值得注意的是，雖然SEM結果顯示AI效能感對技術信任與系統信任均具有較強的預測力，但組態分析表明AI效能感並非所有高信任路徑的必要條件。在知識驅動－風險規避型組態中，高AI知識與低風險感知共同構成了高信任的充分條件。這意味著當公眾對生成式AI擁有充分的認知和理解，且對技術風險與社會風險的感知較低時，知識儲備能夠在一定程度上替代AI效能感所提供的能力信心，為信任的形成提供認知基礎。這也說明通往AI信任的道路並非唯一，不同的能力資源之間可能存在替代關係。

第三，在外在風險維度上，SEM表明公眾對生成式AI的隱私風險感知對技術信任和系統信任均存在顯著的負向影響，而社會風險感知僅對技術信任產生負面影響，對系統信任的影響並不顯著。這種差異可能源於公眾對隱私風險和社會風險的責任歸因有所不同。在隱私風險方面，無論是生成式AI的數據處理機制、平台企業的數據管理，還

《傳播與社會學刊》，（總）第78期（2026）

是監管部門的隱私保護措施等都被認為需要對用戶的隱私安全負責。因此，隱私風險被視為技術與系統的共同責任，這可能導致隱私風險感知同時削弱技術信任和系統信任。相形之下，公眾更傾向於將失業、不平等等社會風險歸因於技術本身，而非治理失靈的問題。在這一歸因邏輯中，技術被視作社會風險的根源，而監管體系則常被視為與之無關，甚至成為潛在的問題解決力量。此外，系統信任通常建立在對技術開發者、提供平台和監管機構的長期依賴和認可之上，具有相對穩定性（陳慧敏等，2024）。因此系統信任不易因宏觀的社會風險感知而動搖。同時，路徑係數顯示，隱私風險感知對技術信任和系統信任的影響力均強於社會風險感知。這種差異可能源於其風險特徵的不同。公眾感知到生成式AI的隱私風險的威脅是個人化、直接且具體的，這使其成為影響信任的首要因素。相比之下，生成式AI帶來的社會風險雖然重要，但往往被感知為相對抽象和遠期的集體性風險（Shin, 2021），個體感受到的威脅緊迫性和相關性較低。根據解釋水平理論（construal level theory），低解釋水平的隱私風險促使個體進行具體歸因，將責任同時歸於技術和系統的具體失職，而高解釋水平的社會風險促使個體進行抽象歸因，更多歸因於技術發展的必然趨勢而非具體的治理失靈（Rim et al., 2009）。這一結果提示，不同類型的風險可能通過激活不同的認知歸因路徑對技術信任與系統信任產生差異化的影響。因此，在建構公眾對生成式AI信任的過程中，需針對不同的風險感知採取精準化的風險溝通策略。

組態分析的結果與此形成呼應。在多數高信任組態中，低風險感知均作為重要條件出現。以效能驅動–風險規避型、知識驅動–風險規避型為例，這兩類組態在技術信任與系統信任的形成中均穩定存在。其中，前者以高AI效能為核心，後者以高AI知識為核心，但兩者均需與低風險感知協同方能構成高信任的充分條件。這表明能力構建雖是信任形成的重要基礎，但僅憑能力提升不足以獨立驅動信任，還需與積極的風險評估形成互補。也就是說，要提高公眾對生成式AI的信任不應僅提升公眾對生成式AI的理解與掌控感，還需同時弱化其對生成式AI的負面風險感知。

社交媒體環境下公眾對生成式AI信任的形成機制研究

第四，在媒介影響上，研究發現社交媒體接觸是公眾信任生成式AI的重要促成條件。SEM結果顯示，社交媒體接觸雖不直接影響信任，卻能通過提高AI知識或增強AI效能感間接提升公眾對生成式AI的技術信任與系統信任。這表明社交媒體已成為AI信息傳播的重要渠道。當前中國公眾通過社交媒體接觸AI信息的頻率較高($M = 5.15$, $SD = 1.04$)。社交媒體不僅提供大量關於生成式AI的新聞和信息以支持公眾的學習，還通過嵌入AI功能讓用戶在日常使用中潛移默化地了解和掌握AI，從而加速技術馴化過程，促成了更深層的適應性信任(Brause & Blank, 2020)。然而，社交媒體通過風險感知影響信任的路徑在兩個信任維度上存在差異。社交媒體接觸能通過降低隱私風險感知和社會風險感知提升技術信任，卻僅能通過降低隱私風險感知提高系統信任；社會風險感知在社交媒體接觸與系統信任之間的中介效應不顯著。這可能是因為即便社交媒體上有關生成式AI的信息和討論能夠降低公眾對AI社會風險的感知，但若公眾認為現有的制度體系尚不健全、制度建設滯後或治理透明度不足，系統信任仍難以提升。需要注意的是，本研究樣本在年齡和教育程度上存在偏倚，以年輕和高學歷群體為主，這可能限制了結論向其他人群的推廣。未來研究需納入更廣泛的年齡層和教育背景的參與者，以進一步驗證本研究的發現。

fsQCA進一步說明了社交媒體接觸對信任的作用，識別出兩種包含社交媒體接觸的高信任組態。一種是在技術信任與系統信任中均存在的社交媒體賦能型。該組態體現了媒介環境(社交媒體接觸)與內在能力(AI知識和效能感)的協同效應，說明社交媒體接觸能夠與個體能力資源形成合力，共同推動對生成式AI信任的形成。另一種則是技術信任獨有的媒介驅動-風險規避型組態。儘管SEM結果顯示社交媒體接觸對技術信任的直接影響不顯著，但在該組態中，社交媒體接觸卻成為核心條件，與低隱私風險感知與低社會風險感知共同構成了高技術信任的充分條件。這說明，部分公眾可能並不具備較高的AI知識或效能感，但因身處信息豐富、基調積極的社交媒體環境，且對生成式AI的風險感知較低，同樣能夠形成技術信任。這種信任更多源於外部信息環境的塑造，而非內部認知加工的結果。也就是說，當個體的認知資

《傳播與社會學刊》，（總）第78期（2026）

源和認知能力不足時，媒介環境可作為技術信任生成的替代性機制。這體現了社交媒體接觸在技術信任形成中的獨特作用。這一發現也提示我們，社交媒體接觸的作用具有情境依賴性（context-dependent）。它並非普遍有效的獨立驅動因素，而是在與低風險感知搭配時方能發揮作用。換言之，針對低風險感知的群體，強化社交媒體上有關AI的信息傳播可能是提升其技術信任的有效策略。然而，值得警惕的是當前中國媒介有關生成式AI的話語更強調其積極影響，對潛在風險的討論相對較少（Zhang et al., 2024）。這種信息生態雖在一定程度上有助於降低公眾對生成式AI的風險感知、提高信任水平，但也可能導致虛假控制感與脆弱信任。因此，構建平衡、理性的媒介信息生態對於培育穩健的技術信任至關重要。

研究的理論與實踐意義

理論上，一方面，本研究拓展了生成式AI信任的概念內涵。不同於既往研究多將AI信任視為單一構念，本研究通過區分系統信任和技術信任，揭示了信任的多維度與複雜性。兩類信任雖互相關聯，但形成路徑仍有所差異。這有助於更好地理解複雜社會技術系統中的信任及其形成機制。另一方面，本研究整合了媒介環境、個體內在能力與外在風險感知三個維度，為中國語境下公眾對生成式AI信任的形成機制提供了整合性的分析框架。通過SEM與fsQCA的混合方法設計，研究兼顧了線性關係與組態效應兩種分析視角，從而為理解生成式AI信任的多元形成機制提供了更全面的解釋。

實踐上，本研究為提升公眾對生成式AI的信任提供了可行性建議。首先，鑒於系統信任與技術信任的形成路徑差異，應相應地採取差異化的信任建設策略。針對系統信任，需要強化制度建設的可見性，如監管機構定期發佈監管公告、AI服務提供商定期公開算法審查結果等。針對技術信任，則應注重技術能力的展示和技術透明度的提高，如建設可解釋性AI。大模型平台可效仿OpenAI公開其內部人工智能模型的安全評估結果，展示其模型在有害內容生成、模型越獄以及

社交媒體環境下公眾對生成式AI信任的形成機制研究

幻覺現象等方面的測試表現。其次，需要構建多層次的AI素養培育體系，並注重風險溝通的協同。一方面，要提升公眾對生成式AI的了解與應用能力。國家層面，應將AI素養納入全民教育體系，鼓勵和支持公眾學習AI知識和技術。平台層面，AI服務提供商應開發用戶友好的教育模塊，如新手引導、官方課程、工作坊和知識庫等，以幫助公眾掌握AI。例如，AI平台可幫助用戶學會根據自身需求和平台特點設計清晰、具體的提示詞(prompt)，以更精準地獲取所需信息。在社區層面，應建立面對不同群體的體驗和培訓機制，特別關注數字弱勢群體的需求。另一方面，要降低公眾對生成式AI風險的感知，需在社交媒體話語中平衡技術機遇與潛在挑戰的呈現，避免「技術烏托邦」或「AI威脅論」的極端敘事。最後，應充分發揮社交媒體在信任建構中的積極作用。社交媒體平台可通過算法優化推送高質量的生成式AI科普內容和實操培訓，同時建立針對AI生成內容的標註與核查機制。微信公眾平台對AI生成內容的標註要求、抖音對AI生成內容的規範管理等都是積極的探索方向。

研究局限與展望

首先，本研究採用的並非完全意義上的隨機抽樣，導致樣本偏向年輕化的高學歷群體，可能影響結論在中國語境下的普適性。未來研究應採用分層隨機抽樣等更嚴格的抽樣方法，特別關注中老年、低學歷等群體對生成式AI的信任情況。其次，本研究採用的橫截面數據限制了對變量間因果關係的推斷，特別是系統信任與技術信任之間可能存在的雙向影響關係。未來可通過縱向追蹤或實驗研究深入探究信任轉移及動態演化過程。再次，除本研究關注的影響因素外，還有更多潛在的因素(如社會文化背景、技術特徵、個體特點等)可能影響公眾對生成式AI的信任。未來的研究可作進一步探索。最後，不同應用場景(如新聞生產、醫療諮詢、教育輔助)、不同文化背景下的AI信任機制可能存在差異，結合場景進行研究或開展跨文化比較研究將有助於深化對生成式AI信任的理解。

《傳播與社會學刊》，（總）第78期（2026）

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

本文未獲項目研究經費資助。

Funding

This article was not funded by any project grant.

ORCID

高芳芳 (Fangfang GAO) <https://orcid.org/0009-0002-7166-9046>

徐欣妍 (Xinyan XU) <https://orcid.org/0009-0005-1191-737X>

參考文獻

中文部分 (Chinese Section)

朱依娜、邱紫薇 (2023)。〈人工智能公眾態度：概念及測量研究綜述〉。《科學與社會》，第3期，頁127–138。

Zhu, Y., & Qiu, Z. (2023). Public attitudes towards artificial intelligence: A systematic review of conceptualization and measurement. *Science and Society*, 3, 127–138.

杜運周、賈良定 (2017)。〈組態視角與定性比較分析 (QCA)：管理學研究的一條新道路〉。《管理世界》，第6期，頁155–167。

Du, Y., & Jia, L. (2017). Configuration perspective and Qualitative Comparative Analysis (QCA): A new road for management research. *Journal of Management World*, 6, 155–167.

社交媒體環境下公眾對生成式AI信任的形成機制研究

- 宋艷、陳琳、李琴、何嘉欣、汪悅 (2022)。〈人工智能倫理風險感知、信任與公眾參與〉。《科學學研究》，第7期，頁1153–1162。
- Song, Y., Chen, L., Li, Q., He, J., & Wang, Y. (2022). AI ethical risk perception, trust and public participation. *Studies in Science of Science*, 7, 1153–1162.
- 英國皇家學會 (2004)。《公眾理解科學》(唐英英譯)。北京理工大學出版社。
(原書The Royal Society [1985]. *The public understanding of science*. The Royal Society.)
- The Royal Society (2004). *The public understanding of science* (Y. Tang, Trans.). Beijing Institute of Technology Press. (Original work published 1985)
- 唐長樂、張曉娟 (2025)。〈建立數字信任：數據要素可信流通的信任基礎、內在邏輯與實現框架〉。《圖書館論壇》，第9期，頁136–145。
- Tang, C., & Zhang, X. (2025). Building digital trust: Trust foundation, intrinsic logic, and implementation framework for the circulation of trusted data elements. *Library Tribune*, 9, 136–145.
- 陳慧敏、朱姿伊、金兼斌 (2024)。〈知識生產視域下生成式人工智能信任的複合框架、作用機制與未來進路〉。《中國編輯》，第12期，頁60–69。
- Chen, H., Zhu, Z., & Jin, J. (2024). Integrated framework, mechanisms, and future pathways of trust in generative artificial intelligence from a knowledge production perspective. *Chinese Editors Journal*, 12, 60–69.
- 張樂、李森林 (2023)。〈知識、理解與信任：個體對人工智能的信任機制〉。《社會學評論》，第3期，頁59–83。
- Zhang, L., & Li, S. (2023). Knowledge, understanding and trust: Mechanisms of individual trust in artificial intelligence. *Sociological Review of China*, 3, 59–83.
- 游淳惠、金兼斌 (2020)。〈新媒體環境下科學知識對爭議性科技態度的影響——以轉基因為例〉。《國際新聞界》，第5期，頁81–98。
- You, C., & Jin, J. (2020). The impact of scientific knowledge on attitudes towards controversial technology in China: A case study on Genetically Modified Organisms (GMO). *Chinese Journal of Journalism and Communication*, 5, 81–98.
- 喻國明、蘇芳、金麗萍 (2024)。〈缺席的對話：大語言模型的認知想像與差異彌合〉。《現代出版》，第1期，頁20–35。
- Yu, G., Su, F., & Jin, L. (2024). Absent conversations: Cognitive imaginations and bridging differences in large language models. *Modern Publishing*, 1, 20–35.
- 楊恆、金兼斌 (2022)。〈知識缺失還是知識自負？知識水平對公眾食品安全信心的影響研究〉。《華中農業大學學報(社會科學版)》，第6期，頁172–183。

《傳播與社會學刊》，（總）第78期（2026）

Yang, H., & Jin, J. (2022). Deficit or overconfidence? The effect of knowledge on public confidence in food safety in China. *Journal of Huazhong Agricultural University (Social Sciences Edition)*, 6, 172–183.

詹海寶 (2024)。〈個體為何信任生成式人工智能？——一項基於扎根理論的質性研究〉。《新媒體與社會》，第2期，頁135–146。

Zhan, H. (2024). Why do individuals trust generative AI? A qualitative study based on rootedness theory. *New Media and Society*, 2, 135–146.

趙婷婷、劉建飛、高婷婷 (2024)。〈人智信任對交互策略類知識貢獻行為的影響研究——基於生成式AI社區〉。《情報科學》，第11期，頁158–166。

Zhao, T., Liu, J., & Gao, T. (2024). The influence of human-AI trust on interaction strategy knowledge contribution behavior —— Based on the generative AI community. *Information Science*, 11, 158–166.

齊玥、陳俊廷、秦邵天、杜峰 (2024)。〈通用人工智能時代的人與AI信任〉。《心理科學發展》，第12期，頁2124–2136。

Qi, Y., Chen, J., Qin, S., & Du, F. (2024). Human-AI mutual trust in the era of artificial general intelligence. *Advances in Psychological Science*, 12, 2124–2136.

鄭紅娥、張艷敏 (2008)。〈論系統信任：關於中國信任問題的思考〉。《江淮論壇》，第1期，頁116–122。

Zheng, H., & Zhang, Y. (2008). System trust: Reflections on the issue of trust in China. *Jianghuai Tribune*, 1, 116–122.

劉鳴箏、王碩 (2024)。〈技術想像與技術實踐的交織——科學傳播視域下公眾對生成式人工智能的認知、評估與使用〉。《國際新聞界》，第4期，頁32–51。

Liu, M., & Wang, S. (2024). The interweaving of technological imagination and technological practice: The public's cognition, evaluation, and use of generative artificial intelligence from the perspective of scientific communication. *Chinese Journal of Journalism and Communication*, 4, 32–51.

英文部分 (English Section)

Araujo, T. B., Brosius, A., Goldberg, A. C., Möller, J. E., & de Vreese, C. H. (2023). Humans vs. AI: The role of trust, political attitudes, and individual characteristics on perceptions about automated decision making across Europe. *International Journal of Communication*, 17, 6222–6249.

社交媒體環境下公眾對生成式AI信任的形成機制研究

- Bauer, M. W., Allum, N., & Miller, S. (2007). What can we learn from 25 years of PUS survey research? Liberating and expanding the agenda. *Public Understanding of Science*, 16(1), 79–95.
- Bodó, B. (2021). Mediated trust: A theoretical framework to address the trustworthiness of technological trust mediators. *New Media and Society*, 23(9), 2668–2690.
- Brause, S. R., & Blank, G. (2020). Externalized domestication: Smart speaker assistants, networks and domestication theory. *Information, Communication and Society*, 23(5), 751–763.
- Brossard, D., & Scheufele, D. A. (2013). Science, new media, and the public. *Science*, 339(6115), 40–41.
- Chen, B., & Bryer, T. (2012). Investigating instructional strategies for using social media in formal and informal learning. *The International Review of Research in Open and Distributed Learning*, 13(1), 87–104.
- Coleman, J. S. (1990). *Foundations of social theory*. Harvard University Press.
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., ... Wright, R. (2023). “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, 71, 102642.
- Fiss, P. C. (2011). Building better causal theories: A fuzzy set approach to typologies in organization research. *Academy of Management Journal*, 54(2), 393–420.
- Fiske, S. T., & Dupree, C. (2014). Gaining trust as well as respect in communicating to motivated audiences about science topics. In *Proceedings of the National Academy of Sciences*, 111 (Supplement 4), 13593–13597.
- Gefen, D., Karahanna, E., & Straub, D. W. (2003). Trust and TAM in online shopping: An integrated model. *MIS Quarterly*, 27(1), 51–90.
- Gerlich, M. (2023). Perceptions and acceptance of artificial intelligence: A multi-dimensional study. *Social Sciences*, 12(9), 502.
- Gilardi, F., Kasirzadeh, A., Bernstein, A., Staab, S., & Gohdes, A. (2024). We need to understand the effect of narratives about generative AI. *Nature Human Behaviour*, 8(12), 2251–2252.
- Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627–660.
- Henrique, B. M., & Santos, E. (2024). Trust in artificial intelligence: Literature review and main path analysis. *Computers in Human Behavior: Artificial Humans*, 2(1), 100043.

- Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57(3), 407–434.
- Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations* (2nd ed.). Sage Publications.
- Kasperson, R. E., Renn, O., Slovic, P., Brown, H. S., Emel, J., Goble, R., Kasperson, J. X., & Ratick, S. (1988). The social amplification of risk: A conceptual framework. *Risk Analysis: An International Journal*, 8(2), 177–187.
- Kumar, A., Shankar, A., Hollebeek, L. D., Behl, A., & Lim, W. M. (2025). Generative artificial intelligence (GenAI) revolution: A deep dive into GenAI adoption. *Journal of Business Research*, 189, 115160.
- Liu, F., Makady, H., Nah, S., & McNealy, J. (2024). When citizens support AI policies: The moderating roles of AI efficacy on AI news, discussion, and literacy. *Journal of Information Technology and Politics*, 21(4), 493–509.
- Marcus, G., & Davis, E. (2019). *Rebooting AI: Building artificial intelligence we can trust*. Pantheon Books.
- Mariani, M. M., Machado, I., Magrelli, V., & Dwivedi, Y. K. (2023). Artificial intelligence in innovation research: A systematic review, conceptual framework, and future research directions. *Technovation*, 122, 102623.
- Mazurek, G., & Małagocka, K. (2019). Perception of privacy and data protection in the context of the development of artificial intelligence. *Journal of Management Analytics*, 6(4), 344–364.
- Mcknight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems (TMIS)*, 2(2), 1–25.
- Montag, C., Becker, B., & Li, B. J. (2024). On trust in humans and trust in artificial intelligence: A study with samples from Singapore and Germany extending recent research. *Computers in Human Behavior: Artificial Humans*, 2(2), 100070.
- Nisbet, M. C., Brossard, D., & Kroepsch, A. (2003). Framing science: The stem cell controversy in an age of press/politics. *Harvard International Journal of Press/Politics*, 8(2), 36–70.
- Nisbet, M. C., Scheufele, D. A., Shanahan, J., Moy, P., Brossard, D., & Lewenstein, B. V. (2002). Knowledge, reservations, or promise? A media effects model for public perceptions of science and technology. *Communication Research*, 29(5), 584–608.
- OECD (2005, May 27). *The definition and selection of key competencies: Executive Summary*. OECD. <https://www.deseco.ch/bfs/deseco/en/index/02.parsys.43469.downloadList.2296.DownloadFile.tmp/2005.dskcexecutivesummary.en.pdf>
- Oksanen, A., Savela, N., Latikka, R., & Koivula, A. (2020). Trust toward robots and artificial intelligence: An experimental approach to human-technology interactions online. *Frontiers in Psychology*, 11, 568256.

社交媒體環境下公眾對生成式AI信任的形成機制研究

- Pappas, I. O., & Woodside, A. G. (2021). Fuzzy-set qualitative comparative analysis (fsQCA): Guidelines for research practice in information systems and marketing. *International Journal of Information Management*, 58, 102310.
- Park, Y., El Sawy, O. A., & Fiss, P. C. (2017). The role of business intelligence and communication technologies in organizational agility: A configurational approach. *Journal of the Association for Information Systems*, 18(9), 648–686.
- Pathak, A., & Bansal, V. (2024). AI as decision aid or delegated agent: The effects of trust dimensions on the adoption of AI digital agents. *Computers in Human Behavior: Artificial Humans*, 2(2), 100094.
- Ragin, C. C. (2008). *Redesigning social inquiry: Fuzzy sets and beyond*. University of Chicago Press.
- Rim, S., Uleman, J. S., & Trope, Y. (2009). Spontaneous trait inference and construal level theory: Psychological distance increases nonconscious trait thinking. *Journal of Experimental Social Psychology*, 45(5), 1088–1097.
- Saffarizadeh, K., Keil, M., & Maruping, L. (2024). Relationship between trust in the AI creator and trust in AI systems: The crucial role of AI alignment and steerability. *Journal of Management Information Systems*, 41(3), 645–681.
- Scheufele, D. A., & Lewenstein, B. V. (2005). The public and nanotechnology: How citizens make sense of emerging technologies. *Journal of Nanoparticle Research*, 7, 659–667.
- Schwesig, R., Brich, I., Buder, J., Huff, M., & Said, N. (2023). Using artificial intelligence (AI)? Risk and opportunity perception of AI predict people's willingness to use AI. *Journal of Risk Research*, 26(10), 1053–1084.
- Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, 146, 102551.
- Söllner, M., Hoffmann, A., & Leimeister, J. M. (2016). Why different trust relationships matter for information systems users. *European Journal of Information Systems*, 25(3), 274–287.
- Stewart, K. J. (2003). Trust transfer on the World Wide Web. *Organization Science*, 14(1), 5–17.
- Stewart, K. J. (2006). How hypertext links influence consumer perceptions to build and degrade trust online. *Journal of Management Information Systems*, 23(1), 183–210.
- Sztompka, P. (1999). *Trust: A sociological theory*. Cambridge University Press.
- Wang, Y.-Y., & Chuang, Y.-W. (2024). Artificial intelligence self-efficacy: Scale development and validation. *Education and Information Technologies*, 29, 4785–4808.
- Zarifis, A., Kawalek, P., & Azadegan, A. (2021). Evaluating if trust and personal information privacy concerns are barriers to using health insurance that explicitly utilizes AI. *Journal of Internet Commerce*, 20(1), 66–83.

《傳播與社會學刊》，（總）第78期（2026）

- Zhan, Q., & Fu, L. (2024). Structural knowledge and subjective knowledge, not factual knowledge, promotes corrective and restrictive actions towards healthy eating misinformation in China: A multigroup comparison of extended cognitive mediation model based on altruism. *Current Psychology*, 43, 27886–27900.
- Zhang, R., Li, H., Chen, A., Liu, Z., & Lee, Y.-C. (2024). AI privacy in context: A comparative study of public and institutional discourse on conversational AI privacy in the US and Chinese social media. *Social Media + Society*, 10(4), 1–19.
- Zhou, T. (2011). The impact of privacy concern on user adoption of location-based services. *Industrial Management and Data Systems*, 111(2), 212–226.
- Zhu, Y. (2019). How Chinese students' scientific competencies are influenced by their attitudes? *International Journal of Science Education*, 41(15), 2094–2112.

本文引用格式

高芳芳、徐欣妍 (2026)。〈社交媒體環境下公眾對生成式AI信任的形成機制研究〉。《傳播與社會學刊》，第78期，頁111–142。

Citation of This Article

Gao, F., & Xu, X. (2026). Exploring public trust in generative AI and its formation mechanism on social media. *Communication and Society*, 78, 111–142.