

專輯論文

LLMs 模擬意見動態和偏見信息的影響： 基於「氣候變化」、「轉基因食品」和 「疫苗安全」議題的多代理研究

陶煒^a、向安玲^b

^a 華中農業大學文法學院廣告與傳播學系，武漢市，中國大陸

^b 中央民族大學新聞與傳播學院智能傳播系，北京市，中國大陸

摘要

在生成式人工智慧逐步介入資訊生產與傳播的背景下，經典傳播理論中有關有限理性、啟發式加工與意見極化的若干預設，亟需在新型認知主體條件下重新檢視。本研究探討大語言模型 (large language models, LLMs) 在模擬人類意見動態與科學議題公共傳播中的表現，特別關注偏誤資訊介入後群體科學共識的變化。研究以多個虛擬代理為對象，圍繞「氣候變化」、「轉基因食品」、「疫苗安全」三項具有科學共識的議題展開模擬測試。實驗一在無偏誤資訊條件下，觀察代理經

陶煒，華中農業大學文法學院廣告與傳播學系副研究員。研究興趣：智能廣告、大模型數據仿真模擬、大數據與網絡新媒體。電郵：wtao@mail.hzau.edu.cn
向安玲，中央民族大學新聞與傳播學院智能傳播系副教授。研究興趣：智能媒體 (AIGC 與媒介大數據)、網絡輿論、國際傳播。電郵：anlingxiang@muc.edu.cn
論文投稿日期：2024年8月25日。論文接受日期：2026年3月10日。

《傳播與社會學刊》，(總)第78期(2026)

由社交網絡進行日常互動後的信念狀態與傳播路徑；實驗二調整初始信念分佈，以分析初始條件對結果的影響；實驗三則引入10%代理發佈偏誤資訊，以檢驗其對信念動態的干擾。50個代理於30天內共生成194,699條對話。研究結合依存語義圖譜、Lotka-Volterra模型與混合效應模型進行分析。結果顯示，在無偏誤資訊情況下，LLM代理整體傾向與科學共識保持一致；即使在不同初始信念分佈下，各代理群體最終仍趨向科學共識，顯示模型具有一定穩定性。引入偏誤資訊後，意見分佈雖更趨多樣，且信念得分隨時間下降，但最終整體仍朝向科學共識發展。研究表明，LLM驅動的多代理系統可作為一種「可控社會模擬」工具，為理解科學共識形成及傳播理論重構提供方法論啟示。

關鍵詞：語言模型、意見動態、偏誤資訊、社會現象

Special Issue Article

Simulating Opinion Dynamics and the Impact of Biased Information Using LLMs: A Multi-Agent Study on Climate Change, GM Foods, and Vaccine Safety

Wei TAO ^a, Anling XIANG ^b

^aDepartment of Advertising and Communication, School of Humanities and Law, Huazhong Agricultural University, Wuhan, Mainland China

^bDepartment of Intelligent Communication, School of Journalism and Communication, Minzu University of China, Beijing, Mainland China

Abstract

Generative artificial intelligence has introduced new cognitive agents into information production and dissemination, calling for a reexamination of classical communication theories on bounded rationality, heuristic processing, and opinion polarization. This study explores the performance of large language models (LLMs) in simulating human opinion dynamics and public communication on scientific issues. Using multiple virtual agents, we examined

Wei TAO (Associate Research Fellow). Department of Advertising and Communication, School of Humanities and Law, Huazhong Agricultural University. Research interests: intelligent advertising, large-model data simulation, big data, new media on the Internet. Email: wtao@mail.hzau.edu.cn

Anling XIANG (Associate Professor). Department of Intelligent Communication, School of Journalism and Communication, Minzu University of China. Research interests: intelligent media (AIGC and media big data), online public opinion, international communication. Email: anlingxiang@muc.edu.cn

Article History: Received on 25 August 2024. Accepted on 10 March 2026.

Communication and Society, 78 (2026)

discussions on three topics with established scientific consensus: climate change, genetically modified foods, and vaccine safety. Across three experiments, we tested interactions under normal conditions, different initial belief distributions, and the introduction of misleading information by 10% of agents. In total, 50 agents generated 194,699 dialogues over 30 days. Results show that LLM agents generally tended toward scientific consensus. Although misleading information increased opinion diversity and lowered belief scores over time, the overall direction still moved toward consensus. The findings suggest that LLM-driven multi-agent systems can serve as a tool for controllable social simulation and provide insights into the formation of scientific consensus in intelligent media environments.

Keywords: language model, opinion dynamics, biased information, social phenomena

引言

隨著人工智慧 (artificial intelligence, AI) 技術的快速發展，人工智慧成為科技界關注的熱點話題，同時也為科學傳播帶來新的機遇與挑戰。理解個體在社會影響下改變意見的機制，對於多個研究領域都具有重要意義，從公共衛生宣傳、衝突調解到錯誤資訊的減輕(陳慧敏等，2023；Jiang et al., 2015)。對這些動態的準確建模，將使研究者能夠預測未來趨勢，如潛在的意見兩極化，並制定有針對性的干預措施，以減輕負面影響。儘管關於意見形成與資訊加工的傳播理論已發展數十年，諸如精細化可能性模型 (elaboration likelihood model, ELM)、啟發式—系統加工理論以及議程設定理論等經典框架，仍然深刻影響著學界對公眾意見演化機制的理解。然而，這些理論大多建立在人類認知有限性、情緒動機、社會壓力與注意力稀缺等前提之上，其解釋力在高度演算法化、智能化的資訊環境中正面臨新的挑戰。一方面，隨著資訊生產與互動主體逐漸從「人—人」擴展至「人—機」甚至「機—機」，傳統理論中有關啟發式線索依賴、從眾效應與議程競爭的核心機制是否仍然成立，尚缺乏可控且可重複的檢驗路徑；另一方面，這些理論在現實社會中往往難以被嚴格證偽，其經驗研究結果亦高度依賴具體情境與樣本結構，從而限制了理論前提本身的反思與修正空間。

基於代理的模型 (agent-based model, ABM) 是科學傳播中意見動態研究的基石 (林慧嫻、陸靜曄，2015；Epstein, 2012; Smaldino, 2023)。ABM 使用數學方程來表示個體，描述意見可能如何透過代理間的交流發生變化。當模擬代理互動時，這些計算可以闡明群體意見動態的演變。例如，當模擬代理在信念更新中包含確認偏誤時，它們往往會趨向於意見集群，而非達成共識 (Flache et al., 2017)。近年來，LLMs 在模擬複雜社會現象方面展現出巨大的潛力。根據近期有關生成代理群體的研究 (Park et al., 2023)，模擬跨主題的多代理對話並控制確認偏誤等因素、研究其對意見演變的影響，具有重要意義。不同於人類輿論形成，基於價值對齊的 LLM 代理或許不具備人類社群所謂「流量傾向性」或「沉默的螺旋效應」——即無論這些代理扮演何種角色，其行為不會體現出人類在面對事實性觀點時所表現出的抗拒與迴避。因此，

《傳播與社會學刊》，（總）第78期（2026）

建立更穩定且更有利於科學共識和真相傳播的LLM代理是可行的，但仍有待透過實驗加以證實與優化。

在此背景下，本文並不試圖對經典理論進行直接驗證或替代，而是將其視為一組以「人類有限理性」為核心假設的經典傳播理論譜系。LLM代理的引入，恰恰為重新檢視這些理論的隱含前提提供了一種方法論契機：當資訊互動主體具備高度語言生成與推理能力，卻不具備人類的情緒迴避、社會身分壓力與現實利益動機時，意見動態是否仍會呈現出極化、議程鎖定或錯誤資訊放大等經典現象？透過建構以LLM為基礎的多代理互動系統，本文嘗試在一個「去人類心理噪音」的認知環境中，考察意見演化機制本身的結構性特徵，從而反向揭示經典傳播理論在多大程度上依賴於特定的人類認知與社會條件。這一視角使本文的研究動機不僅源於技術可行性或資料可得性，更根植於傳播理論在智能媒介環境下面臨的解釋邊界與重構需求。

本研究透過實驗探討LLMs在模擬人類意見動態和社會現象（如極化和錯誤資訊傳播）中的表現。具體而言，研究關注三個主要問題：LLMs是否能夠準確模擬人類的意見動態、LLMs在模擬社會現象（如極化和錯誤資訊傳播）時的表現，以及加入偏誤資訊後，LLMs生成的意見動態是否會發生變化。透過使用GPT-4o模擬多個虛擬代理，並在不同初始條件和干預機制下進行實驗，研究期望揭示LLMs在社會科學研究中的潛在應用價值，並為後續相關研究提供實證支持。與既有研究多聚焦於短期、單議題場景下的觀點生成不同，本文強調建構一個多議題、長週期、具有語境變化的虛擬社交環境，觀察LLM代理在連續互動中如何對資訊進行再加工、如何發生觀點修正與立場演化。研究亦將關注生成文本的語義聚類結構，以探討語言模型驅動的社交系統是否會自然呈現出「語義趨同」與「共識聚合」的模式。這一嘗試為理解數位語境中科學資訊的傳播邏輯提供了一種新的模擬框架，也為社會語言行為建模提供了方法上的創新範式。這不僅在方法層面提供了將傳播理論與生成式人工智慧融合的可能性，更在理論層面為「人類意見如何在複雜語境中被建構、干預與傳播」這一核心議題提供了跨學科的理論延展與實踐啟發。

文獻綜述

意見動態與意見動力學

意見動態是指個體或群體的觀點、態度和信念在一定時間內的變化過程(張霽陽等, 2023)。意見的形成、傳播和變化是一個複雜過程, 通常受到個人經驗、知識以及社會環境等多重因素的共同影響。社交網絡加速了資訊的傳播速度, 擴大了其傳播範圍, 而資訊傳播的路徑和速度則受到網絡結構與節點影響力的影響(Ginossar et al., 2022)。在意見動態中, 個體往往會受到周圍人的影響, 尤其是那些被認為具有權威性或可信度的人(Cialdini & Goldstein, 2004; Katz & Lazarsfeld, 1955)。這種影響可以是正面的, 也可以是負面的, 具體取決於所傳播資訊的內容和傳播者的意圖(Festinger, 1954)。

從意見動態出發, 意見動力學(opinion dynamics)是傳播學和社會科學中研究個體或群體意見隨時間變化的理論體系, 重點關注資訊傳播過程中意見形成、擴散、演變及社會影響的機制(Castellano et al., 2009)。該領域探討不同個體間的互動如何導致群體內意見趨同、分化或極化等多樣化社會現象, 涵蓋社會影響力模型、認知衝突理論以及資訊傳播路徑等關鍵內容。意見動力學理論強調, 資訊在社交網絡中的傳播過程不僅是單向傳遞, 更是一個複雜的動態回饋系統, 受到認知偏誤、信念更新機制和社會結構等多重因素交織影響(Flache et al., 2017)。

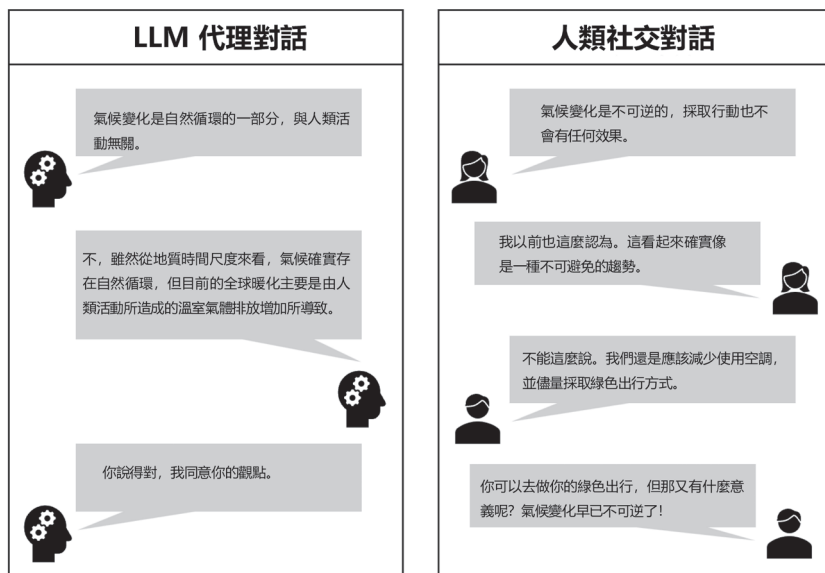
本研究的理論基礎融合了意見動力學中有關個體信念更新與群體互動機制的核心觀點, 同時借鑑資訊傳播理論對於資訊語境和媒介環境影響的系統性思考。透過引入LLMs作為虛擬代理, 本文將傳統基於規則的意見動力學模型擴展為具備自然語言語義理解與生成能力的認知互動模擬, 實現從結構性、靜態模型向動態、語義豐富模擬的轉變。此一方法不僅能夠更細緻地刻劃資訊傳播中的語言表達與語境變化, 也能捕捉資訊傳播過程中的認知衝突與信念調整機制, 從而為傳播學理論提供新的實證工具與理論視角。

理解意見動態在現代社會中具有重要意義, 尤其是在公共衛生宣傳、衝突調解和錯誤資訊遏制等領域。在公共衛生宣傳中, 及時且有

《傳播與社會學刊》，（總）第 78 期（2026）

效的意見傳播可以提高疫苗接種率，促進健康行為（郭娜等，2024）。在衝突調解方面，了解不同群體的意見動態有助於制定有效的溝通策略，減少誤解與衝突（Budak et al., 2016）。此外，錯誤資訊的傳播是當今社會面臨的重大挑戰，精準的意見動態模型可以幫助識別並遏制錯誤資訊的擴散（Loomba et al., 2021; Pennycook et al., 2020）。

圖一 LLM代理對話與人類社交對話的兩種模式圖



社交網絡的普及使得資訊的傳播速度和範圍顯著提高，社會影響力在意見形成和傳播中發揮了關鍵作用（賀琴等，2024）。在傳播學中，資訊傳播理論（communication theory）從更廣泛的視角研究資訊的生成、傳遞和接受過程，涵蓋經典的傳播模型（如拉斯韋爾模型〔Lasswell's Model of Communication〕、香農－韋弗模型〔Shannon-Weaver Model of Communication〕）、媒介理論以及社會建構主義等範式。現代資訊傳播理論強調資訊的語義內涵、媒介環境和受眾的能動性，特別關注傳播過程中資訊真實性、可信度以及受眾認知加工機制（McQuail, 2010）。在數位時代背景下，資訊傳播理論逐漸融合計算社會科學方法，拓展至大數據、社交媒體及人工智慧環境下的傳播機制研究（Tandoc et al., 2018）。

LLMs 模擬意見動態和偏見信息的影響

既有文獻回顧了意見動態和意見動力學理論，強調個體信念受到社會環境、權威影響及網絡結構的作用。傳統ABM模型透過數值或離散狀態模擬群體意見演化，但在捕捉自然語言互動、語境變化和多議題長期演化方面存在局限。尤其是在模擬連續互動、資訊語義與認知衝突的過程中，ABM難以準確呈現現實中個體意見的微妙變化。本研究的理論基礎融合了意見動力學中有關個體信念更新與群體互動機制的核心觀點，同時借鑑資訊傳播理論對於資訊語境和媒介環境影響的系統思考。透過引入LLMs作為虛擬代理，本文將傳統基於規則的意見動力學模型擴展為具備自然語言語義理解與生成能力的認知互動模擬，實現從結構性、靜態模型向動態、語義豐富模擬的轉變。這一方法不僅能夠更細緻地刻劃資訊傳播中的語言表達和語境變化，還能夠捕捉資訊傳播過程中的認知衝突和信念調整機制，從而為傳播學理論提供新的實證工具與理論視角。

因此，結合前文分析的理論缺口，產生了以下問題：LLM代理能否在多輪、多議題互動中準確模擬人類意見動態。提出這一問題，是為了檢驗LLM在補充傳統意見動力學模型局限性方面的可行性，同時為驗證理論機制提供新工具。雖然傳統意見動力學模型能夠描述群體內意見趨同與極化的演變，但這些模型主要依賴於固定規則或數值化信念，難以捕捉語言表達和語境變化對意見形成的影響。因此，本研究嘗試引入LLM代理，透過自然語言互動模擬多輪、多議題的意見動態，以驗證其是否能夠重現傳統理論所描述的信念演化模式。由此提出研究問題一(RQ1)：LLM代理是否能夠準確模擬人類在社交網絡中的意見動態，尤其在多輪對話和多議題互動中，是否呈現與經典意見動力學模型一致的信念演化規律？

LLMs、生成代理與社會模擬

LLMs是基於深度學習技術建構的自然語言處理(natural language processing, NLP)模型，透過訓練大量文本資料來生成和理解人類語言(郭佳霖等，2024)。LLMs能夠捕捉和生成語言的複雜模式，在文本生成、翻譯、問答、摘要、情感分析等多種NLP任務中表現出色。例

《傳播與社會學刊》，（總）第78期（2026）

如，GPT-4系列模型是OpenAI開發的最新一代LLMs，具有數百億到上千億級別的參數（Brown et al., 2020）。其技術特點包括預訓練與微調、多任務學習、生成品質和理解推理能力。例如，GPT-4在文本生成方面表現優異，能夠生成連貫、流暢且具有上下文關聯的文本，使其在對話系統、內容創作和自動化寫作等應用中具有高實用價值。

LLMs在模擬人類行為和意見動態中展現出巨大的潛力。這些模型可以處理複雜的自然語言任務，生成逼真的文本，從而模擬個體在各種社會情境中的互動和行為。透過訓練和微調，LLMs能夠捕捉到人類語言中的細微變化和語境依賴性，因此在社會科學和社會模擬中表現出色（謝天等，2023）。LLMs不僅能夠生成連貫的對話，還能夠在模擬中反映社會影響和認知偏誤，為研究社會動態提供了強大的工具。

值得注意的是，本文所採用的基於LLMs建構代理群體進行意見動態模擬的方法，在傳播學和人機互動領域已有一定的應用基礎。隨著人工智慧技術的發展，利用LLMs模擬個體或群體的認知與互動行為，已成為探索複雜資訊傳播機制的新途徑。該方法能夠彌補傳統基於固定規則的代理模型（如ABM）在處理語言理解與生成方面的局限，尤其在模擬多樣化意見形成和動態調整中表現出較強的靈活性和適應性。因此，從理論與技術角度來看，LLM驅動的模擬在研究複雜社會資訊傳播過程中具備合理性與創新性。

文獻回顧表明，LLMs具備生成自然語言和捕捉語義關係的能力，能夠模擬人類複雜社會互動。然而，現有研究多集中於單議題或短期對話實驗，缺少系統評估其在長週期、多議題互動中的表現，尤其是如何體現社會影響力和認知偏差的作用。結合前文對LLM能力的分析和ABM在語言與認知處理上的局限，研究者自然提出問題：在多輪、多議題互動中，LLM代理能否呈現社會影響力和確認偏誤對意見演化的作用。這個問題不僅源於理論缺口，也回應了智能媒介環境中意見動力學模型需要擴展和驗證的現實需求。

因此，本文第二個關注點是LLMs在模擬社會現象（如極化和錯誤資訊傳播）時的表現。目前已有研究嘗試使用LLMs生成虛擬代理，模擬多主題對話。例如，Park等人（2023）透過GPT-4建立了一組虛擬代理，使其圍繞不同話題進行互動和討論。研究發現，這些生成的代理能夠有效模擬人類在社交網絡中的意見傳播和變化過程。透過操弄代

LLMs 模擬意見動態和偏見信息的影響

理的初始信念和認知偏誤，研究人員能夠觀察到意見極化、資訊傳播路徑和社會影響力的作用。尤其是在應對多樣性強、動態性高的社會現象時，LLMs 展現出巨大的模擬潛力。LLM 代理能夠生成語義連貫、邏輯一致的文本，使其在模擬多輪互動和複雜社會現象時更接近真實社交網絡中的意見傳播過程。這為檢驗經典傳播理論在智能媒介環境下的適用性提供了新的工具。因此，本研究提出研究問題二 (RQ2)：LLM 代理在模擬社會現象 (如意見極化和偏誤資訊傳播) 時，其行為是否能夠反映社會影響和確認偏誤機制？

偏誤資訊與意見極化

偏誤資訊 (或虛假資訊) 是指那些不準確或具有誤導性的資訊，通常被有意或無意地傳播，旨在影響公眾的觀點或行為 (Chuang et al., 2024)。社交網絡的興起顯著加速了此類資訊的傳播速度和範圍，因為平台具備快速、高效的內容分享和傳播機制 (胡泳, 2024)。偏誤資訊對公眾意見和社會動態具有深遠影響。研究發現，COVID-19 疫苗相關的偏誤資訊顯著降低了公眾的接種意願，加劇公共健康危機 (陳瑛瑋等, 2021)。偏誤資訊不僅會誤導公眾，還會削弱科學共識，激化社會不信任和對立情緒。偏誤資訊與確認偏見交織，加劇意見極化，深化群體分歧 (夢非、朱慶華, 2021)。這種極化現象嚴重阻礙公共政策的有效制定。

在此基礎上，確認偏見是指個體傾向於接受和記憶支持其既有信念的資訊，同時忽視或否認與其信念相悖的資訊 (Nickerson, 1998)。此一認知偏誤會影響個體對新資訊的處理方式，在意見動態中發揮重要作用。研究指出，確認偏見不僅會鞏固個體既有觀點，還會導致不同意見之間的交流變得困難 (高嘉誠等, 2024)。人們因此更傾向於在社交網絡中尋找並分享支持其觀點的資訊，從而形成「資訊繭房」。從社會層面來看，確認偏見促使群體內部意見趨同，同時加劇群體間的分裂與極化。研究表明，確認偏見透過社交媒體的傳播機制，放大了群體內的意見一致性，同時加劇了群體間的意見對立 (何一清等, 2015)。偏誤資訊與確認偏見的疊加效應，進一步加劇了這一現象，使不同群體間的分歧越加顯著，進而影響公眾對科學與政策的信任。

《傳播與社會學刊》，（總）第78期（2026）

前文指出，偏誤資訊加劇意見極化，並與確認偏誤交織，使群體意見趨向分化。傳統ABM模型難以在多議題、長週期、自然語言語義層面上模擬這種複雜互動；而初步LLM代理研究也未充分探索偏誤資訊對意見動態的長期影響。基於這些理論和方法上的不足，研究自然產生問題：引入偏誤資訊後，LLM代理的意見動態是否發生顯著變化？科學共識是否仍能長期穩定形成？提出這個問題的邏輯是：透過LLM代理實驗，可以檢驗經典傳播理論（如社會影響、確認偏誤、意見極化）在智能媒介環境下是否仍然適用，以及科學共識形成機制在偏誤資訊干擾下的穩定性。因此，本文的第三個關注點，是透過人工引入偏誤資訊，分析LLMs生成的意見動態是否發生變化。由於偏誤資訊在現實社會中會擾亂公眾認知，並透過確認偏誤加劇意見極化，本研究透過在LLM代理系統中引入人為偏誤資訊，檢驗其對群體信念演化的影響。因此，提出研究問題三（RQ3）：在引入偏誤資訊後，LLM代理生成的意見動態是否會發生顯著變化？

現有研究的貢獻與不足

現有研究在模擬意見動態和社會現象方面作出了重要貢獻。Gilbert和Terna (2000) 透過基於ABM，展示了如何利用數學方程和計算方法模擬個體間的互動和意見傳播。Smaldino (2023) 在此基礎上進一步拓展，強調社會影響和網絡結構在意見動態中的作用。這些研究不僅為理解複雜社會現象提供了理論框架和模擬工具，也揭示了社會影響力及意見傳播路徑的關鍵因素。

基於上述文獻綜述，可以發現，現有意見動力學和資訊傳播研究對個體信念更新、社會影響和偏誤資訊傳播提供了堅實理論基礎，但在應對高度語言化、語境依賴的社交互動時仍存在局限。同時，LLM驅動的多代理模擬在語言理解與生成能力方面展現出潛力，卻缺乏與傳統理論系統對話的實證分析。同時，ABM和早期LLM代理研究在結構化模擬、社會網絡分析和語言互動上各有優勢，但仍無法同時解決多議題、長週期、語義豐富、偏誤資訊干擾的複雜模擬問題。這一缺口提示研究需要一個新的研究框架，將LLM代理、多議題互動、偏誤

LLMs 模擬意見動態和偏見信息的影響

資訊干預整合進實驗設計中，以驗證科學共識形成機制並擴展經典理論。因此，本研究目標是探索LLM代理在複雜社會現象模擬中的潛力，檢驗理論機制在多議題、長週期互動下的適用性，為優化科學傳播策略提供理論與方法支持。本文旨在填補兩者之間的空白，將LLM代理模擬與經典傳播理論（如意見動力學、社會影響理論、確認偏誤和資訊傳播路徑理論）結合，提出明確的研究問題與假設，以指導實驗設計與資料分析。

無論是基於現實的小樣本仿真模擬，還是傳統的基於ABM，在捕捉人類語言細微差別和上下文關聯方面均表現不足。傳統基於規則的代理基礎模型（ABM）在模擬社會互動與宏觀行為演化方面已有廣泛應用，其優勢在於能夠透過明確規則控制代理行為，並觀察系統層面的湧現模式（emergent pattern）。然而，ABM在處理語義豐富、上下文依賴的意見傳播過程時存在一定局限。例如，傳統ABM代理的行為決策往往依賴於事先設定的固定規則，難以模擬現實社交行為中語言表徵與觀點建構的複雜動態。因此，當研究對象轉向「話語生成」、「觀點變遷」與「語義傳播」這類高度語言化的問題時，基於規則的ABM在表達層面顯得力有未逮。

相較之下，基於LLM的代理系統具備強大的語言理解與生成能力，能夠動態地根據語境生成語義連貫、觀點邏輯一致的表達。這使得LLM代理不僅能模仿人類在網絡空間中的語言互動行為，更能模擬多輪對話中意見的微妙變化與傳播路徑。因此，採用LLM驅動的模擬方法，有望突破ABM在社會語言行為模擬中的結構性瓶頸，更貼近真實社交平台中意見生成與演化的複雜機制。當下的LLMs能夠生成高品質的自然語言文本，透過生成虛擬代理模擬意見動態，為解決這一難題提供了新的路徑。LLMs能夠捕捉和再現複雜的語言互動，模擬個體在社交網絡中的真實對話，進而更準確地反映意見形成和傳播的過程（許志偉等，2024）。

隨著GPT等語言模型的興起，研究者開始嘗試將LLMs嵌入社會模擬框架中，用以生成虛擬代理的語言行為（Park et al., 2023）。相較於傳統基於規則或淺層機率機制的模擬方法，LLM驅動代理系統能夠藉助其預訓練語言知識與對話上下文處理能力，更真實地重現人類在面

《傳播與社會學刊》，（總）第78期（2026）

對社會議題時的語言互動過程。這種模擬方式不僅保留了多代理系統在結構層面的擴展性，更大幅提升了語義表達的真實性與社會行為的可解釋性，為意見演化建模提供了新路徑。

本研究在吸收上述LLM代理模擬範式的基礎上，進一步引入多議題設定與長期互動機制，嘗試探索LLM代理在多輪互動中逐步形成集體信念結構的過程，旨在從語言演化的視角審視科學共識在社交網絡中傳播與聚合的可能路徑。具體研究思路來源於上述理論與方法的發展，旨在探討LLMs在模擬意見動態中的應用潛力。具體而言，研究將使用GPT-4o生成虛擬代理，透過模擬多話題對話，研究社會影響、確認偏見和偏誤資訊傳播對意見動態的影響。透過設計系列實驗，測試代理在不同初始條件和干預措施下的表現。本研究不僅為理解和預測如政治極化、公眾輿論變化等社會現象提供了新方法和工具，也有助於揭示資訊在社交網絡中的傳播機制，進一步識別科學資訊傳播的障礙並優化傳播策略。

本文綜合歸納的研究問題為：

RQ1：LLM代理是否能夠準確模擬人類在社交網絡中的意見動態，尤其在多輪對話和多議題互動中，是否呈現與經典意見動力學模型一致的信念演化規律？

RQ2：LLM代理在模擬社會現象（如意見極化和偏誤資訊傳播）時，其行為是否能夠反映社會影響和確認偏誤機制？

RQ3：透過操控LLM代理的初始信念和偏誤資訊比例，能否揭示科學共識在智能媒介環境中的自我糾正機制及其穩定性？

實驗設計及步驟

代理設置、初始信念分佈、話題選擇

本研究使用GPT-4o模擬多個智能體（虛擬代理），以探索LLMs在模擬意見動態和社會現象中的應用潛力。每個智能體代表一個虛擬代理，實驗共設置50個代理，以確保樣本的多樣性和代表性。每個代理被賦予

LLMs 模擬意見動態和偏見信息的影響

特定特徵，包括初始信念和社交活躍度。初始信念分為三種狀態：接受、拒絕和中立。此種設定確保了觀點的多樣分佈，便於觀察不同初始信念在社交網絡中的互動和演變過程。社交活躍度分為：非常活躍、相對活躍和不活躍，設定非常活躍代理每天發佈 10 至 20 條訊息，相對活躍代理每天發佈 5 至 10 條訊息，不活躍代理每天發佈 1 至 5 條訊息。

本研究的核心變項聚焦於「代理信念」(agent belief)，即透過 LLMs 模擬的個體代理在特定科學議題上的觀點和態度。這一變項不僅是本研究考察的因變項，也是理解資訊傳播與意見動態的關鍵切入點。代理信念的概念化基於傳播學中意見形成與傳播理論，具體體現為代理在多輪互動與資訊影響下，信念的演變過程。其操作化透過對代理輸出內容的語義分析與主題聚類進行量化，反映信念的方向性和強度變化。聚焦代理信念，有助於揭示資訊環境中知識傳播與認知調整的機制，從而深入探討科學傳播的效果與路徑。研究試圖透過 LLMs 這一新興工具，拓展傳統傳播理論的應用邊界，尤其是在數位化、智能化資訊環境下，模擬和理解公眾意見形成的複雜性。透過對核心變項的理論釐清與實證考察，研究希望為傳播學中的意見動力學提供新的視角和方法論補充，具備理論推進與實踐啟示的雙重價值。以下是代理三種活躍度區分邏輯的代碼：

```
for person in range(num_people):
    sender = f'代理人_{person + 1}'
    activity_level = 0
    if person + 1 in very_active_agents:
        activity_level = random.randint(10, 20) # 非常活躍的代理人每天發佈
10-20 條訊息
    elif person + 1 in somewhat_active_agents:
        activity_level = random.randint(5, 10) # 相對活躍的代理人每天發佈
5-10 條訊息
    else:
        activity_level = random.randint(1, 5) # 不活躍的代理人每天發佈 1-5 條
訊息
```

《傳播與社會學刊》，（總）第 78 期（2026）

為確保討論具有科學依據與現實意義，本研究選取三個具有廣泛科學共識的話題作為討論對象。如表一所示，這些話題包括「氣候變化」、「轉基因食品」以及「疫苗安全」，當中不包括 COVID-19 疫苗。其中，「氣候變化」話題主要關注全球暖化的事實及其影響；「轉基因食品的安全性」聚焦其對健康與環境的影響；而「疫苗的有效性和安全性」則著重於疫苗效用等相關議題。為提升模擬語境的複雜性與表達多樣性，研究進一步將每個一級議題劃分為多個子議題。此一劃分主要基於兩個維度：其一，現實社交平台中使用者討論的高頻焦點與爭議點，例如「氣候變化是否由人為因素造成」以及「可再生能源是否具備替代可能」；其二，政策文件與學術文獻中所揭示的意見分歧，例如疫苗議題下的「副作用爭議」或「政府強制接種措施」。透過此種維度劃分，研究力求在維持議題代表性的同時，引導代理生成內容於不同子主題之間展開，從而模擬現實中更具多維性的意見表達環境。

表一 話題類型示例

一級議題	分議題
氣候變化	溫室氣體排放
	氣候變化的證據
	氣候變化的影響
	應對氣候變化的措施
	國際合作
轉基因食品	健康安全性
	環境影響
	農業生產力
	經濟與市場影響
	倫理與社會問題
疫苗安全	疫苗的科學原理
	臨床試驗與審批
	疫苗的副作用與風險
	群體免疫
	疫苗的公共衛生影響

實驗設計和步驟

為模擬信念的更新過程，實驗透過多輪對話與資訊交換進行迭代，整體持續30天。主要測量變項包括：信念狀態（每個代理每日的信念狀態，如接受、拒絕、中立）、互動鏈路、非偏誤資訊傳播路徑（記錄資訊在社交網絡中的傳播路徑和範圍）、引入偏誤資訊後每個代理在不同時間步的信念狀態，以及偏誤資訊的傳播範圍等。控制變項包括：初始信念狀態比例、科學共識議題、是否加入偏誤資訊，以及實驗持續天數。三個階段的變項配置如表二所示。

表二 三階段實驗的核心變項控制

實驗階段	信念狀態比例	科學共識議題	偏誤資訊	天數
第一階段	設置一組	3個	無	30
第二階段	設置三組	3個	無	30
第三階段	設置一組	3個	有	30

在實驗初始化階段，研究首先為每位虛擬代理分配初始立場（支持、反對或中立），並以此為依據構造其首輪資訊發佈內容。具體而言，透過設定具備明確立場傾向的提示詞，引導大語言模型生成具有代表性的言論。每位代理的初始發言數量控制在五條以內，以確保模擬在內容豐富度與模型負載之間取得平衡。此外，代理間的社交網絡結構在模擬啟動前即完成建構，採用類似弱連結小世界的結構，以模擬現實社交平台上的資訊擴散路徑。代理的信念狀態由其在每一輪生成內容中的立場傾向所決定。具體操作上，使用GPT-4o對代理發言進行「零樣本學習」（zero-shot learning）語義編碼，判定其對核心議題的態度為「支持」、「反對」或「中立」，並為每個判定結果標註置信度（0-1），用以刻劃其信念強度變化。所有代理生成文本均映射到核心變項「代理信念」上，透過語義量化與聚類分析，形成可操作化的信念指標，以保證實驗設計中「初始信念」與「偏誤資訊」操弄的可驗證性。同時，為確保「支持」、「反對」或「中立」的初始信念在代理生成文本中獲得忠實反映，研究對實驗一與實驗二的初始信念分佈設定進行事後驗證，確認生成文本與預設立場高度一致（文本中立/支持/反對比例與初始設定吻合）。

《傳播與社會學刊》，（總）第78期（2026）

該信念變項將在時序維度上進行追蹤，從而辨識代理的立場是否發生變遷。同時，研究亦使用嵌入表示計算不同代理之間語言表達的聚合程度，以衡量多輪互動中是否出現群體意見極化現象。所有實驗均在相同網絡結構和代理配置下重複運行，以檢驗模型輸出的穩定性。此外，所有模擬均在無額外輸入干預的封閉資訊系統中運行，以確保模型輸出完全由代理間的互動行為所推動，從而建構出一個自組織的動態傳播環境。

I. 實驗一：無偏誤資訊下的意見動態模擬

第一階段，實驗在無偏誤資訊條件下模擬意見動態。初始條件基於預設分佈設定各代理的信念狀態：30%的代理接受科學共識，30%的代理拒絕科學共識，40%的代理持中立態度。代理透過模擬社交網絡進行對話和資訊交換，以模擬現實中人們透過社交媒體互動的情境。高頻次的互動設計旨在捕捉資訊傳播與意見動態的細微變化。

II. 實驗二：初始意見分佈對結果的影響

第二階段，實驗重點探討不同初始意見分佈對結果的影響。研究設置三種初始信念分佈方案。方案一為均衡分佈（30%接受、30%拒絕、40%中立），方案二為多數拒絕科學共識（20%接受、60%拒絕、20%中立），方案三為多數接受科學共識（60%接受、20%拒絕、20%中立），而對話持續天數、代理人數量與科學共識議題均保持不變。此一設計旨在模擬不同社會環境下的意見動態，並觀察初始信念分佈如何影響最終的意見演變。

III. 實驗三：引入偏誤資訊後的意見動態模擬

第三階段，實驗在無偏誤資訊的基礎上引入偏誤資訊，以觀察其對意見動態的影響。初始條件與第一階段相同，但增加了偏誤資訊干預。具體而言，研究設定在對話過程中投放佔總資訊量10%的偏誤資訊，用以模擬與科學共識相悖的虛假內容，刻意誤導代理進行討論，並評估其對社交網絡中意見動態的影響。表三為無偏誤與有偏誤資訊狀態下，50個代理在30天內就三個科學共識議題展開的對話總量，共

LLMs 模擬意見動態和偏見信息的影響

計 194,699 條。對於被設定為「偏誤資訊傳播者」的代理，研究對其生成內容進行了語義與事實性分析。結果顯示，這些代理生成文本中包含與科學共識相悖陳述的比例，顯著高於無偏誤條件，表明偏誤資訊操弄具備有效性。

表三 無偏誤與有偏誤資訊狀態下三個議題的代理對話總量

	氣候變化	轉基因食品	疫苗安全
無偏誤	26,007	25,995	25,943
有偏誤	38,941	38,758	39,055

在實驗設計中，為避免潛在混淆變項對模擬結果產生干擾，所有代理生成發言時所接收的議題提示文本均保持內容一致，僅在呈現順序上進行隨機化處理，以避免因提示語差異影響生成傾向。同時，代理之間的社交網絡結構在實驗初始化時即已設定為固定的小世界結構，以確保不同實驗之間的資訊傳播路徑一致。最後，所有模型生成均採用一致的溫度 (temperature) 參數 (設為 .7)，在保持語言多樣性的同時抑制過度隨機性，以增強實驗穩定性與結果的可解釋性。雖然 LLM 模擬中仍不可避免地存在部分語義噪音，但透過上述設計，研究已盡可能減少非研究變項對實驗系統性輸出的干擾。

綜上所述，透過三個階段的實驗設計，本研究系統考察了 LLMs 在模擬意見動態中的應用潛力。在無偏誤資訊條件下，研究觀察意見的自然演變過程，分析社交網絡中的互動模式與資訊傳播路徑；在引入偏誤資訊後，評估其對意見動態的干擾效應，理解偏誤資訊的傳播機制及其影響；在不同初始信念分佈條件下，則探討初始意見環境對最終意見分佈的影響。

數據分析

固定初始信念狀態下的意見動態與信念演變

實驗一模擬了無偏誤資訊條件下的意見動態。在三個議題上，初始信念分佈均為 15 個支持、15 個反對、20 個中立。經過 30 天的對話

《傳播與社會學刊》，（總）第78期（2026）

後，各議題的信念狀態分佈：(1) 氣候變化：3個反對、4個中立、43個支持；(2) 轉基因食品的安全性：1個反對、1個中立、48個支持；(3) 疫苗的有效性和安全性：0個反對、1個中立、49個支持。

圖二、圖三與圖四的橫座標為時間，縱座標為代理編號(1-50)。圖示上半部分中，每個圓點的顏色代表各代理每日的信念狀態，其中深黑色是支持，灰色是中立，淺灰色是反對；圓點顏色的變化對應信念的變化。圖示下半部分的橫座標為時間，縱座標為信念得分，箱型圖的粗細代表代理信念的聚集程度。隨著時間推移，絕大部分代理的信念逐漸聚集於1.0(支持科學共識)附近。因此，數據表明，在無偏誤資訊條件下，三個議題中的代理信念均朝向科學共識集中。尤其是在「疫苗的有效性和安全性」議題上，幾乎所有代理最終均轉向支持科學共識。大部分信念態度的轉換發生在第五天前後，表明資訊傳播的早期階段對信念更新具有關鍵作用。

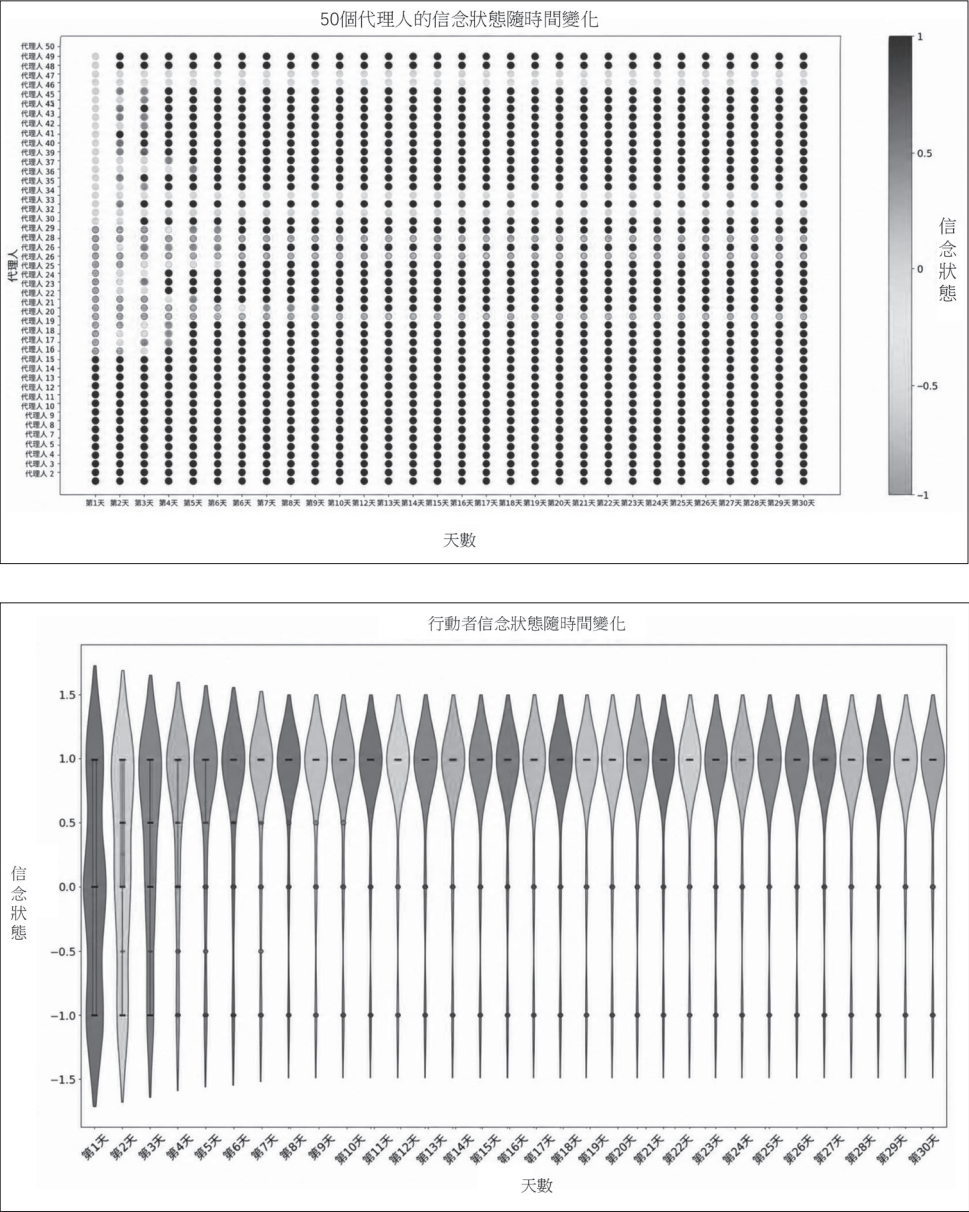
以上結果表明，在無偏誤資訊條件下，社交網絡中的高頻互動與資訊交換能夠有效促進科學共識的傳播與接受。本實驗透過模擬無偏誤資訊條件下的社交網絡互動，初步驗證了高頻互動對意見動態與科學共識傳播的積極作用，尤其突顯出資訊傳播早期階段的重要影響。

透過潛在狄利克雷分配(latent dirichlet allocation, LDA)主題建模，可以分析50個代理圍繞三個科學共識議題所展開的具體討論內容。LDA旨在從大量文本中挖掘隱藏的主題結構。圍繞三個典型科學共識議題所進行的LDA分析，有助於增強模擬內容的真實語境。例如，在疫苗議題中設置了「疫苗副作用」、「疫苗有效性」、「政府疫苗政策」三個子議題，作為LLMs代理在生成發言時的上下文提示輸入。這些子議題在實驗階段具有雙重作用：一方面，它們限定了代理在不同回合中討論的具體話題範圍，增強了模擬的語義多樣性；另一方面，子議題的語境提示亦作為生成控制變項，使得LDA分析時能在一定程度上追溯話題生成來源。

如圖五所示，由上至下依次為「氣候變化」、「轉基因食品」、「疫苗安全」三個科學議題的主題趨勢演變及主題佔比圖，橫座標為日期，縱座標為對話數量，對應不同主題的變化趨勢。

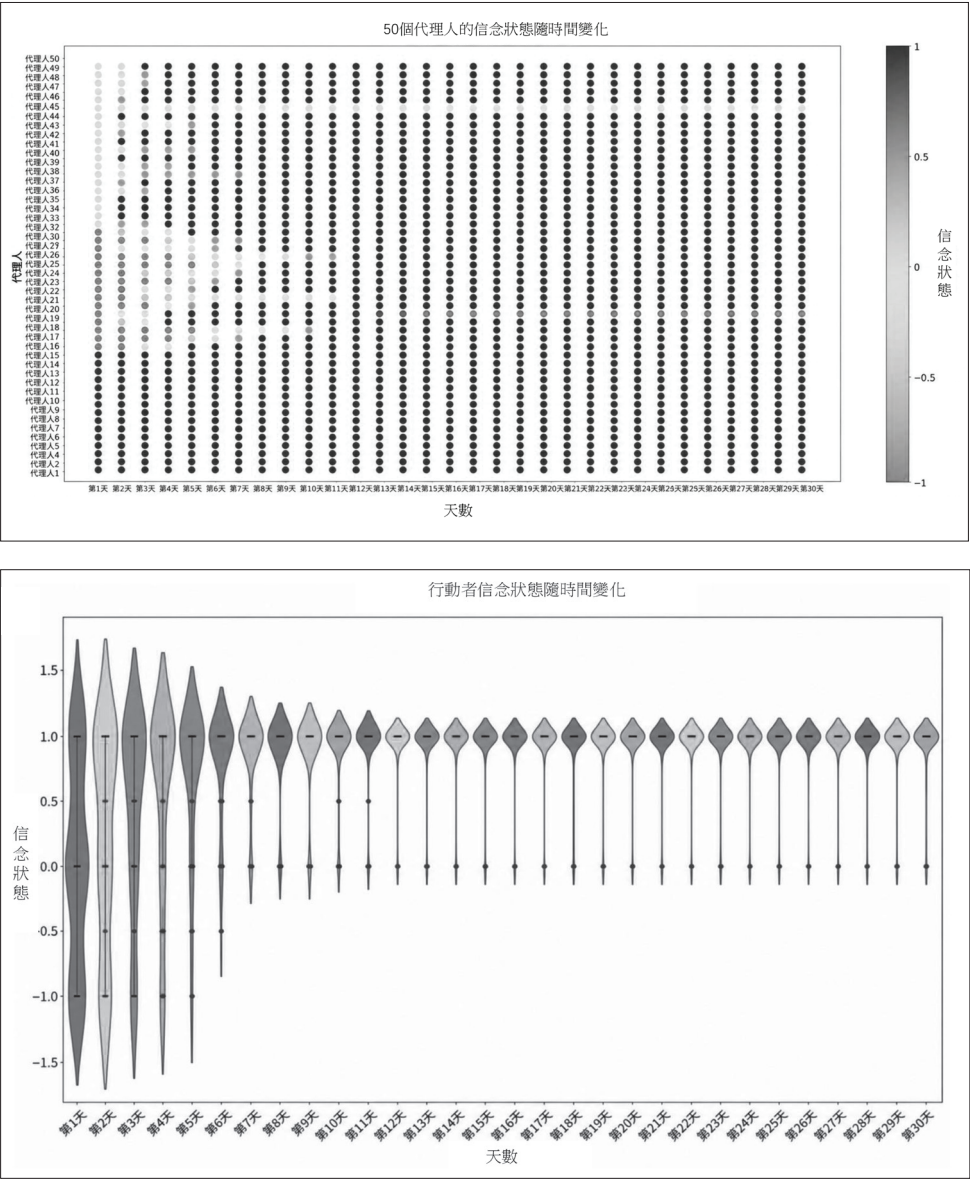
LLMs 模擬意見動態和偏見信息的影響

圖二 50個代理就「氣候變化」的信念狀態演變



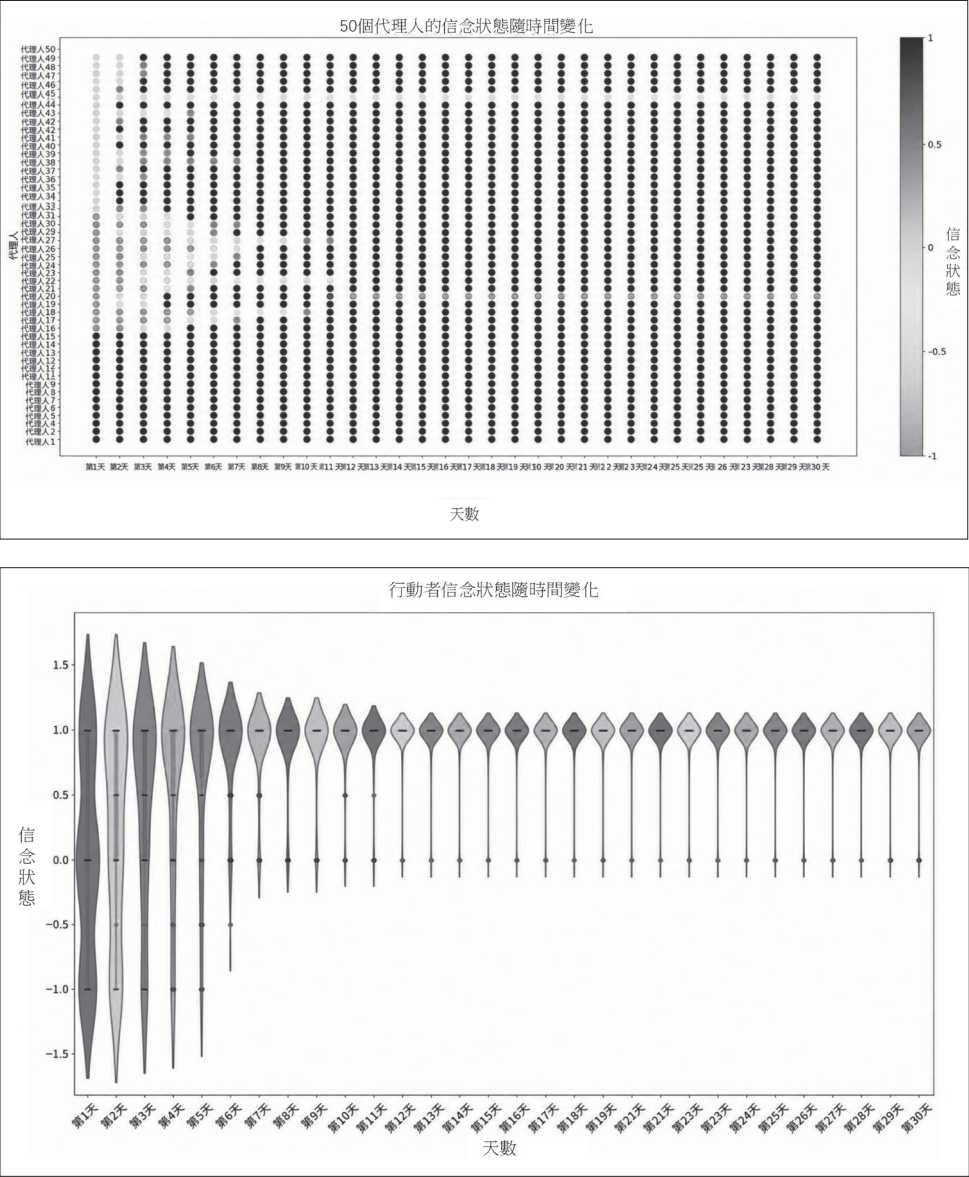
《傳播與社會學刊》，（總）第78期（2026）

圖三 50個代理就「轉基因食品」之信念狀態演變



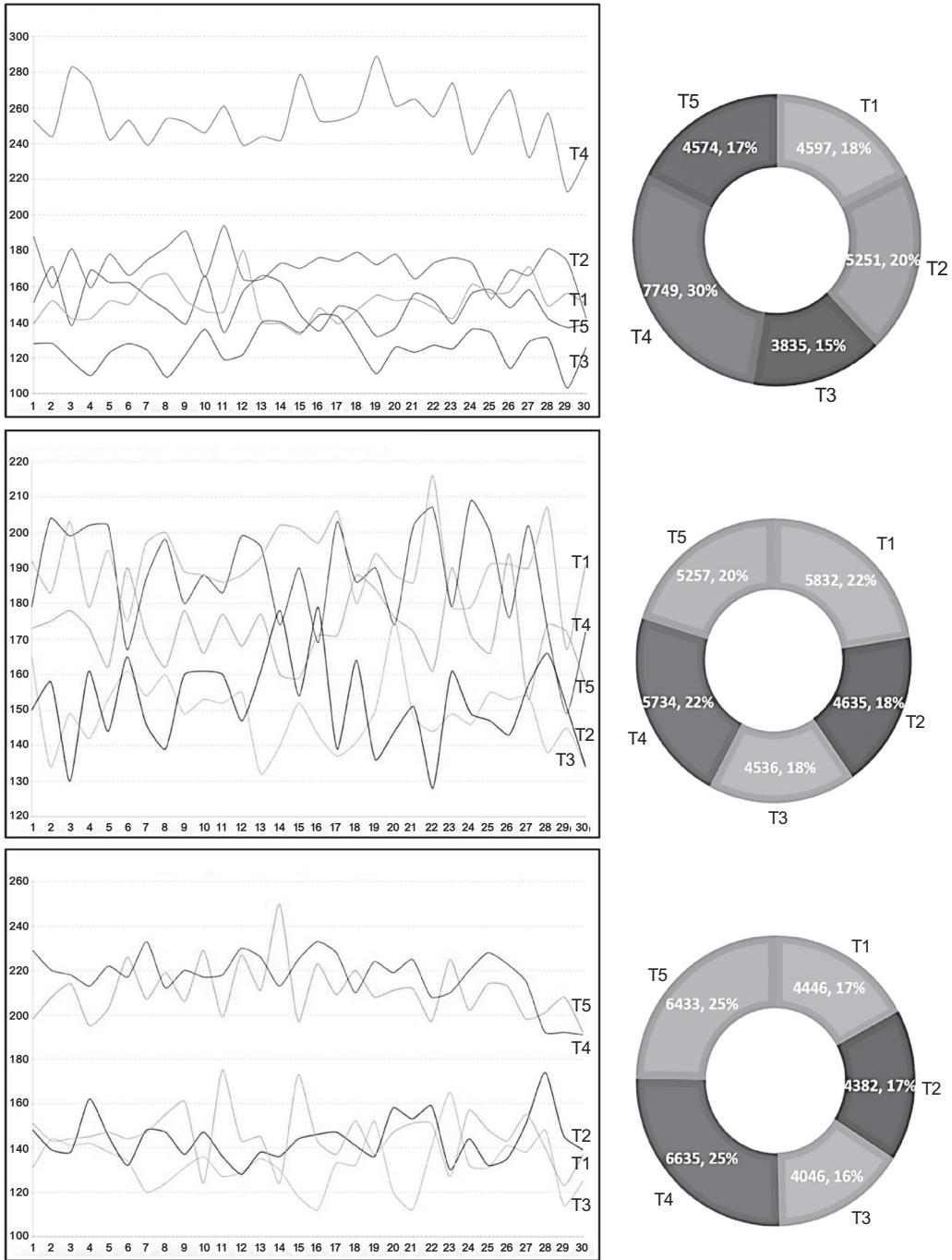
LLMs 模擬意見動態和偏見信息的影響

圖四 50個代理就「疫苗安全」之信念狀態演變



《傳播與社會學刊》，（總）第 78 期（2026）

圖五 三個話題的 LDA 主題建模趨勢圖與分話題佔比圖



LLMs 模擬意見動態和偏見信息的影響

具體而言，「氣候變化」議題被識別出五個主要主題：氣候變化對健康和環境的影響(4,597條)、氣候變化的原因及科學證據(5,251條)、應對氣候變化的措施(3,835條)、對氣候變化的不同看法(7,749條)，以及個人在應對氣候變化中的角色(4,574條)。這些主題涵蓋了氣候變化問題的各個層面，從其影響、成因到應對措施及不同觀點，體現了社交網絡對氣候議題的多維關注。其中，「對氣候變化的不同看法」為討論量最大的主題，共計7,749條訊息。特別是在第19日當天，討論量達到峰值289條，主要包括氣候變化對農業的影響、企業減排責任及環保生產工藝等。

「轉基因食品」議題下包含五個主要主題：政府和企業的角色(5,832條)、安全性和科學證據(4,635條)、公眾參與和科學研究(4,536條)、監管和環境影響(5,734條)，以及安全性爭議和觀點(5,257條)。這些主題涵蓋了轉基因食品在安全性、監管、科學研究和公眾參與等方面的主要討論點。在第22日的討論中，「政府和企業的角色」主題的討論數量最多，共計216條訊息，主要包括政府在轉基因食品監管中的措施、轉基因食品在提高農業產量和抗病性方面的潛力，以及可能帶來的實際影響。這些討論反映了代理對政府和企業在轉基因食品推廣與監管中所起作用的集中關注。

「疫苗安全」議題下包含五個主要主題：疫苗的安全性和有效性(4,446條)、疫苗的科學進步和監管措施(4,382條)、公眾態度和科學證據(4,046條)、疫苗的社會影響和長期影響(6,635條)，以及疫苗的爭議和不同觀點(6,433條)。這些主題涵蓋了疫苗在安全性、科學進步、公眾態度、社會影響和爭議觀點等方面的主要討論點。在第14天的討論中，「疫苗的爭議和不同觀點」主題最為熱烈，共計250條訊息。內容包括涉及接種疫苗的觀點、對疫苗安全性的質疑，以及疫苗技術對醫學發展的影響等。這些討論反映了代理對疫苗安全性和科學研究方法的高度關注，而多元化與爭議性的觀點則成為該議題的核心焦點。需要指出的是，由於LDA模型是基於輸出語料的無監督主題提取方法，其所識別的主題與實驗預設的子議題之間並非一一對應。研究並不試圖直接映射兩者，而是希望透過LDA揭示在自由演化語境下，代

《傳播與社會學刊》，（總）第78期（2026）

理實際討論內容的自然聚類結構，並藉此檢驗LLMs代理是否會圍繞預設語義核心產生內容聚合效應。因此，子議題主要在發言生成階段提供結構引導，而LDA主題分析則用於檢驗語義聚合趨勢，兩者在方法上互為補充。

實驗一證明，在無偏誤資訊環境中，LLM代理能夠更穩定地接受和傳播科學共識，降低資訊失真和極化的可能性。這一發現顯著有別於社會自然狀態下的人類意見動態。在現實社交網絡中，人類輿論往往受到意見領袖、謠言和流量等多種因素的影響，資訊失真和意見極化現象頻繁發生。人類認知中的確認偏誤，則進一步強化了人們對既有信念一致資訊的接受傾向，加劇資訊極化。相比之下，在無偏誤資訊條件下，LLM代理能夠透過穩定的資訊傳播機制，有效促進科學共識的傳播，顯示出其在控制資訊失真和減少意見極化方面的潛力。

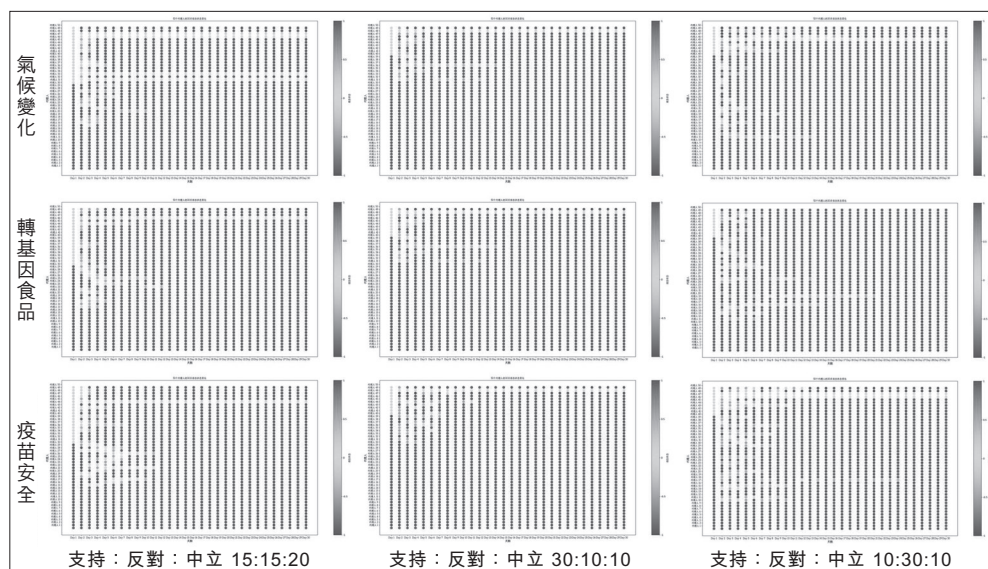
不同初始信念分佈下的信念基數效應和真實趨向性

實驗二探討了不同初始意見分佈對結果的影響，透過設置三種不同的初始信念分佈方案進行具體評估。圖六為「3 × 3」分組的實驗結果。數據表明，無論初始信念如何分佈，代理的最終信念均趨向科學事實。也就是說，不論初始信念分佈如何，代理最終都傾向於趨同於科學共識，在一定程度上呈現出科學資訊在網絡傳播中的「自我糾正」趨勢。不過需注意，結果主要反映的是大語言模型在封閉系統中生成的代理行為所呈現的內在一致性，而非對現實社會意見演化過程的直接映射。因此，上述信念演化趨勢更適合被理解為在模型結構假設下的「機制性發現」。相關結果為後續從「人工智慧生成行為」出發，探討複雜意見形成機制提供了理論啟發與技術路線，具有重要的方法論意義。Berger和Milkman (2012)指出，科學資訊因其高可信度和公共利益屬性，往往更容易被廣泛傳播和接受。Friedkin和Johnsen (2011)的研究也表明，社交網絡中的意見動態往往趨向於理性共識，為本研究發現提供了進一步支持。

LLMs 模擬意見動態和偏見信息的影響

儘管大多數代理最終趨向科學事實，實驗仍發現最終信念分佈與初始信念的基數相關；當初始反對者比例較高時，實驗最終反對者數量也相對較多。這表明初始信念分佈對意見動態存在基數效應。Lewandowsky 等人 (2013) 的研究指出，初始信念和資訊處理偏誤對最終信念形成具有顯著影響。Allcott 和 Gentzkow (2017) 的研究表明，虛假資訊和錯誤認知在初始階段的廣泛存在，會對資訊傳播和最終信念形成產生深遠影響。

圖六 三種不同初始信念狀態分佈下的代理信念變化圖



不同議題的對比結果顯示，初始條件對意見演變速度和最終分佈具有顯著影響。例如，在疫苗議題中，即便初始條件為多數反對，最終反對者比例仍相對較低，這可能與該議題的科學共識更強，以及公眾對健康安全問題的關注度更高有關。而在轉基因食品議題中，最終反對者比例相對較高，反映出該議題的爭議性和資訊不對稱現象更為明顯。因此，當初始信念分佈為「多數接受科學共識」時，意見動態趨向科學共識的速度最快，說明當更多代理一開始就接受科學共識時，有助於科學共識更迅速且穩定地擴散與被採納。而在「多數拒絕科學共

《傳播與社會學刊》，（總）第 78 期（2026）

識」的初始條件下，儘管總體趨勢仍趨向於科學共識，但最終仍有部分代理保持反對態度，驗證了基數效應的存在。此外，中立者在意見演變過程中表現出較高的可轉化性，更容易受到其他群體的影響而改變立場，凸顯了中立者在意見動態中的關鍵作用。

為進一步分析不同初始信念方案對群體信念演化的影響，本研究對三種議題（氣候變化、轉基因食品、疫苗安全）在三種初始意見分佈方案下的代理人信念數據進行了單因子變異數分析 (analysis of variance, ANOVA)，並在 ANOVA 顯著的情況下，進一步使用 Tukey HSD 事後檢驗，以比較各方案之間的具體差異。ANOVA 能夠檢驗整體是否存在顯著差異，而 Tukey HSD 透過多重比較校正，能夠識別哪些特定方案組合存在顯著差異，從而提供更精細的條件比較。

表四 事後分析 ANOVA 值

Topic	ANOVA_p
氣候變化	.63
轉基因食品	.04
疫苗的安全	.67

如表四所示，不同初始信念方案對「轉基因食品」議題的最終信念分佈具有顯著影響 (ANOVA $p = .04$)；而對於「氣候變化」和「疫苗安全」議題，不同方案之間的整體差異則未達顯著水準 ($p > .60$)。具體而言，「氣候變化」與「疫苗安全性」在三種初始方案下的平均數差異較小，說明在這兩類議題上，初始信念分佈對最終意見演化的影響相對有限，可能受到既有社會共識或代理互動機制穩定作用的約束。相較之下，「轉基因食品」議題呈現出更高的敏感性，不同初始方案下的最終信念出現可檢測的差異，表明在某些科學議題中，初始意見分佈可能會顯著塑造群體信念動態。

LLMs 模擬意見動態和偏見信息的影響

表五 不同信念分佈的事後檢驗結果

Topic	Group1	Group2	MeanDiff	p_adj	Lower	Upper	Reject
氣候變化	均衡分佈	多數拒絕科學共識	-.10	.68	-.38	.18	FALSE
氣候變化	均衡分佈	多數接受科學共識	-.10	.68	-.38	.18	FALSE
氣候變化	多數拒絕科學共識	多數接受科學共識	.23	1.61	-.28	.28	FALSE
轉基因食品	均衡分佈	多數拒絕科學共識	.14	.14	-.03	.31	FALSE
轉基因食品	均衡分佈	多數接受科學共識	.18	.04	.01	.35	TRUE
轉基因食品	多數拒絕科學共識	多數接受科學共識	.04	.85	-.13	.21	FALSE
疫苗安全	均衡分佈	多數拒絕科學共識	.02	.95	-.14	.18	FALSE
疫苗安全	均衡分佈	多數接受科學共識	-.04	.83	-.20	.12	FALSE
疫苗安全	多數拒絕科學共識	多數接受科學共識	-.06	.66	-.22	.10	FALSE

如表五所示，進一步的 Tukey HSD 事後分析揭示，在「轉基因食品」議題中，均衡分佈方案與多數接受科學共識方案之間存在顯著差異 ($p_{adj} = .04$)，而其他方案組合之間的差異則未達顯著水準。這表明，即便整體差異有限，事後分析仍能識別出特定條件下的顯著變化，強調了初始信念分佈在意見形成過程中的細微作用。此外，這些結果也支持在後續研究中繼續使用事後檢驗的方法，以便更細緻地比較不同條件對意見演化的影響。

偏誤資訊引入下的信念演變、語義圖譜和活躍代理

實驗三透過引入偏誤資訊來觀察其對意見動態的影響。初始條件與第一階段相同，但增加了 10% 的代理作為偏誤資訊傳播者。根據第一階段各議題的資訊總量，在這一階段每個主題加入了 2,600 條偏誤資訊。例如，議題一包括「氣候變化是個騙局」、「氣候變化是自然現象，不受人類影響」等；議題二包括「轉基因食品會導致癌症」、「轉基因食

《傳播與社會學刊》，（總）第78期（2026）

品含有有毒成分」等；議題三包括「疫苗會導致自閉症」、「疫苗中含有有毒成分」、「疫苗會導致不孕不育」等。需要說明的是，為保證實驗結果的內在一致性與可控性，所有代理的言論生成完全依賴其所接收到的歷史對話上下文和既定信念狀態，未再額外引入外部擾動變項（如突發新聞事件、平台干預政策或意見領袖發佈），以維持初步模型在封閉語境下的可控性與邏輯演化的清晰性。大模型引入偏誤資訊的邏輯：

```
for _ in range(activity_level):
    receiver = f'代理人_{random.randint(1, num_people)}'
    belief = belief_states[person, day - 1]

    if random.random() < 0.1: # 10% 的機率引入干擾資訊
        message = random.choice(error_messages)
    else:
        if belief == 1:
            message = random.choice(support_statements)
        elif belief == -1:
            message = random.choice(oppose_statements)
        else:
            message = random.choice(neutral_statements)
```

引入偏誤資訊後，對話數量顯著增加；相較於無偏誤狀態下，如表三所示，每個科學共識議題的對話數量均增長約40%，顯示出偏誤資訊在社交網絡中的強傳播力與高影響力。少部分代理傾向於接受符合其既有信念的資訊，忽視或拒絕與其信念相悖的資訊，從而導致意見進一步分化。這與Anagnostopoulos等人（2015）的研究一致，該研究發現，社交媒體上的虛假資訊傳播會引發大規模的討論與分化。Allcott和Gentzkow（2017）也指出，社交媒體上的虛假資訊能夠顯著影響公眾輿論，增加意見極化與分裂。虛假資訊的傳播速度與範圍往往超過真實資訊，進一步加劇社會認知裂隙。LLM引入圖譜依存關係的邏輯：

LLMs 模擬意見動態和偏見信息的影響

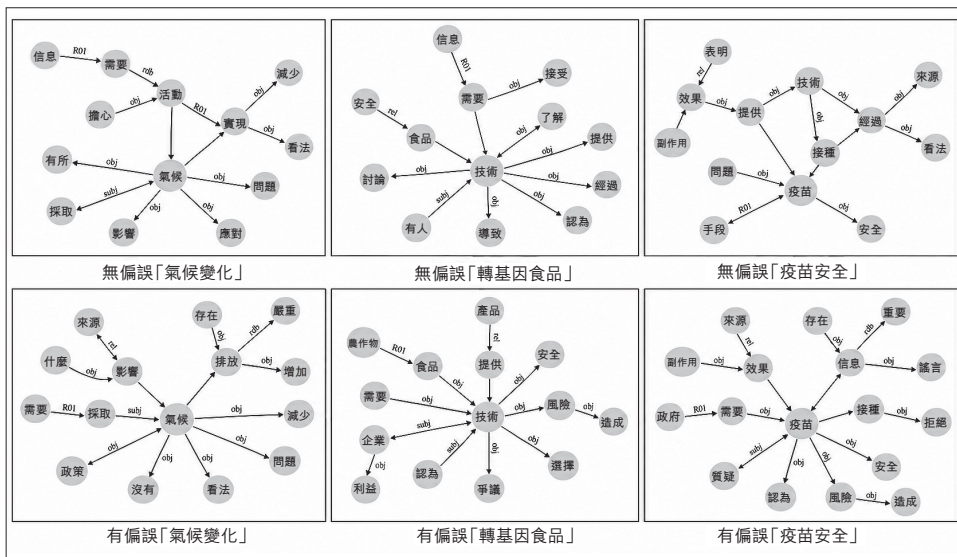
```
# 添加依存關係到圖中
for token in doc:
    if token.dep_ in ('nsubj', 'dobj', 'pobj', 'attr', 'ROOT'):
        for child in token.children:
            if child.dep_ in ('nsubj', 'dobj', 'pobj', 'attr', 'prep') and token.head
            != child:
                G.add_edge(token.head.text, child.text, label=token.dep_)
```

選擇圖中最有代表性的40個實體（基於節點的度）

```
top_nodes = sorted(G.nodes(data=True), key=lambda x: G.degree(x[0]),
reverse=True)[:40]
```

為觀測偏誤資訊加入後50個代理意見動態的具體呈現，研究基於依存關係構造三個科學共識議題的語義圖譜，並與無偏誤狀態進行對比。對於每個科學共識議題，選取最具代表性的40個實體（基於節點度數）。在NLP中，依存關係（dependency relations）用於描述句子中詞與詞之間的關係，幫助揭示句子的句法結構，包括名詞性主語（nsubj）、直接賓語（dobj）、介詞賓語（pobj）、屬性（attr）與根依存關係（ROOT）等。「3 × 2」結構的語義圖譜對比結果見圖七。

圖七 有偏誤話題關鍵詞依存關係圖譜

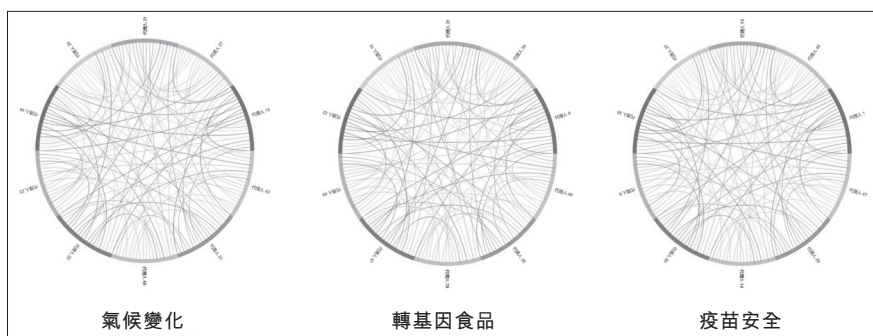


《傳播與社會學刊》，（總）第78期（2026）

在無偏誤資訊的情況下，三個議題的語義圖譜相對簡潔，各節點之間的連結較少，呈現出較為單一的意見動態。這表明，在無偏誤資訊環境中，代理的討論更集中於科學共識，資訊傳播路徑較為直接，且焦點明確。相反地，加入偏誤資訊後，語義圖譜變得更加複雜，各節點之間的連結顯著增加，顯示出更多的意見分化和多樣化的討論路徑。由此可見，偏誤資訊的引入不僅使討論更為頻繁、話題更趨多樣，也體現出確認偏誤和資訊噪音對意見動態的顯著影響。

偏誤資訊對議題語義網絡的結構造成了擾動。具體來看，加入偏誤資訊後，「氣候變化」、「轉基因食品」和「疫苗安全」三個議題的語義圖譜中，中心節點的數量增加，且這些節點之間的連結變得更加密集。例如，在「氣候變化」議題中，偏誤資訊的加入催生了更多反對科學共識的討論節點，說明其有效觸發了對立觀點的傳播。「轉基因食品」與「疫苗安全」議題的語義圖譜亦呈現出類似趨勢：偏誤資訊擴展了討論的廣度與深度，同時也帶來更多爭論與分歧。研究進一步對有偏誤狀態下三個科學共識議題中的活躍代理進行分析。具體做法是計算每個代理的互動次數，並選取排名前十的代理，過濾出僅包含這些高頻代理及其關聯邊的資料，再使用Holoviews將圖形轉換為弦圖(chord diagram)，從而得到有偏誤狀態下的活躍代理共現圖，如圖八所示。

圖八 有偏誤社交網絡的活躍代理共現圖



LLMs 模擬意見動態和偏見信息的影響

儘管討論議題不同，三個議題的共現圖卻呈現出一定的一致性。代理 29、代理 47 等在多個議題中均處於高頻互動位置，表現出跨議題的活躍度與影響力。這一現象表明，部分代理在社交網絡中具有較強的傳播能力與影響力，不僅能在單一議題中引導討論，還能跨議題擴散資訊，從而可能促進偏誤資訊在網絡中的廣泛傳播與討論。

本研究進一步採用線性混合效應模型 (mixedLM) 來分析信念變化情況，其中因變項為信念分值 (belief_score)。模型中包含隨機效應代理編號 (agent_id)，即每個代理皆具有其自身的隨機截距，用以控制個體差異。此外，模型設定了若干固定效應：時間變項天數 (day)、資訊條件的有偏誤與無偏誤 (misinfo_condition)，以及支持、反對、中立 (nitial_belief)。為探究條件間可能存在的交互作用，模型亦納入三類交互效應：

- 1) misinfo_condition $\times$ initial_belief、day $\times$ misinfo_condition
- 2) day $\times$ initial_belief

假設資料包含 50 個代理、30 天、2 種條件，共計 3,000 筆紀錄。隨機效應的變異數 (group var = .211) 顯示，不同代理之間存在較大的個體差異。模型截距為 .52 ($p < .001$)，表示參照組的平均初始信念水準。此處的參照組為偏誤資訊條件 (bias) 且初始信念為中立 (neutral) 的代理。換言之，在實驗起始日 (day 0)，中立立場的代理在有偏誤資訊條件下的平均信念約為 .52。這一截距為後續解釋固定效應與交互效應提供了基準。

如表六、表七所示，在無偏誤資訊條件下，截距為 .89；而在有偏誤資訊條件下，截距為 .85，表明在缺乏偏誤資訊時，代理的初始信念平均值略高，較傾向於接受科學共識。時間固定效應在兩種條件下均為負值，但在有偏誤資訊條件下下降幅度更大 ($-.008$ vs $-.006$)，顯示偏誤資訊對信念具有顯著負面影響，並加速信念分值隨時間的下降。交互效應分析表明，時間乘以偏誤資訊條件的交互作用達顯著水準 ($p < .001$)，說明在偏誤資訊干擾下，信念隨時間下降的速度更快；時間乘以初始信念分佈的交互效應亦達顯著水準 ($p < .01$)，表明初始信念分佈會調節信念演化的速度與方向。此外，偏誤資訊條件乘以初始信念分佈的交互效應顯示，在不同初始信念組合下，偏誤資訊對群體信念皆具有負面影響，但其程度存在差異。

《傳播與社會學刊》，（總）第 78 期（2026）

表六 無偏誤「氣候變化」代理信念狀態 LMM 分析結果

Model	MixedLM		Dependent Variable			信念得分	
觀察值數	1,500.00		Method			REML	
組別數	30.00		Scale			0.32	
最小組別樣本數	50.00		Log-Likelihood			−1,294.80	
最大組別樣本數	50.00		Converged			Yes	
平均組別樣本數	50.00						
參數	Coef.	Std.Err.	z	P> z	[0.025	0.975]	
截距	.89	.04	20.72	.00	.80	.97	
日期	−.01	.00	−6.33	.00	−.01	−.00	
組間變異	.03	.02					

組間變異 (group var) 方面，在無偏誤條件下為 .03，而在有偏誤條件下為 .02，顯示在無偏誤情境中，個體信念變化的差異性更大。這可能意味著，在缺乏偏誤資訊干擾的環境下，個體更容易受到多元資訊的影響，從而呈現出較高的變異程度；相對地，在存在偏誤資訊的環境中，偏誤資訊的傳播使個體信念變化趨於一致，差異性因而減小。這一結果提示，偏誤資訊不僅影響個體信念的方向，還可能使信念變化呈現出更高的同質化趨勢。換言之，在有偏誤資訊條件下，組間變異的降低意味著個體信念更趨一致，顯示偏誤資訊不僅影響信念方向，也可能壓縮意見分佈，導致集體認知出現同質化傾向。

表七 有偏誤「氣候變化」代理信念狀態 LMM 分析結果

Model	MixedLM	Dependent Variable			信念得分	
觀察值數	1,500.00	Method			REML	
組別數	30.00	Scale			.39	
最小組別樣本數	50.00	Log-Likelihood			−1440.10	
最大組別樣本數	50.00	Converged			Yes	
平均組別樣本數	50.00					
參數	Coef.	Std.Err.	z	P> z	[0.025	0.975]
截距	.85	.04	20.26	.00	.77	.93
日期	−.01	.00	−7.05	.00	−.01	−.01
組間變異	.02	.01				

LLMs 模擬意見動態和偏見信息的影響

表八 有偏誤「氣候變化」代理信念狀態 LMM 分析結果

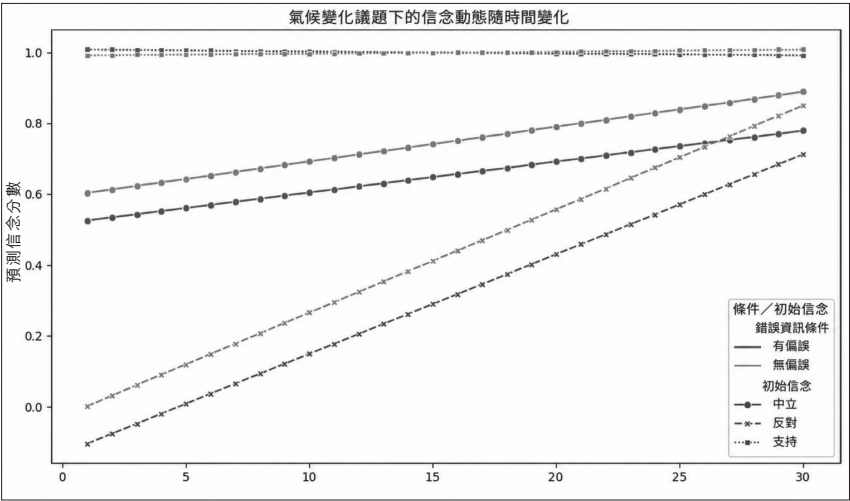
	coef	std_err	z_value	p_value	ci_lower	ci_upper
組間變異		.47	4.92	.00	1.40	3.24
截距	.52	.11	4.92	.00	.31	.72
日期	.01	.00	7.34	.00	.01	.01
日期 × 初始信念 (反對)	.02	.00	12.68	.00	.02	.02
日期 × 初始信念 (支持)	-.01	.00	-6.04	.00	-.01	-.01
日期 × 偏誤資訊條件 (無偏誤)	.00	.00	.86	.39	-.00	.00
初始信念 (反對)	-.65	.16	-4.08	.00	-.96	-.34
初始信念 (支持)	.49	.16	3.08	.00	.18	.80
偏誤資訊條件 (無偏誤)	.08	.03	2.90	.00	.02	.13
偏誤資訊條件 (無偏誤) × 初始信念 (反對)	.03	.03	1.09	.28	-.02	.08
偏誤資訊條件 (無偏誤) × 初始信念 (支持)	-.09	.03	-3.51	.00	-.15	-.04

表七、八的數據顯示，截距項為 .52 ($p < .001$)，表明參照組 (中立初始信念、有偏誤資訊條件) 在實驗起始階段的平均信念水準約為 .52。偏誤資訊條件效應達顯著水準 (無偏誤資訊組較有偏誤資訊組的信念平均高出 .08, $p = .004$)，提示偏誤資訊會略微降低代理對科學共識的信念。初始信念效應亦表現顯著：初始支持者的信念高於中立者 (係數 .49, $p = .02$)，初始反對者的信念低於中立者 (係數 -.65, $p < .001$)，說明初始信念在信念演化過程中發揮重要作用。

交互效應分析發現，對於初始支持者而言，在無偏誤資訊條件下其信念略微下降 (係數 -.09, $p < .001$)；而對於初始反對者而言，偏誤資訊的調節作用並不顯著 ($p = .277$)，說明偏誤資訊對不同初始信念群體的影響存在差異。時間效應整體達顯著水準 (係數 .00, $p < .001$)，且時間與初始信念的交互作用亦達顯著水準：反對者的信念隨時間增加 (每天增加 .02, $p < .001$)，逐漸向科學共識靠攏；而支持者的信念則隨時間略微下降 (每天下降 .02, $p < .001$)，反映群體討論可能引發一定程度的中性化趨勢。時間與偏誤資訊的交互作用不顯著 ($p = .39$)，表明信念隨時間變化的速度在不同偏誤資訊條件下差異並不明顯。隨機效應變異數為 .21，說明代理之間存在較大差異，模型有效控制了個體間的異質性。

《傳播與社會學刊》，（總）第 78 期（2026）

圖九 「氣候變化」議題下代理信念隨時間變化的趨勢



圖九顯示了氣候議題下代理信念隨時間變化的趨勢。可以看出，不同初始信念組的信念水準存在明顯差異：初始支持者的信念較高且保持穩定；中立者的信念居中並略有上升；而初始反對者的信念起點較低，但隨時間顯著上升，逐漸向科學共識靠攏。偏誤資訊條件對信念演化具有調節作用：在無偏誤資訊條件下，反對者與中立者的信念上升速度更快，而支持者的信念變化不大。整體而言，時間、初始信念與資訊環境共同作用，形塑出不同群體的信念動態，並直觀驗證了線性混合效應模型結果中固定效應與交互效應的顯著性。

總體而言，LMM 模型能夠合理刻劃實驗中信念隨時間變化的動態特徵，並系統呈現固定效應及交互效應的統計顯著性。儘管引入了偏誤資訊，實驗結果顯示 LLM 代理的意見動態在最終仍呈現出真相趨向性。長期的高頻互動與資訊交換使科學共識逐步佔據主導地位，並對偏誤資訊產生自我糾正效應。這與 Nyhan 和 Reifler (2010) 在現實生活中的研究結果相符；他們指出，儘管虛假資訊在短期內可能擾亂公眾認知並引發混亂，但從長期來看，基於 LLM 代理的科學共識依然具有較強的自我糾正能力，透過持續的事實查核與科學傳播，公眾最終能夠修正錯誤認知並接受科學事實。數據結果表明，時間、偏誤資訊以及初始信念對信念水準均具有顯著影響，且不同初始信念群體在不同

LLMs 模擬意見動態和偏見信息的影響

資訊環境下的信念變化趨勢存在明顯差異。這不僅驗證了實驗設計的有效性，也為後續針對意見動態與資訊干擾機制的理論討論提供了堅實的數據支持。

總結

當下，傳統基於代理模型的意見動力學研究，雖然在模擬社會網絡內的傳播路徑和影響力擴散方面具有重要價值，但其在認知複雜性和語言表達方面的局限性日益顯現 (Macy & Willer, 2002)。而近年來，LLMs 因其強大的語言理解與生成能力，逐漸成為模擬人類語言行為和社會互動的新興工具 (Bender et al., 2021; Bommasani et al., 2021)。

Smith 和 Conrey (2007) 指出，個體意見動態在社交互動中具有較高敏感性，早期互動尤為關鍵。Sunstein (2009) 進一步強調，社交網絡中的資訊同質化與回聲室效應能夠加速共識的形成，高頻互動能顯著改變使用者的信念。Vosoughi 等人 (2018) 則指出，資訊的真實性與可靠性在意見動態中扮演重要角色，無偏誤資訊條件下的互動更能促進科學共識的接受。本文以意見動力學和資訊傳播理論為理論支撐，透過 LLM 模擬人類社交網絡中的意見動態，重點關注「氣候變化」、「轉基因食品」、「疫苗安全」三個具有科學共識議題的信念變化。實驗結果表明，即便在偏誤資訊干擾下，科學共識仍能在資訊傳播和群體互動中表現出較強的糾錯能力和真相趨向性。相關研究 (Vosoughi et al., 2018) 亦支持這一發現，指出社交網絡互動和資訊傳播對意見形成與變化具有顯著影響。這一現象說明，持續的科學傳播和事實查核能有效對抗偏誤資訊，促進真實資訊的傳播。

三個不同的實驗設置進一步驗證了大語言模型較好的適應性和穩定性：它們指出，儘管虛假資訊短期內可能擾亂公眾認知、引起混亂，但從長期來看，基於 LLM 代理的科學共識仍具有較強的自我糾正能力；透過持續的事實查核和科學傳播，公眾最終能夠修正錯誤認知，接受科學事實。從理論上看，這一結果提示，經典傳播理論中對啟發式加工主導、意見極化不可避免性的強調，可能在特定認知主體條件下具有明顯的情境依賴性。基於 LLM 代理的模擬顯示，當資訊互

《傳播與社會學刊》，（總）第78期（2026）

動主體不受情緒迴避、身分認同與社會壓力等因素影響時，意見演化更容易呈現出事實趨向與共識聚合特徵。因此，本文的發現並非否定既有理論，而是透過對比性模擬揭示這些理論賴以成立的人類認知與社會前提，為其在智能媒介環境中的再理解提供了新的視角。

本研究使用大語言模型模擬意見極化與動態變化，為理解和預測複雜社會現象（如政治極化、公眾輿論演變）提供了新的理論方法和分析工具。透過探討不同初始信念分佈和偏誤資訊干預下科學共識的傳播機制，研究識別了科學資訊傳播的障礙，提出了改進傳播策略的理論依據。同時，本研究亦存在一些局限。首先，研究所使用的GPT無法代表其他大語言模型，不同模型在相同議題上的模擬效果可能存在差異，仍需進一步驗證。其次，儘管實驗顯示在10%偏誤資訊干預下，代理意見依舊趨於真實，但偏誤資訊的比例和種類相對有限；同時，代理模型在模擬真實社交網絡中人類的複雜情感與認知反應方面仍存在不足，未能全面考慮現實世界中資訊失真的多樣性與複雜性，或許在超過某一閾值後，代理模型將趨於幻覺。因此，未來研究可考慮擴展偏誤資訊的類型與比例，並在更大規模和更長時間尺度上開展模擬，以便更全面地揭示科學共識在公眾輿論中的傳播效應與規律。

值得注意的是，儘管模擬結果表明，不論初始信念分佈如何，代理最終均傾向趨同於科學共識，在一定程度上呈現出科學資訊在網絡傳播中的「自我糾正」趨勢，但該結果主要反映的是LLMs在封閉系統中生成之代理行為所展現的內在一致性，而非對現實社會意見演化過程的直接映射。LLMs本質上依賴大量語言資料的統計規律，其「代理」表現更多反映的是預訓練語料中的語言模式，而非真正具備獨立信念、動機或社會目標的個體。由於LLMs在生成語言內容時，受到其訓練語料、對齊機制及機率建模方式的深刻影響，所模擬的代理在面對資訊時的反應更傾向於理性一致、事實導向，這與現實個體在社交環境中所表現出的認知偏誤、社群極化、情緒感染等非理性因素存在顯著差異。因此，本研究所揭示的信念聚合趨勢，應被視為模型機制下的一種結構性輸出，而非對現實社交系統的泛化性結論。

從理論層面看，本文的發現為經典傳播理論在智能媒介環境中的適用邊界提供了新的反思視角。以ELM、啟發式加工與議程設定理論

LLMs 模擬意見動態和偏見信息的影響

為代表的傳統框架，普遍假設個體在資訊處理過程中受到認知負荷、情緒動機與社會壓力的深刻制約，從而更容易依賴線索化判斷並形成意見極化。然而，基於LLM代理的模擬結果顯示，在缺乏人類情緒迴避、身分認同與現實利益驅動的條件下，意見動態呈現出更強的事實趨向性與共識聚合特徵。這一差異並非否定上述經典理論的解釋力，而是提示其核心機制高度依賴於特定的人類心理與社會條件。當資訊互動主體部分轉向具備語言生成與推理能力、卻不具備人類情緒結構的智能系統時，經典理論中有關啟發式主導與議程競爭的解釋邏輯需要被重新界定。因此，本文的理論貢獻並不在於對既有傳播理論進行再次驗證或替代，而在於透過LLM驅動的多代理模擬，提供了一種檢驗經典理論隱含前提的「對照性認知環境」。這種方法使研究者能夠在去除人類情緒噪音與社會壓力變項的條件下，觀察意見演化機制的結構性表現，從而反向揭示哪些傳播效應源於資訊結構本身，哪些則高度依賴於人類認知與社會互動特性。透過這一視角，本文為理解科學共識形成、錯誤資訊干預以及意見極化問題提供了一種跨越傳統實證研究限制的理論補充路徑，也為傳播學理論在生成式人工智慧廣泛介入的資訊環境中實現再詮釋與重構奠定了方法論基礎。

需要指出的是，LLMs作為科學傳播載體的有效性，確實依賴於其在社會中的採納率及公眾對其資訊可信度的認可。因此，本研究的結論適用於那些LLMs已被一定規模使用者群體接受並頻繁使用的場景，能夠為理解和預測新興智能技術在科學資訊傳播中的作用提供重要參考。例如，部分專業領域的從業者，如法律、醫療、教育等行業人士，已開始將大型語言模型作為輔助決策和資訊檢索的重要工具；普通公眾中也出現了使用聊天機器人輔助日常學習、寫作和資訊獲取的趨勢。在這些場景下，LLMs不僅影響個體的知識獲取，也逐漸成為資訊傳播和意見形成的新節點，因而具備重要的科學傳播意義。與此同時，本研究方法突破了傳統代理模型在語言理解和複雜意見動態模擬上的限制，豐富了科學傳播的研究工具箱，對推動傳播學理論與智能技術的深度融合具有積極意義。隨著LLMs的普及和應用深化，相信相關模擬研究將在更廣泛的社會語境中發揮更大價值，為科學傳播策略的優化提供更加精準的理論支撐。

《傳播與社會學刊》，（總）第78期（2026）

基於上述討論，未來研究可進一步探索引入人類參與的混合代理模擬框架，或建構能夠隨資訊環境與互動規則動態調整的模型機制，以增強對現實社會意見演化複雜性的刻劃能力。需要強調的是，大語言模型雖為研究社會意見動態提供了前所未有的技術路徑，但其模擬結果仍深受模型訓練語料、對齊策略及生成機制的影響，難以完全涵蓋現實社會中情緒動員、權力結構與社會身分等關鍵因素。因此，相關研究應將LLM代理視為一種用於檢驗理論假設與揭示機制邊界的分析工具，而非對現實社會過程的直接替代。在持續拓展方法可能性的同時，研究者亦需保持必要的理論自覺與規範反思，審慎評估模型偏差、適用情境及潛在倫理風險，以推動傳播學理論在智能媒介深度介入背景下的穩健發展。

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

本文為湖北省教育科學規劃課題「人機共生下的『AI四能教育』改革路徑與倫理風險研究」（HBGX25007C18451）和中央高校基本科研業務費項目「基於AIGC的價值機制路徑分析與幻覺風險研究」（2662025WFQD002）之研究成果。

Funding

This article was funded by the Hubei Provincial Education Science Planning Project, “Research on the Reform Path and Ethical Risks of ‘AI Four-Competency Education’ under Human–Machine Symbiosis” (HBGX25007C18451),

LLMs 模擬意見動態和偏見信息的影響

and the Fundamental Research Funds for the Central Universities project, “Analysis of the Value Mechanism Path and Hallucination Risks Based on AIGC” (2662025WFQD002).

ORCID

陶煒 (Wei TAO) <https://orcid.org/0000-0002-6941-0353>

向安玲 (Anling XIANG) <https://orcid.org/0000-0003-1690-1586>

參考文獻

中文部分 (Chinese Section)

- 何一清、崔連廣、王迎軍 (2015)。〈知識架構、信息篩選與群體對立認知機制〉。《科學學研究》，第 10 期，頁 1447–1455。
- He, Y., Cui, L., & Wang, Y. (2015). Knowledge schema, information screening, and the cognitive mechanism of group opposition. *Studies in Science of Science*, 10, 1447–1455.
- 林慧嫻、陸靜曄 (2015)。〈基於代理的信息傳播網絡中的初始節點選擇〉。《微型機與應用》，第 4 期，頁 55–57。
- Lin, H., & Lu, J. (2015). Initial node selection in an agent-based information diffusion network. *Microcomputer and its Applications*, 4, 55–57.
- 胡泳 (2024)。〈人工智能驅動的虛假資訊：現在與未來〉。《南京社會科學》，第 1 期，頁 96–109。
- Hu, Y. (2024). Artificial intelligence-driven misinformation: The present and the future. *Nanjing Journal of Social Sciences*, 1, 96–109.
- 高嘉誠、劉鑰、田立寧 (2024)。〈互聯網使用與性別偏見認同感——理論分析與實證檢驗〉。《統計學報》，第 2 期，頁 86–94。
- Gao, J., Liu, Y., & Tian, L. (2024). Internet use and gender bias identity: Theoretical analysis and empirical test. *Journal of Statistics*, 2, 86–94.
- 郭佳霖、智敏、殷雁君、葛湘巍 (2024)。〈圖像處理中 CNN 與視覺 Transformer 混合模型研究綜述〉。《計算機科學與探索》，第 1 期，頁 1–18。
- Guo, J., Zhi, M., Yin, Y., & Ge, X. (2024). A review of CNN and visual transformer hybrid models in image processing. *Journal of Frontiers of Computer Science and Technology*, 1, 1–18.

《傳播與社會學刊》，（總）第78期（2026）

郭娜、陳佳鵬、姜小峰、李惠文、朱萍、白玥（2024）。〈疫苗猶豫對北京市老年人新冠病毒疫苗接種情況影響〉。《中國公共衛生》，第4期，頁446–449。

Guo, N., Chen, J., Jiang, X., Li, H., Zhu, P., & Bai, Y. (2024). Impact of vaccine hesitancy on COVID-19 vaccination among the elderly in Beijing: A cross-sectional survey. *Chinese Journal of Public Health*, 4, 446–449.

陳瑛瑋、許吉祥、王一、閻慧慧、高俊嶺（2021）。〈中國公眾COVID-19疫苗選擇偏好與接種意願分析〉。《復旦學報（醫學版）》，第5期，頁578–585、617。

Chen, Y., Xu, J., Wang, Y., Yan, H., & Gao, J. (2021). Public preference and vaccination willingness for COVID-19 vaccines in China. *Fudan University (Journal of Medical Sciences)*, 5, 578–585, 617.

陳慧敏、朱澤宇、于北辰、金兼斌（2023）。〈公共衛生危機中政府信任、科學家信任與社會焦慮——基於新冠肺炎疫情網絡輿論的觀察〉。《當代傳播》，第3期，頁44–52。

Chen, H., Zhu, Z., Yu, B., & Jin, J. (2023). Government trust, scientist trust, and social anxiety during public health crises: An observation based on online public opinion during the COVID-19 pandemic. *Contemporary Communication*, 3, 44–52.

許志偉、李海龍、李博、李濤、王嘉泰、謝學說、董澤輝（2024）。〈AIGC大模型測評綜述：使能技術、安全隱患和應對〉。《計算機科學與探索》，第9期，頁2293–2325。

Xu, Z., Li, H., Li, B., Li, T., Wang, J., Xie, X., & Dong, Z. (2024). A survey of AIGC large model evaluation: Enabling technologies, security risks, and countermeasures. *Journal of Frontiers of Computer Science and Technology*, 9, 2293–2325.

張霽陽、張鵬、李思佳、劉靜、竇雲蓮、蘭月新（2023）。〈基於實時社會網絡分析的突發輿情事件動態意見領袖識別方法研究〉。《情報雜誌》，第9期，頁109–116、126。

Zhang, J., Zhang, P., Li, S., Liu, J., Dou, Y., & Lan, Y. (2023). Research on dynamic opinion leader identification in public opinion emergencies based on real-time social network analysis. *Journal of Intelligence*, 9, 109–116, 126.

賀琴、劉志中、趙美悅、孟令強（2024）。〈社交網絡中動態信任感知的觀點演化模型〉。《南京大學學報（自然科學）》，第3期，頁357–369。

LLMs 模擬意見動態和偏見信息的影響

- He, Q., Liu, Z., Zhao, M., & Meng, L. (2024). Dynamic trust-aware opinion evolution model in social networks. *Journal of Nanjing University (Natural Sciences)*, 3, 357–369.
- 夢非、朱慶華 (2021)。〈社交網絡信息傳播中意見偏差的國外研究進展〉。《情報理論與實踐》，第10期，頁193–201。
- Meng, F., & Zhu, Q. (2021). A review of international research on opinion bias in social network information dissemination. *Information Studies: Theory and Application*, 10, 193–201.
- 謝天、邱林、李雨瞳、羅殷、劉盼 (2023)。〈大模型時代的社會科學，何去何從？〉。《圖書情報知識》，第6期，頁6–9、30。
- Xie, T., Qiu, L., Li, Y., Luo, Y., & Liu, P. (2023). The landscape of social science in the era of large language models? *Documentation, Information and Knowledge*, 6, 6–9, 30.

英文部分 (English Section)

- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211–236.
- Anagnostopoulos, A., Bessi, A., Caldarelli, G., Del Vicario, M., Petroni, F., Scala, A., ... & Quattrociocchi, W. (2014). Viral misinformation: The role of homophily and polarization. *arXiv*. <https://arxiv.org/abs/1411.2893>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (pp. 610–623). ACM.
- Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, 49(2), 192–205.
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., ... Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv*. <https://arxiv.org/abs/2108.07258>.
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, 33, 1877–1901.
- Budak, C., Goel, S., & Rao, J. M. (2016). Fair and balanced? Quantifying media bias through crowdsourced content analysis. *Public Opinion Quarterly*, 80(S1), 250–271.

《傳播與社會學刊》· (總) 第 78 期 (2026)

- Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2), 591–646.
- Chuang, Y.-S., Goyal, A., Harlalka, N., Sures, S., Hawkins, R., Yang, S., Shah, D., Hu, J., & Rogers, T. T. (2024). Simulating opinion dynamics with networks of LLM-based agents. *arXiv*. <https://doi.org/10.48550/arXiv.2311.09618>.
- Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, 55, 591–621.
- Epstein, J. M. (2012). *Generative social science: Studies in agent-based computational modeling*. Princeton University Press.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7(2), 117–140.
- Flache, A., Mäs, M., Feliciani, T., Chattoe-Brown, E., Deffuant, G., Huet, S., & Lorenz, J. (2017). Models of social influence: Towards the next frontiers. *Journal of Artificial Societies and Social Simulation*, 20(4), 2.
- Friedkin, N. E., & Johnsen, E. C. (2011). *Social influence network theory: A sociological examination of small group dynamics*. Cambridge University Press.
- Gilbert, N., & Terna, P. (2000). How to build and use agent-based models in social science. *Mind and Society*, 1(1), 57–72.
- Ginossar, T., Cruickshank, I. J., Zheleva, E., & Sulskis, J. (2022). Cross-platform spread: Vaccine-related content, sources, and conspiracy theories in YouTube videos shared in early Twitter COVID-19 conversations. *Human Vaccines & Immunotherapeutics*, 18(1), 1–13.
- Katz, E., & Lazarsfeld, P. F. (1955). *Personal influence: The part played by people in the flow of mass communications*. Free Press.
- Lewandowsky, S., Gignac, G. E., & Oberauer, K. (2013). The role of conspiracist ideation and worldviews in predicting rejection of science. *PLoS One*, 8(10), e75637.
- Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K., & Larson, H. J. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature Human Behaviour*, 5(3), 337–348.
- Macy, M. W., & Willer, R. (2002). From factors to actors: Computational sociology and agent-based modeling. *Annual Review of Sociology*, 28, 143–166.
- McQuail, D. (2010). *McQuail's mass communication theory* (6th ed.). Sage Publications.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220.
- Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), 303–330.
- Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. In

LLMs 模擬意見動態和偏見信息的影響

- Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)* (pp. 1–22). ACM.
- Pennycook, G., Bear, A., Collins, E. T., & Rand, D. G. (2020). The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management Science*, 66(11), 4944–4957.
- Smaldino, P. E. (2023). *Modeling social behavior: Mathematical and agent-based models of social dynamics and cultural evolution*. Princeton University Press.
- Smith, E. R., & Conrey, F. R. (2007). Agent-based modeling: A new approach for theory building in social psychology. *Personality and Social Psychology Review*, 11(1), 87–104.
- Sunstein, C. R. (2009). *Going to extremes: How like minds unite and divide*. Oxford University Press.
- Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining “fake news”: A typology of scholarly definitions. *Digital Journalism*, 6(2), 137–153.
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151.
- Xia, L.-L., Jiang, G.-P., Song, B., & Song, Y.-R. (2015). Rumor spreading model considering hesitating mechanism in complex social networks. *Physica A: Statistical Mechanics and its Applications*, 437, 295–303.

本文引用格式

陶煒、向安玲 (2026)。〈LLMs 模擬意見動態和偏見信息的影響：基於「氣候變化」、「轉基因食品」和「疫苗安全」議題的多代理研究〉。《傳播與社會學刊》，第 78 期，頁 1–45。

Citation of This Article

Tao, W., & Xiang, A. (2026). Simulating opinion dynamics and the impact of biased information using LLMs: A multi-agent study on climate change, GM foods, and vaccine safety. *Communication and Society*, 78, 1–45.