ISSN 2759-7830 (Online)
ISSN 2760-2508 (Print)

JOURNAL OF GLOBAL TRENDS IN SOCIAL SCIENCE

MARCH 2026 VOLUME 03 NUMBER 02

JANDOO PRESS
株式会社間渡出版

https://jandoopress.com./journal/jgtss

Issue

Volume 3, Number 2 (March 2026)

Publisher

株式会社間渡出版 (Jandoo Press Co., Ltd.)
Tokyo office: 1-53-13 Nishigahara, Kita City, Tokyo 114-0024, Japan
Email: contact@press.jandoo.ac

Journal info

Portal: https://jandoopress.com/journal/jgtss
ISSN 2759-7830 (Online)
ISSN 2760-2508 (Print)

Journal statement

1. The views, interpretations, and conclusions expressed in the articles published in the Journal of Global Trends in Social Science (JGTSS) are solely those of the authors and do not necessarily represent those of the Editorial Board, the Editorial Office, or the publisher. Authors bear full responsibility for the accuracy, integrity, and legality of the content they submit and publish.
2. JGTSS is an open access journal. Unless otherwise stated, all articles published in the journal are made available under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License (https://creativecommons.org/licenses/by/4.0/). Under this license, anyone may read, download, copy, distribute, reproduce, print, adapt, and reuse the published material in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are clearly indicated.
3. Authors retain copyright in their work. By submitting a manuscript to JGTSS and agreeing to its publication upon acceptance, authors grant the journal the right of first publication and a non-exclusive license to publish, disseminate, archive, index, preserve, and display the work in electronic, online, and related formats for academic communication and public access.
4. Authors must ensure that their submissions are original, do not infringe any copyright or other legal rights of third parties, and are not under consideration elsewhere at the time of submission, unless clearly disclosed to the journal. Where images, tables, long quotations, or other third-party materials are included, authors are responsible for obtaining any necessary permissions prior to submission and for providing appropriate acknowledgements where required. Any dispute arising from copyright infringement, academic misconduct, or other legal deficiencies remains the responsibility of the author(s).
5. All submissions are subject to editorial review and peer review in accordance with the journal's policies. The journal reserves the right to make necessary editorial revisions to accepted manuscripts for clarity, consistency, language quality, formatting, referencing style, and house style, provided that such revisions do not alter the academic substance of the work. Substantive changes affecting content or interpretation will be made only in consultation with the author(s).
6. A DOI is assigned to each published article to support persistent identification, citation, dissemination, and long-term accessibility. Once an article is formally published, the digital version hosted on the journal's official website shall be regarded as the version of record.
7. Manuscripts must be submitted through the journal's official submission channel or website. Submitted materials are generally not returned. If no decision is communicated within three months of submission, authors may withdraw the manuscript and submit it elsewhere.
8. JGTSS is committed to the principles of academic integrity and publication ethics. In cases of plagiarism, fabrication, falsification, duplicate submission, improper authorship, or other forms of academic misconduct, the journal reserves the right to reject, retract, correct, or otherwise address the publication in accordance with established editorial and ethical standards.

Official print edition notice

1. In addition to the digital edition published on the journal's official website, the journal may produce a limited number of official print copies primarily for submission, archiving, and review purposes. Official printed copies may also be provided to article authors upon request.
2. The availability of official print copies does not alter the open-access status of the content. All published materials remain available under the Creative Commons Attribution 4.0 International License (CC BY 4.0), and may therefore be downloaded, reproduced, printed, and reused in accordance with the terms of that license.
3. Where authors request official printed copies, the related production and delivery costs shall be borne by the requesting author(s). Charges are determined according to the length of the issue concerned. Please contact the Editorial Office for further details.
4. For citation, indexing, and version-control purposes, the digital version published on the journal's official website shall be regarded as the version of record. Printouts made from downloaded PDF files do not constitute official print editions.

JOURNAL OF GLOBAL TRENDS IN SOCIAL SCIENCE

The Journal of Global Trends in Social Science (JGTSS) is an international, peer-reviewed, open-access venue committed to identifying and analyzing transformative global trends within the social sciences. JGTSS distinguishes itself by championing interdisciplinary synthesis, specifically fostering the nexus between social scientific inquiry and technological advancement.

Official Partner

Center for Surrounding Communication Studies, Peking University
China Institute of Surrounding Communication of Geographical Indications

Contents

JOURNAL OF GLOBAL TRENDS IN SOCIAL SCIENCE

J G T S S

Research article

https://doi.org/10.70731/385e0a78

Falling Leaves, Unrooted Lives: Media Use, Power Asymmetries, and the Cross-cultural Adaptation of Vietnamese Brides in China

Tsun Wong [1,*], **Xunfan Chen** [2], **Feilong He** [3]

Received 1 February 2026

Accepted 9 March 2026

Published 31 March 2026

Abstract: The increasing flow of marriage migration from Vietnam to China has drawn scholarly attention to the figure of the Vietnamese bride, women whose lives are suspended between two nations, cultures, and systems of belonging. While prior studies have examined their economic adaptation and social identity, less is known about how their everyday media practices shape and are shaped by underlying power asymmetries. Through a year-long ethnographic study in Daxin County, Guangxi, including in-depth interviews with 12 Vietnamese brides, this paper explores how media serve as both a resource for and a constraint on cross-cultural adaptation. We introduce the twin lenses of strategic and affective media engagement to analyze how these women navigate the double bind of host-society exclusion and home-society stigma. Findings reveal that media use not only facilitates linguistic acquisition, social networking, and identity re-articulation but also reinforces structural marginalization through digital divides and symbolic violence. The study identifies three distinct patterns of media engagement: strategic integration, affective preservation, and resistant appropriation. By framing media practice within a hierarchy of citizenship and cultural power, this research contributes to a critical reappraisal of cross-cultural communication, one that situates technology within the geopolitics of intimacy and the political economy of marriage migration. The findings challenge linear models of migrant adaptation and call for more nuanced understanding of digital agency within constrained circumstances.

Keywords: Media engagement; Cross-cultural adaptation; Vietnamese brides; Strategic and affective practice; China-Vietnam borderlands

ISSN 2759-7830 (Online)
ISSN 2760-2508 (Print)

Introduction

"My phone is the only place where I can be both Vietnamese and not an outsider." This reflection from a 29-year-old Vietnamese bride in Guangxi encapsulates the paradoxical role of media in the lives of cross-border marriage migrants. Over the past three decades, the rising tide of cross-border marriages between Vietnamese women and Chinese men has reshaped demographic and social landscapes in China's southern borderlands, embodying what Constable (2005, p. 10) terms "spatial hypergamy", which refers to a gendered mobility pattern where women from economically marginalized regions marry into households in more developed economies. In Daxin County, Guangxi, a region sharing a 40-kilometer border with Vietnam, Vietnamese brides have become a highly visible yet socially peripheral group, their membership in both host communities and home societies remaining partial, contested, and perpetually negotiated.

This phenomenon unfolds amid a context of structural inequality: demographic imbalances and rural male marriage squeezes in China, paired with regional economic disparities and informal kinship networks bridging the China-Vietnam border, have normalized commercialized matchmaking processes that often reduce marital relations to transactional arrangements (Peng Y. 2018). Most Vietnamese brides arrive with limited Mandarin proficiency and minimal exposure to local customs, rendering media an indispensable tool for daily navigation. However, their media practices remain insufficiently explored as they oscillate between assimilationist demands and the need for cultural continuity, particularly when these practices intersect with power asymmetries embedded

[1] Xiamen University, Xiamen 361005, China; [2] Jinan University, Guangzhou 510632, China; [3] China West Normal University, Nanchong 637009, China.
* Corresponding author. Email: wongtsun@163.com

within state policies, patriarchal norms, and the governance of digital platforms.

This gap is notable given the growing scholarly focus on migration and digital media. Existing studies highlight how transnational migrants use media to maintain homeland ties and facilitate integration (Harpaz & Mateos, 2019), while others document how migrant women deploy subtle tactics, what Scott (1990) terms "hidden transcripts", to resist familial control (Shen, 2008; Wang, 2001). However, these works often adopt an instrumental lens, neglecting the contradictory ways media simultaneously empower and constrain, or fail to account for the unique geopolitical and cultural dynamics of Sino-Vietnamese marriage migration. Moreover, existing research on cross-border marriage primarily centers on economic motivations and legal precarities (Kim & Kilkey, 2018), sidelining the role of media in shaping everyday experiences of belonging.

This paper addresses this gap by investigating how Vietnamese brides in rural China engage with media amid structural constraint. We introduce the twin concepts of strategic and affective media engagement to analyze their responses to "double stigmatization": they are racialized as cultural outsiders in China while being framed as "traitors" or "materialists" in Vietnamese public discourse (Chiu & Yeoh, 2021). Drawing on Giddens' structuration theory and a critical media engagement perspective, we examine how media practices both reproduce and resist power asymmetries, and how these women navigate the tension between assimilation and cultural preservation.

This study focuses on Daxin County for two key reasons. First, the region's high concentration of Vietnamese brides (accounting for approximately 15% of local marriages) and long-standing cross-border ties provide a rich context for exploring media's role in adaptation. Second, the coexistence of traditional kinship networks and digital platforms in the region allows for analysis of how old and new communication infrastructures intersect to shape migrant experiences. Through a year-long ethnography involving 12 in-depth interviews and participatory observation, we pose three core research questions:

RQ1. How do Vietnamese brides engage with media strategically to secure social resources, legal recognition, and cultural legitimacy in China?

RQ2. In what ways do affective attachments to Vietnam shape their digital practices, and how do these practices negotiate domestic and transnational power dynamics?

RQ3. What patterns of media engagement emerge from these negotiations, and how do they reflect the tension between agency and structural constraint?

By centering the voices of Vietnamese brides, this research contributes to debates on media, migration, and identity, while challenging Western-centric models of transnational media use. It also offers insights for policymakers seeking to support marriage migrants in an era of digital connectivity and globalized intimacy.

Literature Review

Media, migration, and the paradox of agency

Research on migration and digital media has long emphasized a dual role: media as a tool for maintaining transnational ties and as a resource for host-society integration (Harpaz & Mateos, 2019). For migrant women, in particular, digital platforms offer spaces to alleviate homesickness, sustain familial bonds, and mitigate the uncertainty of new environments (Fortier, 2017). However, these practices are not without contradiction. State-led initiatives, such as adaptation classes for foreign brides, often position media as disciplinary instruments, framing digital literacy as a means to assimilate migrant women into normative roles as "dutiful wives" and "modern citizens" (Sara Liao, 2019).

This paradox reflects broader tensions in scholarship on migrant agency. On one hand, scholars highlight how migrant women use media to exercise tactical resistance: forming secret networks to share grievances, curating online personas to counter stereotypes, and leveraging digital tools to access labor rights (Scott, 1990; Dymess & Sepúlveda, 2020). On the other hand, critical scholars point out that media platforms themselves are embedded with power structures, including algorithmic bias, linguistic hierarchies and geopolitical constraints, which exacerbate marginalization. (van Dijck, 2013). For example, Chinese platforms like WeChat and Douyin prioritize Mandarin content and national narratives, while accessing Vietnamese platforms like Zalo often requires VPNs, imposing additional digital labor on migrants (Nguyen et al., 2023).

These dynamics are amplified in the context of cross-border marriage migration in Asia, where citizenship regimes rooted in jus sanguinis and patriarchal family structures place migrant women in legally ambiguous positions (Chiu & Yeoh, 2023). Existing studies on Vietnamese brides in Taiwan, China and Korea document how media use becomes a site of negotiation: women may use local platforms to comply with familial expectations while clandestinely accessing homeland media to preserve cultural identity. Yet, these works rarely explore the interplay of infrastructural, algorithmic, and patriarchal power in shaping media practices—a gap this study addresses.

Double stigmatization and the politics of belonging

Cross-border marriage migrants face a unique form of "double stigmatization" that shapes their media engagement. In host societies, they are often racialize as cultural outsiders, with their loyalty and cultural competence questioned (Kim & Kilkey, 2018). In home societies, they may be stigmatized as "abandoning" their culture for economic gain, framed as threats to national identity. This dual othering is compounded by the commercialized nature of cross-border matchmaking, which reduces their agency to transactional choices and reinforces stereotypes of passivity.

Scholars have noted how migrant women navigate this double bind through identity work, but few have linked this to media practice. Fortier (2017) argues that affective attachments to homeland media constitute a form of "affective citizenship," allowing migrants to maintain a sense of belonging amid displacement. Harpaz and Mateos (2019) complement this with the concept of "strategic citizenship," where migrants leverage media to accumulate cultural capital and secure social acceptance. Our twin concepts of strategic and affective media engagement build on these insights, framing them as overlapping, dynamic practices that respond to intersecting hierarchies of citizenship, cultural prestige, and familial authority.

De-westernizing the study of transnational media use

Existing research on migrant media use is predominantly rooted in Western contexts, where liberal citizenship regimes and open digital infrastructures shape engagement patterns (Harpaz & Mateos, 2019). In non-Western contexts, however, state control over digital platforms, patriarchal family structures, and economic disparities produce distinct dynamics. For example, in Latin America, scholars have documented how market-driven media initiatives frame women's empowerment through consumption, while neglecting structural inequality (Verónica Schild, 2015). In Asia, studies on soap operas and social media suggest that media are often mobilized to reinforce national identity and gender norms, rather than challenge them.

This study contributes to de-Westernizing the field by focusing on Sino-Vietnamese marriage migration, a context where state power, platform governance, and patriarchal norms intersect in unique ways. By examining how Vietnamese brides navigate survival within China's digital ecosystem, we offer fresh insights into the constraints and possibilities of digital agency within constrained environments.

Study Context and Methodological Approach

Research site: Daxin county, Guangxi

Daxin County, located in the Guangxi Zhuang Autonomous Region, is a key site for Sino-Vietnamese cross-border marriages, shaped by its geographic proximity to Vietnam, regional economic disparities, and informal kinship networks. Over the past decade, local government records indicate that approximately 15% of marital unions in the county involve Vietnamese brides, with numbers rising due to intensified cross-border economic ties. Marriages are typically arranged through informal matchmakers, former Vietnamese immigrants or local relatives, characterized by rapid matchmaking with minimal pre-marital interaction. Most brides arrive with limited Mandarin proficiency and little exposure to local customs, making media central to their daily adaptation.

Data collection

This study employs a multi-method qualitative approach, grounded in twelve months of ethnographic fieldwork (June 2023–May 2024). Data sources include in-depth interviews with 12 Vietnamese brides and participatory observation in community spaces, households, and online environments. Participants were selected via purposive sampling to ensure heterogeneity across age (19–50 years), duration of residence (under 5 years to over 20 years), educational background (primary to university level), and socioeconomic status (Table 1). All participants resided in Daxin County and had entered marriages with Chinese men through either broker-facilitated or kinship-mediated arrangements.

Semi-structured interviews (60–120 minutes) were conducted in Mandarin or Vietnamese with bilingual interpreters, focusing on daily media routines, platform preferences (WeChat, Douyin, Zalo, etc.), communication patterns with Vietnamese and Chinese networks, and experiences of inclusion/exclusion. Interviews were transcribed verbatim and analyzed using NVivo software, with iterative coding to identify themes related to media agency, structural constraint, and affective practice. Pseudonyms were used to protect confidentiality, and location details were generalized.

Participatory observation complemented interview data: the research team accompanied participants to community events, household gatherings, and online interactions, documenting how media use intersects with daily life. This approach fostered "situated empathy" (Chiu, Yeoh, 2023), centering participants as experts in their lived experiences while critically engaging with structural forces.

Data analysis

Analysis drew on structuration theory (Giddens, 1984) to explore how media practices are shaped by—and reshape—social structures (gender norms, immigration policies, platform algorithms). We also employed comparative thematic analysis (Charmaz, 2006) to identify patterns of media engagement across participants, contrasting interview data with observational field notes to triangulate findings. Special attention was paid to how participants navigated platform-specific features (e.g., WeChat groups, Douyin algorithms) and technical barriers (e.g., VPN use for Vietnamese platforms), and how these practices reflected negotiations of power (**Table 1**).

Findings: Navigating Double Stigmatization Through Media

Strategic media engagement: Calculative assimilation and symbolic capital

Strategic media engagement emerged as a core practice for securing social acceptance and resources, reflecting what we term "calculative agency", defined as deliberate efforts to accumulate cultural and social capital through digital tools. For linguistic acquisition, participants leveraged Douyin as a "self-directed language laboratory," focusing on colloquial speech rather than formal Mandarin. XLY (29), a secondary

Table 1 | Key Demographic Information of In-Depth Interviewees

No.	Code	Age	Education Level	Occupation	Years in China
1	PSY	43	Primary	Farming	Over 10 years
2	ZQQ	48	Junior High	Farming	Over 10 years
3	HYY	34	Primary	Office Worker	Over 10 years
4	LKS	21	Junior secondary	Freelancer	Under 5 years
5	NTT	50	Primary	Farming	Over 10 years
6	NM	38	Primary	Office worker	Under 10 years
7	ZZW	39	Primary	Farming	Over 10 years
8	XLY	29	Secondary school	Civil servant	Over 10 years
9	ZK	28	Primary	Freelancer	Less 10 years
10	CXF	19	High school	Dropped out	Less 5 years
11	RZL	27	High school	Office worker	Less 10 years
12	XJY	50	Primary	Farming	Over 20 years

school graduate, explained: "I watch short videos of Chinese families cooking, chatting, even arguing. I repeat the phrases and practice the tones. It's like having a window into how people really talk, not just textbook language." This aligns with Nguyen's (2023) concept of "linguistic shadowing," a tactic to reduce accent bias and integrate into local discourse.

Participants also curated social media personas to counter stereotypes of "perpetual foreigners," engaging in "performative assimilation" (Scott, 1990) as a means to present themselves as dutiful wives and community members. HYY (34), a mother of two, joined seven local WeChat parenting groups to learn Chinese child-rearing norms and demonstrate commitment to "raising children the Chinese way." XLH (29) described curating her WeChat Moments to "only share photos of me cooking Chinese dishes, helping my husband in the fields, or celebrating Chinese festivals—never anything about missing Vietnam." This self-censorship was a strategic choice to deflect suspicion and assert belonging, particularly among in-laws and neighbors.

Notably, strategic engagement correlated with length of residence: women living in China for over a decade (e.g., XJY, 50) were more likely to prioritize Chinese platforms, viewing media as a tool for socioeconomic advancement. XJY, who had resided in China for 20 years, stated: "I only use Chinese apps now—WeChat for neighbors, Douyin for recipes, Meituan for shopping. To be part of this community, you have to think and act Chinese, even online." Yet, this assimilation came at a psychological cost: several participants reported "affective dissonance"—the strain of suppressing cultural attachments to maintain a strategic facade. One woman confessed: "Sometimes I feel like I'm erasing myself to be accepted. But what choice do I have? My children's future is here."

Affective media engagement: Emotional lifelines and transnational bonds

In contrast, affective media engagement centered on preserving cultural identity and emotional connection to Vietnam, functioning as "emotional lifelines" amid displacement. All participants used Zalo or Facebook (via VPN) to maintain daily contact with family, while consuming Vietnamese dramas, music, and vlogs to sustain cultural continuity. For newer arrivals like CXF (19), who struggled with intense homesickness, this was non-negotiable: "I spend hours every day watching Vietnamese variety shows on YouTube and talking to my family via Zalo. It's the only thing that keeps me from feeling completely lost."

These practices created "affective sanctuaries", digital spaces where participants could escape Hochschild's (1983) "emotional labor" and express their authentic selves. NM (38), who lived in a mountainous village, shared: "At night, I watch Vietnamese dramas and cry without judgment. It's my secret therapy." However, such engagement often sparked domestic conflict, as husbands and in-laws interpreted frequent Vietnam-focused media use as "disloyalty" or "failure to adapt." CXF (19) recounted how her husband monitored her Zalo messages, forcing her to use coded language and delete chat histories. This reflects how media use becomes a site of negotiation over belonging, with transnational ties policed as threats to familial authority (Shen, 2008).

Affective engagement also required technical workarounds: participants developed "digital border-crossing skills," using multiple VPNs to access geographically restricted content and switching between Chinese and Vietnamese platforms. One woman explained: "I use one VPN for Facebook, another for Zalo, then switch to WeChat for daily life. It's exhausting, but necessary to stay connected to who I am." This digital labor underscored the emotional and technical burdens of maintaining transnational identity in a constrained media environment.

Resistant appropriation: Media as tools of collective agency

Beyond assimilation and emotional solace, participants employed "resistant appropriation", a practice of adapting Chinese platforms to challenge marginalization and build community. Drawing on Scott's (1990) "hidden transcripts," women formed private WeChat groups with fellow Vietnamese brides to share grievances and strategies. ZZW (39) organized a clandestine network to address underpayment in local farms: "We use WeChat to share job information and negotiate wages together—alone, we have no power, but as a group, we can push back."

Others used Douyin to subvert stereotypes, posting videos of themselves succeeding in local businesses to refute narra-

tives of passivity. RZL (27), who ran a small grocery store, explained: "I post videos of myself managing the store, talking to customers in Mandarin. It shows I'm not just a 'Vietnamese bride'—I'm a business owner." These practices created "third spaces" (Bhabha, 1994), hybrid digital environments where participants blended Vietnamese and Chinese cultural elements to craft new forms of identity. One woman described teaching Chinese neighbors Vietnamese recipes via WeChat while sharing Chinese parenting tips with Vietnamese friends, creating a "digital kinship" network (Dymess & Sepúlveda, 2020) that transcended national boundaries.

Yet, resistant practices operated within strict limits. Participants self-censored to avoid retaliation, keeping groups small and private. ZZW warned: "We have to be careful—if our groups get too visible, people might complain to the village committee. We're like a digital family, but we can't let anyone outside in." This caution reflects the structural constraints of their environment, where even subtle resistance carries risks of social sanction or familial conflict.

Power geometries in digital space: Infrastructural, algorithmic, and patriarchal control

Vietnamese brides' media practices are embedded in what Massey (1994) terms "power geometries"—overlapping systems of state, platform, and familial power that shape digital engagement. Three interconnected dimensions of power emerged from our analysis, highlighting the paradox of media as both enabling and constraining.

First, infrastructural power manifested in digital inequality: brides from low-income households relied on outdated smartphones and limited data plans, restricting access to video calls and data-intensive content. Those in remote villages faced erratic internet service, exacerbating isolation. NM (38) noted, "When the weather is bad, the internet disappears. I can't even send a simple message to my family, let alone see their faces." This "algorithmic violence" (Hanping Feng) reproduces socioeconomic marginalization, limiting participants' ability to accumulate "digital capital" (Bourdieu)—the skills and resources needed for full digital participation.

Second, algorithmic power shaped content access through platform biases. Chinese platforms like WeChat and Douyin prioritize Mandarin content and national narratives, creating "algorithmic enclosure", a filtered reality that marginalizes transnational identities. RZL (27) expressed frustration: "Even when I follow Vietnamese accounts, the platform shows me mostly Chinese content. It's like it wants me to forget where I came from."

Third, patriarchal power regulated media use within households, operating through subtle social pressure rather than overt censorship. Participants engaged in "anticipatory compliance"—modifying behavior to avoid conflict. ZQQ (48) explained: "I only call my family when my husband is at work and my mother-in-law is napping. No questions, no suspicious looks." This internalization of disciplinary gazes (Gillian Rose) reflects the "patriarchal bargain" (feminist scholars), a dynamic where women negotiate within gendered hierarchies and sometimes reinforce them through adaptive strategies.

Together, these power structures create "double power geometries", a layered system where global digital infrastructures, national platform policies, and intimate family dynamics intersect to shape contradictory media experiences. Participants' "calibrated navigation" of this terrain—balancing strategic assimilation, affective preservation, and resistant appropriation—reflects their agency within constraint.

Conclusion and Discussion

This study examines how Vietnamese brides in China navigate cross-cultural adaptation through media, revealing a complex interplay of agency and structural constraint. The three identified patterns of media engagement, namely strategic integration, affective preservation, and resistant appropriation, illustrate that digital tools are neither inherently empowering nor constraining; instead, they serve as contested terrains for negotiating belonging.

Theoretically, this research contributes to two key debates. First, it extends scholarship on migration and media by developing the twin concepts of strategic and affective engagement, moving beyond instrumentalist frameworks to capture the emotional and political dimensions of digital practice. By framing media use within power geometries of citizenship, culture, and patriarchy, we highlight how marginalized groups navigate contradictory demands in constrained environments. Second, it advances de-Westernization efforts by centering Sino-Vietnamese marriage migration, offering a non-Western case that challenges assumptions about digital agency rooted in liberal democratic contexts. Unlike Western migrants, who may use media to claim citizenship rights, Vietnamese brides deploy media as a "safe" form of resistance—low-risk, clandestine, and adaptive to state and familial control.

Practically, the findings offer insights for policymakers and support organizations. Digital literacy programs should acknowledge both strategic and affective needs, teaching technical skills while validating cultural continuity. Platform designers could incorporate multilingual features and reduce algorithmic bias to accommodate transnational lives. Community initiatives—such as peer mentorship networks led by long-resident brides—could foster safe spaces for sharing media strategies and addressing discrimination.

Limitations of this study include its focus on one county, limiting generalizability, and its emphasis on brides' experiences (excluding husbands' and in-laws' perspectives). Future research could explore regional variations, track media practices over time, and examine how commercial marriage intermediaries shape digital expectations. Longitudinal studies would also illuminate how media use evolves with residency duration and generational shifts.

The metaphor of "falling leaves, unrooted lives" captures the duality of these women's experiences: media allows them to "put down roots" in China through strategic engagement while remaining "tethered" to Vietnam via affective practice. Their precise, adept, invisible and enduring navigation demonstrates their capacity for autonomous action amid global inequality, national regulation and intimate geopolitics.

In the end, their media practices reveal a profound truth: belonging is not a fixed state, but an ongoing negotiation, one click, one video, one secret chat at a time.

References

1. Bhabha, H.K. (1994). The Location of Culture (2nd ed.). Routledge. https://doi.org/10.4324/9780203820551
2. Charmaz, K. C. . (2006). Constructing grounded theory: a practical guide through qualitative analysis. International Journal of Qualitative Studies on Health and Well-Being,1(3). https://doi.org/10.3402/qhw.v1i3.4932
3. Chiu, T. Y., & Yeoh, B. S. A. (2021). Marriage migration, family and citizenship in Asia. Citizenship Studies, 25(7), 879–897. https://doi.org/10.1080/13621025.2021.1968680
4. Chiu, T.Y., & Yeoh, B.S.A. (Eds.). (2023). Marriage Migration, Family and Citizenship in Asia (1st ed.). Routledge. https://doi.org/10.4324/9781003391869
5. Constable, N. (Ed.). (2005). Cross-Border Marriages: Gender and Mobility in Transnational Asia. University of Pennsylvania Press. http://www.jstor.org/stable/j.ctt3fhv66
6. Dyrness, Andrea, and Enrique Sepúlveda. Border Thinking: Latinx Youth Decolonizing Citizenship. University of Minnesota Press, 2020. JSTOR, https://doi.org/10.5749/j.ctvz0h9tq
7. Fortier, A. (2017). The psychic life of policy: Desire, anxiety and 'citizenisation' in Britain. Critical Social Policy, 37, 21 - 3. http://dx.doi.org/10.1177/0261018316655934
8. Giddens, A. (1984) The Constitution of Society. Outline of the Theory of Structuration. University of California Press, Berkeley.
9. Harpaz, Y., & Mateos, P. (2019). Strategic citizenship: negotiating membership in the age of dual nationality. Journal of Ethnic and Migration Studies, 45(6), 843–857. https://doi.org/10.1080/1369183X.2018.1440482
10. Hochschild, A. R. (1983). The managed heart: Commercialization of human feeling. University of California Press.
11. Iwabuchi, K. (Ed.). (2004). Feeling Asian Modernities: Transnational Consumption of Japanese TV Dramas. Hong Kong University Press. http://www.jstor.org/stable/j.ctt2jc5b9
12. Kim, G. and Kilkey, M. (2018), Marriage Migration Policy in South Korea: Social Investment beyond the Nation State. Int Migr, 56: 23-38. https://doi.org/10.1111/imig.12350
13. Liao, S. (2019). Wang Hong Fashion Culture and the Postfeminist Time in China. Fashion Theory, 25(5), 663–685. https://doi.org/10.1080/1362704X.2019.1638158
14. Massey, D. . (1994). Power-geometry and a progressive sense of place.
15. Nguyen, T. H. et al. (2023). Climate Impacts on Hydropower in Vietnam's Red River Basin. Energy for Sustainable Development, 75, 1-14. https://doi.org/10.48550/arXiv.2301.10343
16. Peng, Y. Migrant Mothering in Transition: A Qualitative Study of the Maternal Narratives and Practices of Two Generations of Rural-Urban Migrant Mothers in Southern China. Sex Roles 79, 16–35 (2018). https://doi.org/10.1007/s11199-017-0855-7
17. Scott, J. C. (1990). Domination and the Arts of Resistance: Hidden Transcripts. Yale University Press. http://www.jstor.org/stable/j.ctt1np6zz
18. Schild, Veronica. (2015). Emancipation as Moral Regulation: Latin American Feminisms and Neoliberalism. Hypatia. 30. https://doi.org/10.1111/hypa.12162
19. Shen, H. H. (2008). The Purchase of Transnational Intimacy: Women's Bodies, Transnational Masculine Privileges in Chinese Economic Zones. Asian Studies Review, 32(1), 57–75. https://doi.org/10.1080/10357820701870759
20. van Dijck, Jose, The Culture of Connectivity: A Critical History of Social Media (New York, 2013; online edn, Oxford Academic, 24 Jan. 2013), https://doi.org/10.1093/acprof:oso/9780199970773.001.0001
21. Wang, H. (2001) Academic mentorship: An effective professional development strategy for medical reference librarians. Medical Reference Services Quarterly, 20, 23-31. https://doi.org/10.1300/J115v20n02_03

JOURNAL OF GLOBAL TRENDS IN SOCIAL SCIENCE

J G T S S

Research article

https://doi.org/10.70731/c2k8kj23

China's Open Government Data: A Legal Perspective

Layton Ren [1,*], Fei-hung Yu [2]

Received 4 February 2026

Accepted 2 March 2026

Published 31 March 2026

ISSN 2759-7830 (Online)
ISSN 2760-2508 (Print)

Abstract: This article focuses on an innovative system in China's digital government construction from a legal perspective, called open government data. This system can improve government administrative efficiency and promote the use of data to expand the data factor market. However, in practice, the indiscriminate use of the concepts of open government data (OGD) and government information disclosure (GID), as well as legal problems such as the direct transplantation of existing systems, seriously hinder the effective advancement of OGD. To explore current status of OGD practices, this article selects 4,629 judicial documents of litigation related to GID and 4 types of indexes about OGD platforms as empirical analysis samples, aiming to clarify meanings and relationships between two concepts at the legal level. A deep analysis of these samples reveals litigation dilemma, platform development dilemma, and legal dilemma in the process of OGD. In response, this article proposes some measures including regulating OGD through separate law, replacing the three-tier system with the three-tier & dual system, and establishing the collaborative rights protection system. In this way, the ultimately goal that promoting the efficient and orderly advancement of OGD will be achieved.

Keywords: Open Government Data (OGD); Government Information Disclosure (GID); Concept Clarification; Regulation Improvement; Rights Protection

Introduction

In the epoch-making transformation of digital governance model, government information disclosure (GID) has shifted to open government data (OGD), which is a new requirement for the construction of digital government [1]. In this trend, countries continue to explore and improve the institutional basis, operational logic and guarantee system of OGD. In 2015, the State Council issued the *Outline of National Action for Facilitating Big Data Industry Development* under the goal of strengthening the data power, and the *Opinions on Building a Data Infrastructure System to Better Play the Role of Data Elements* (hereinafter referred to as the *Twenty Data Measures*) issued by the Central Committee of the Communist Party of China and the State Council at the end of 2022, both emphasize the promotion of effective opening up and flow of "public data," including "government data." In 2016, the State Council issued the *Outline of the 13th Five-Year Plan for National Economic and Social Development of the People's Republic of China*, also directly focus on the term "government data," and explicitly require "the formulation of government data sharing and opening up catalogs, and push forward the opening up of data resources to the community in accordance with the law." Such significant policy documents provide directional guidance about the implementation of OGD for government departments around China. In the process of implementing OGD in full swing, Shanghai launched the first OGD platform (named the "Shanghai Government Data Portal" [2]) and issued the *Interim Procedures of Shanghai City on the Opening of Public Data* to standardize the order of public data openness, all of which provide valuable experience for OGD. If OGD operates efficiently, it can not only enhance the construction of service-oriented, digitalized, and efficient government, but also influence other subjects, such as increasing the total factor productivity of enterprises [3], the efficiency of citizens' access to administrative information, and so forth.

However, due to the lack of specific legislation on OGD in China, the specific operational rules have not yet become a unified standard. Simultaneously, in the pilot attempts of governments around China, the platforms are scattered and clearly separated. It can be also seen that all of them have different understandings of OGD, and that the concept is sometimes confused with GID. Therefore, both the regulations governing OGD, and the attempts to promote it, have significant shortcomings. How to clarify the distinction between OGD and GID, how to structure a system for implementing and

[1] Shandong University, Qingdao 266237, China; [2] South China Normal University, Guangzhou 510631, China.
* Corresponding author. Email: 202410436@mail.sdu.edu.cn

guaranteeing OGD are crucial issues to the construction of a digital government and the flow of data elements, and should be discussed again.

Inheritance But Not Replace: The Relationship Between OGD and GID

For a long time, although there is no clear comparison and clarification of the relationship between OGD and GID in practice and legislation, there has been a general debate on this issue in academia, attempting to emulate the relatively mature system behind GID and apply it to OGD. The first step in how to operate is to clarify the two major concepts, thereby understanding their connection and links.

A review of previous academic arguments on the two concepts reveals three main theories that establish a logical connection between them, **1)** inheritance theory, which states that GID lays the legal foundation for OGD, and in the big data era, it has become a topic of data openness [4]; **2)** subordination theory, which states that data openness, in terms of concepts, laws, values, or management, is all part of information disclosure [5]; **3)** mutual exclusivity theory, which states that the two concepts are mutually exclusive because their institutional foundations and goals are fundamentally different. However, all three types of theories have shortcomings, to varying degrees, in their incomplete explanation of the relationship between the two concepts. For example, the inheritance theory merely focus on the connection between the two concepts and mentions fewer of the differences. The subordination theory is even more so. On the contrary, the mutual exclusivity theory makes the two concepts exclusive and does not seek common ground from differences. As mentioned above, the three theories are impossible to clarify relationship between OGD and GID, and they are all not reasonable enough to become the theoretical foundation of the institutional framework. Therefore, it is necessary to explore the real relationship between the two concepts from their extension and intension.

From the perspective of extension, data is the result of digitization of information, which is also applicable to the administrative field for conceptual explanation. Extraterritorial law has made a definition of government data, and in the *Open Government Data Act of the United States*, data is defined as "recorded information." Evidently, in the context of the data-based administration, data and information do not have a strict delimitation of their extensions. In the practice of various countries, administration digitization is interpreted as the evolution from GID to OGD, with the former being the basis for the adaptation of the latter, and the latter being the future evolution of the former. For example, the United States has gradually promoted OGD within the framework of GID, and in the Memorandum of the *Transparency and Open Government*, the authorities requested the coordination among relevant departments to develop the *Open Government Directive* within the framework of the *Freedom of Information Act*. China's *2015 Key Points for Government Information Disclosure Work* specifies that OGD is one of the paths to further expanding GID. Therefore, as the inheritance theory suggests, OGD is the expanded and upgraded version of GID in the big data era and there is a logic of inheritance between the two concepts. Further, although the conceptual contents of government data and government information may have minor differences, the administrative acts carried out through openness or disclosure means are all a kind of transparency in government affairs, which means they have formal correspondence and content relevance. Therefore, the two concepts are closely related and can be applied through institutional transfer.

From the perspective of intension, OGD and GID share the same macroscopic objectives of enhancing government transparency, fostering the construction of digital government and utilizing data and information resources to promote the development of the national economy. However, there is a core difference in specific application. In GID, besides the option of proactive disclosure, the government also bears the obligation of disclosure under the application-based disclosure system. In this legal context, the government is positioned as the supervisor and unilateral obligor in order to protect citizens' right to know, but the two sides are likely to have an adversarial relationship and antagonistic sentiments [6]. In OGD, the government is in a proactive position, with no passive application for openness. Meanwhile, OGD is not merely stop at the level of knowing, but further moves towards integration and utilization. For example, opening up data related to people's livelihoods, such as transportation and weather, allows for collaborative discussions with businesses and social organizations to solve problems in traffic management and agricultural governance. The relationship between the government and the public at this example is non-confrontational and mutually beneficial. It can be noted that GID tends to guarantee the people's right to know, while OGD tends to promote the combination and utilization of data, and to move towards the effective flow of data. So the above two concepts are not mutually replaceable.

In a nutshell, "inheritance but not replace" best summarizes the relationship between OGD and GID. Furthermore, GID can serve as a template and foundation for OGD, but the theoretical improvement of its shortcomings needs to focus on further practical operations.

Data Analysis on Judicial Documents and Indexes

To explore the ideal institutional framework for OGD, this article leverages data analysis as its research method. On the one hand, it focuses on judicial practice and summarizes the insights and experiences from litigation related to GID, while also revealing the necessity and urgency of improving OGD. On the other hand, it conducts an in-depth analysis of the current status of local OGD platforms, summarizing the shortcomings of existing platforms from three dimensions about time, region, and effectiveness, so as to provide feasible ideas for good laws and sound governance in the field of OGD in the future.

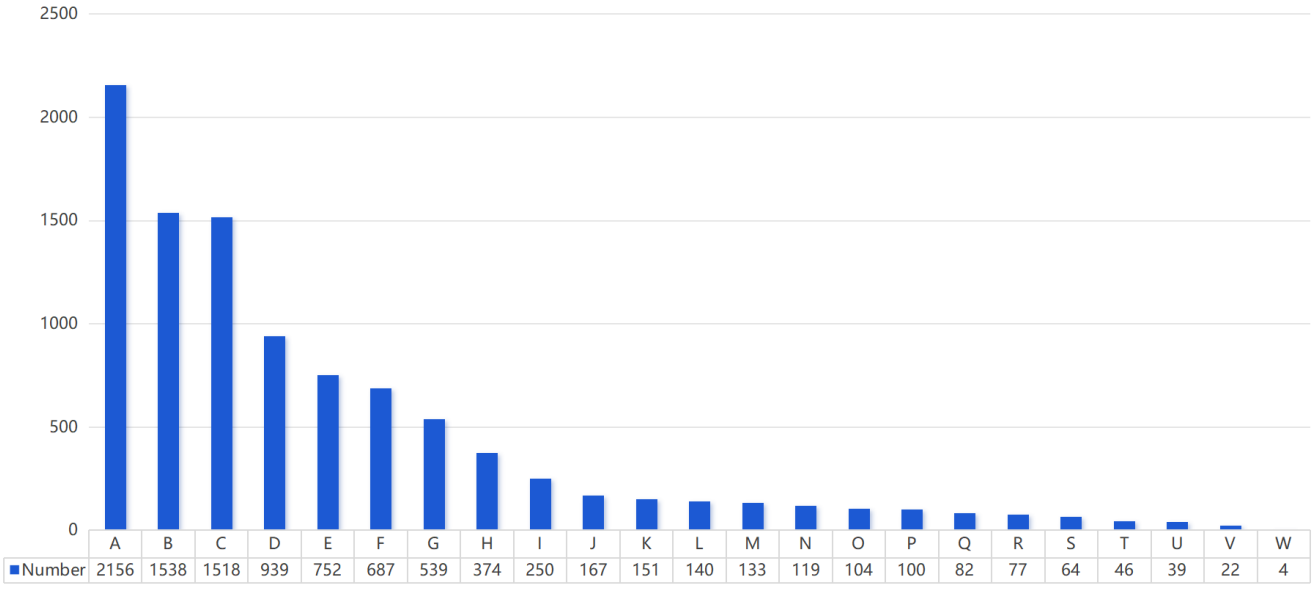

Figure 1 | Statistics on the types of information requested for disclosure in litigation related to GID

Letters from A to W represent different types of information that citizens are requesting to be disclosed in the litigation. Specifically, A: Planning and related policies; B: Land and housing expropriation and requisition; C: Housing demolition; D: Government finances; E: Archives; F: Major construction projects; G: Education; H: Social security; I: Economic and social development statistics; J: Environmental protection; K: Food, medicine, and products; L: Land bidding; M: Employment; N: Government procurement; O: Poverty alleviation; P: Work safety; Q: Public health; R: Public emergencies; S: Administrative agency office information; T: Civil service examinations; U: Real estate transactions; V: Medical care; W: Administrative and institutional fees.

Data analysis on judicial documents of litigation related to GID

This article selects judicial documents of litigation related to GID that occurred between January 1, 2012 and September 17, 2021. A total of 4,629 judicial documents were identified, including 4,045 judgments and 584 orders. The following section presents a statistical analysis of the sample from three important dimensions that reflect the substantive characteristics of litigation, including the requests for disclosure, the outcomes of litigation, and the reasons in litigation.

The requests for disclosure

Among the 4,629 judicial documents, the requests for disclosure can be categorized into 23 types (Figure 1). Data shows that the top three categories—planning and related policies, land and housing expropriation and requisition, and housing demolition—are centered around real estate and constitute the majority of the requests for disclosure. In practice, these requests often coexist. For example, in a single case, an applicant may simultaneously request disclosure of regional planning schemes, land and housing expropriation and requisition policies, and housing demolition policies.

The outcomes of litigation

Among the 4,045 judgments, in the procedure of first instance, only slightly more than half of the judgments ruled in favor of the plaintiffs in all or part of their claims (Figure 2). This indicates that while litigation related to GID, as the final channel for rights remedies, plays a role in protecting the right of natural persons, legal persons, and other organizations to know government information, its role is clearly insufficient. In the procedure of second instance, the ratio of judgments with reversed judgments to judgments with dismissed appeals is approximately 1:10 (Figure 3), indicating that the vast majority of first instance judgments in litigation related to GID are final and effective, with a low rate of reversed judgments in the procedure of second instance. Furthermore, there are only 3 judgments in the retrial of litigation related to GID, and the applicant for retrial is the same one. The result in all such judgments is that the respondent (namely government) is required to make a new response within prescription, which constitutes the relatively insignificant example.

Among the 584 orders, regardless of procedure, orders in litigation related to GID mostly ended with the rejection of the plaintiff's, appellant's, and applicant's requests (Figure 4, 5 & 6). To some extent, this reflects the general tendency in China's judicial practice to "emphasize substance over procedure," and further reveals that administrative bodies have a certain degree of subjective resistance and negative attitude in their work on information disclosure.

The reasons in litigation

Among the 4,629 judgments and orders, the reasons for agreeing to information disclosure were relatively simple. However, a significant number of judicial documents (in total 2,867) rejected citizens' requests or refused to file cases, which were detrimental to the administrative counterparts. In these cases, the courts' reasons for refusal were diverse, falling into eight categories (Table 1), reflecting the complexity of information disclosure review.

https://doi.org/10.70731/c2k8kj23

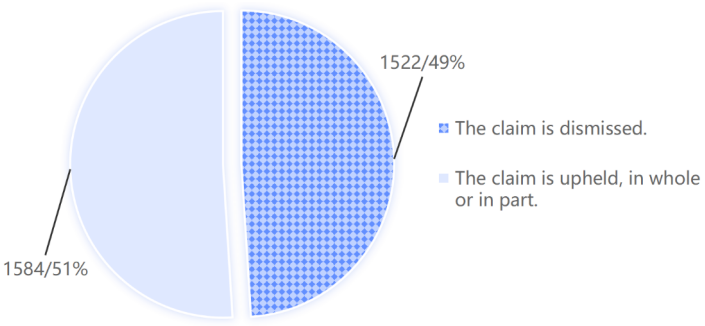

Figure 2 | The composition of the judgments in the procedure of first instance

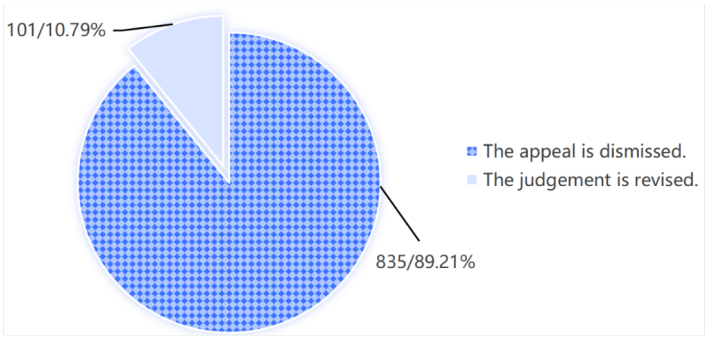

Figure 3 | The composition of the judgments in the procedure of second instance

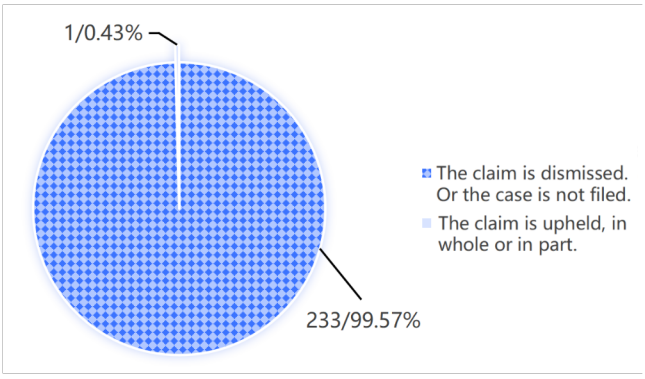

Figure 4 | The composition of the orders in the procedure of first instance

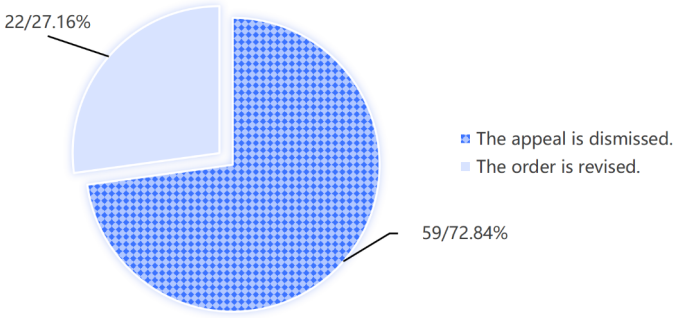

Figure 5 | The composition of the orders in the procedure of second instance

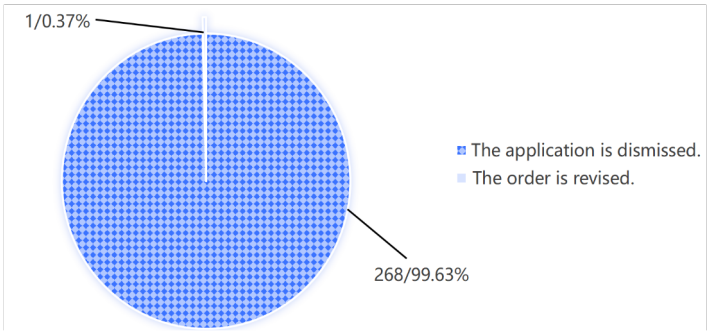

Figure 6 | The composition of the orders in the retrial

Table 1 | Reasons for refusing to disclose information

Reasons	Number	Legal Basis
Making it public could endanger national security, public safety, economic security, and social stability.	70	Article 14 of Open Government Information Regulation of the People's Republic of China
The information involves personal privacy and trade secret. Disclosure of it would harm the legitimate rights and interests of third parties.	871	Article 15 of Open Government Information Regulation of the People's Republic of China
Information that pertains to the internal affairs of administrative agencies may not be made public.	21	Article 16 of Open Government Information Regulation of the People's Republic of China
Information that pertains to process-related information and case files of administrative law enforcement may not be made public.	174	Article 16 of Open Government Information Regulation of the People's Republic of China
The government information in question does not exist.	1364	Article 36.1.4 of Open Government Information Regulation of the People's Republic of China
The administrative agency has no authority to disclose this information.	76	Article 36.1.5 of Open Government Information Regulation of the People's Republic of China
The applicant repeatedly requested for access to the same government information.	87	Article 36.1.6 of Open Government Information Regulation of the People's Republic of China
The information falls under categories such as business registration, real estate registration data, and so forth. The relevant laws and administrative regulations have specific rules regarding the acquisition of such information.	204	Article 36.1.7 of Open Government Information Regulation of the People's Republic of China

Data Source: China Judgments Online

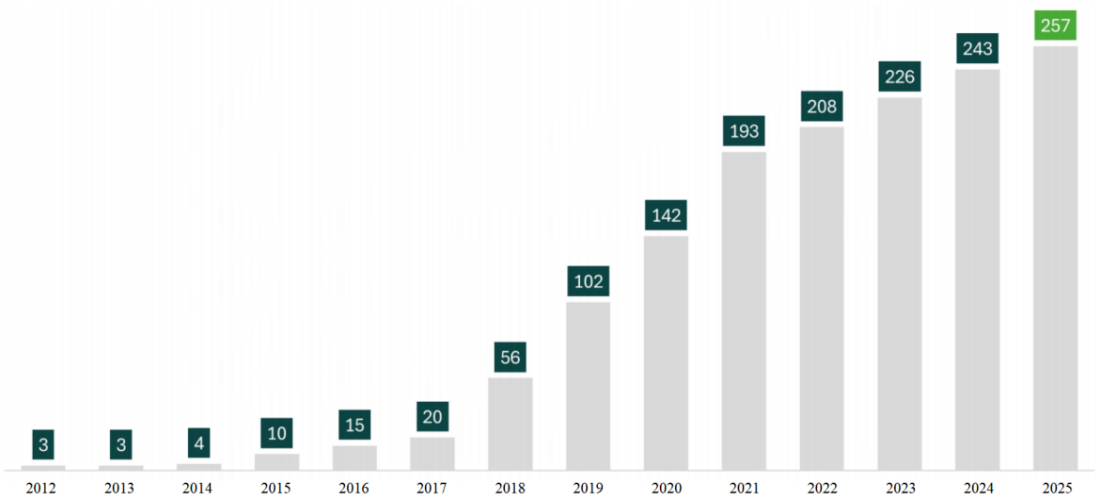

Figure 7 | Growth in the number of OGD platforms at the prefecture-level and above over the years [7]

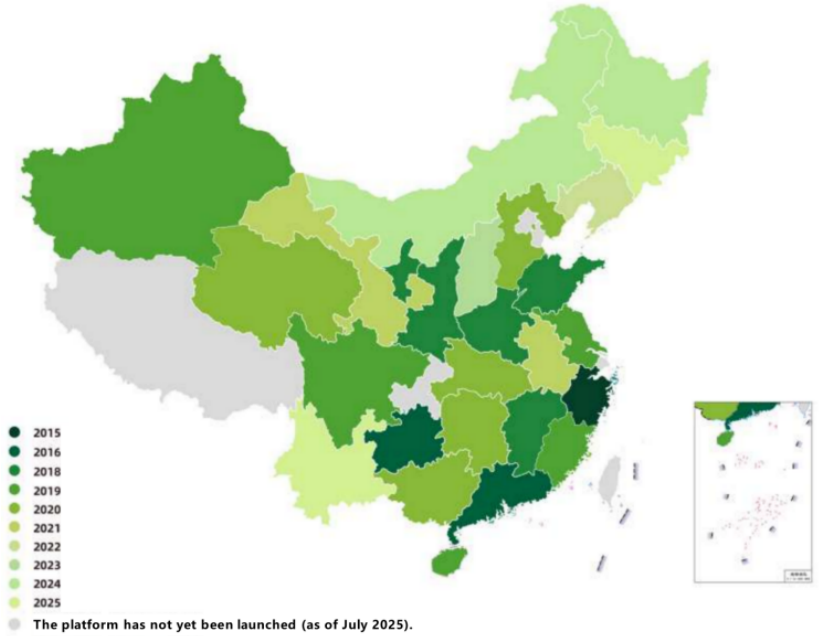

Figure 8 | Geographical distribution of the launch time of provincial-level OGD platforms [7]

Data analysis on indexes about OGD platforms

In addition to analyzing judicial documents, a comprehensive understanding of the current development of China's OGD is necessary to gather sufficient practical evidence for its future inclusion in more laws and regulations. To this end, this section collects 4 types of indexes about OGD platforms, presenting them from three dimensions about time, space, and result, and then drawing some preliminary conclusions.

Time-based analysis: growth

According to the Figure 7, the total number of OGD platforms at the prefecture-level and above in China has been increasing year by year, and entered a period of rapid construction in 2017, which slowed down after 2022. Among them, the platform construction in 2021 was particularly remarkable, with 51 new OGD platforms added compared to 2020, reaching the highest increase in history.

Space-based analysis: distribution

According to Figure 8, the development of China's OGD platform construction project generally shows a trend of radiating from the southeast coast to the northwest. The underlying reason is that the more developed a province's economy and the more frequent its external economic exchanges, the earlier its OGD platform is launched. However, after more than a decade of development, some provinces still haven't launched their OGD platforms, and some platforms in other provinces have reached their limits and become inaccessible.

According to Figure 9, among provinces in China, the number of prefecture-level OGD platforms is relatively small and their distribution is scattered. Taking Henan Province as

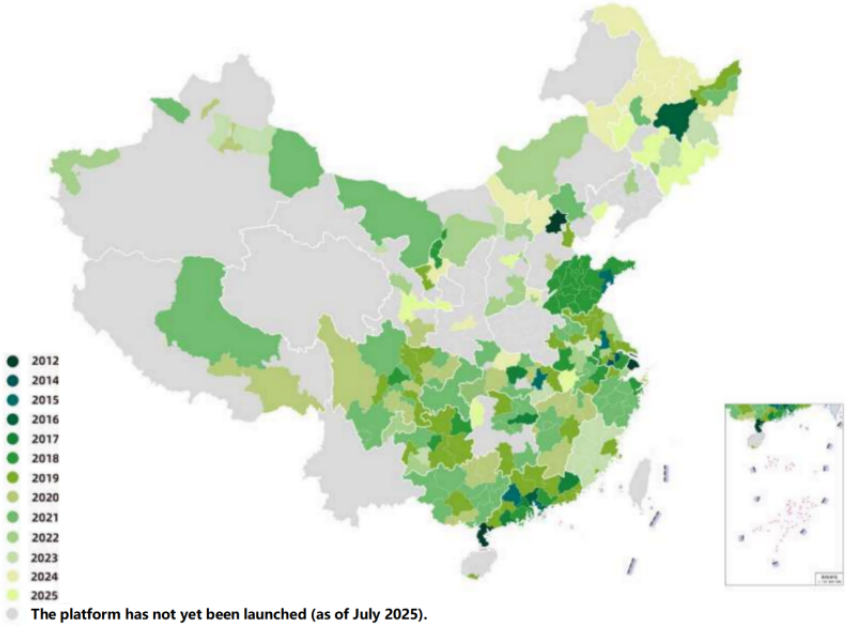

Figure 9 | Geographical distribution of the launch time of prefecture-level OGD platforms [7]

an example, although provincial-level OGD platforms started relatively early in Henan Province, only a few prefecture-level cities in northern place of Henan Province have established prefecture-level OGD platforms, while most other prefecture-level cities have not yet launched their own OGD platforms.

Results-based analysis: how effective was it?

The China Open Data Index is a statistical index used to assess the development level of OGD platforms. It evaluates platforms from four aspects about preparedness, platform layer, data layer, and utilization layer. It comprehensively and deeply reflects the objective status of OGD platforms in China. The higher the index, the better the effectiveness. According to **Figure 10**, OGD platforms have achieved effective development nationwide, but the effectiveness varies, and many provinces still have platforms with relatively low effectiveness.

The Existing Dilemmas of OGD

Litigation dilemma: Chaotic handling caused by system confusion

One purpose of OGD is to stimulate the huge social and economic value of government information resources, enhance the national economic growth [8], and allow the source of social wealth to fully flow through the carrier of data. The legislative purpose of the *Regulations on the Disclosure of Government Information* (Article 1) is that "in order to ensure that the natural persons, legal persons and other organizations have access to government information according to the law, to improve the transparency of the work of the government and build a government under the rule of law, and to give full play to the role of government information in serving the people's production, livelihood and economic and social activities." It is obvious that the common purpose of OGD and GID intersect in the economic field that to facilitate economic development and the growth of national wealth. However, this article finds, after empirical analysis on judicial documents, that the overlap of the common purpose can lead to chaotic handing in judicial practice. More notably, such overlap can actually lead to more negative impacts between two similar systems in litigation related to GID.

The negative impact of litigation related to GID on OGD

Firstly, although the purpose of OGD and GID are intertwined in the economic field, which has led to a surge in cases involving requests for access to planning and related policies, which are most closely related to the economy (**Figure 1**), the significant differences between other purposes of OGD and GID in other fields inevitably lead to litigation related to GID being an inefficient, imperfect, but also the only judicial remedy alternative to OGD.

Secondly, the chaotic handling has led to a certain degree of evasion and laxity by some administrative bodies in litigation related to GID, which will likewise affect the judicial fairness of cases related to OGD (**Figure 2**,**3**,**4**,**5**,**6** & **Table 1**). As a result of remedy approach of OGD in China is not yet clear, litigation related to GID is compelled to be used in some of cases related to OGD, and the administrative intervention of judicial decision-making in practice is difficult to completely eliminate in the short term. Under this impact can be predicted that part of the reasonable claims related to OGD, also failed to be supported by the court under the shell of litigation related to GID.

Thirdly, the rights and interests of private subjects will also face challenges in OGD. The cases in which the judgment of litigation related to GID rejected the claims, those involved in personal privacy, trade secret, and the legitimate rights and

https://doi.org/10.70731/c2k8kj23

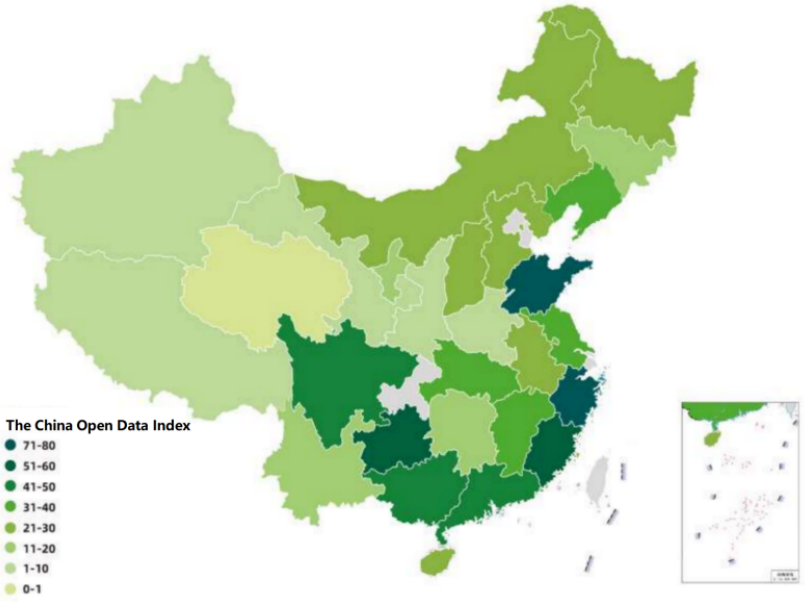

Figure 10 | The China Open Data Index [7]

interests of third parties occupied a considerable proportion (Table 1). It indicate that under the framework of GID, the civil rights and interests of the private subjects may be infringed upon in countless cases, which poses a severe challenge to the judicial defense and leads to serious challenges to the protection of the private rights of civil and commercial entities. Since both government information in GID and government data in OGD possess the same characteristics of general data, namely ease of replication and mobility, thus there is no doubt that the protection of private rights will also be a key area of focus and challenge for laws and regulations in the future field of OGD.

The negative impact of OGD on litigation related to GID

Similarly, due to the chaotic handing of the two systems, and the lack of remedy approach for OGD, the case burden that should be shared by remedy approach of OGD has instead been placed on litigation related to GID. Such phenomenon has two direct negative impacts, **1)** it will lead to the waste of unnecessary judicial resources, affecting the effectiveness of such resources; **2)** a surge in the number of cases reduces the average time and effort that judicial personnel spend on each case, leading to a decline in the quality of litigation and thus increasing the probability of mistrial.

Platform Development Dilemma: Uneven Regional Development and Insufficient Progress

Pilot projects for OGD platforms have been launched in provinces and cities across China. However, as the empirical analysis on indexes about OGD platforms shows, the development of these platforms still faces some dilemmas, hindering the further implementation of OGD.

Uneven regional development

The development of China's OGD platforms generally shows a trend of radiating from the southeast coast to the northwest inland, with progress slowing down the further away from the coastal areas (Figure 8 & 9). This generally aligns with the geographical distribution of economic development levels across China. The more developed the economy, the faster the construction of supporting OGD platform infrastructure. The two complement and promote each other, a point particularly evident in the Yangtze River Delta Urban Cluster and the Pearl River Delta Urban Cluster. This is particularly evident in the Yangtze River Delta and Pearl River Delta urban clusters. For less developed regions, effectively assisting in the construction of OGD platforms, promoting investment attraction, and achieving a positive interaction between OGD platforms and economic development presents a challenge that legislative and administrative bodies need to consider in the future.

Insufficient progress

Although the number of OGD platforms is increasing year by year, their effectiveness varies greatly, some lack maintenance, and most are at a low level (Figure 7 & 10). Improving the quality of these platforms remains a crucial issue for future development. Simultaneously, while most provinces already have OGD platforms (Figure 8), development within these provinces is noticeably hollowed out (Figure 9 & 10), exhibiting a trend of "much ado about nothing." Therefore, how to promote the establishment of prefecture-level OGD platforms within provinces and accelerate the linkage between prefecture-level and provincial-level platforms urgently needs to be addressed.

Table 2 | Status of Local Norms on OGD

Region	Number of Abstract Administrative Acts	Number of Administrative Legislation
Guangdong	62	5
Jiangsu	39	0
Shanghai	10	0
Guizhou	48	8
Sichuan	18	0
Hubei	4	0
Anhui	15	0
Xinjiang	8	0
Hebei	18	18
Heilongjiang	24	0
Fujian	22	1

Data Source: Wolters Kluwer China Law & Reference

Deeper Essence: The Dilemmas Caused by the Lack of Law

The litigation dilemma and platform development dilemma revealed by the analysis of litigation data and platform data have a deeper, undeniable cause: the lack of clear laws. This vagueness results in a lack of explicit rules to resolve disputes over OGD, and also a lack of clear rules to provide clear guidance for promoting OGD.

Difficulties in legislation: The absence of top-level design and local administrative legislation

Firstly, the absence of top-level design on OGD. At present, the only existing laws and administrative regulations related to OGD in China are the *Cybersecurity Law*, the *Data Security Law*, the *Personal Information Protection Law*, and the *Regulations on the Security Protection of Key Information Infrastructures*. However, they are not specific enough to regulate OGD, and fail to systematically and comprehensively construct provisions according to the characteristics and purposes of OGD, which is insufficient to effectively regulate and guide the development of it. What kind of legislative techniques and concepts should be used to formulate specialized departmental regulations, administrative regulations and even laws for OGD in the future is a matter of urgency.

Secondly, the absence of local administrative legislation on OGD. This article has conducted a statistical survey on the number of abstract administrative acts and administrative legislation on OGD in some provinces and autonomous regions from the northeast, central, east and west regions in China. According to Table 2, due to the lack of specialized top-level design mentioned above, local administrative legislation is problematic, which is manifested in the fact that the vast majority of provinces have only stayed at the level of issuing other normative documents regarding the regulation of OGD. Such abstract administrative acts, which only scratch the surface, not only have limited scientific rigor and are unlikely to make a constructive contribution to improving the legal system, but also cause serious conflicts between local legislative ideas, resulting in a situation where "everyone has their own reasons."

Normative vagueness: The confusion regarding the intension and extension of OGD

Through research, this article finds that few laws, regulations, rules, and other normative documents provide a clear textual interpretation of the concepts of OGD and government data. For example, the concept "government data" is defined only in the *Regulations of Guiyang City on the Sharing and Opening of Government Data*, the *Implementation Measures of Guiyang City on the Sharing and Opening of Government Data*, the *Administrative Measures of Panjin City on Co-construction and Sharing of Government Data Resource (Trial)*. Similarly, the concept OGD is only clearly explained in the *Regulations of Guizhou Province on Open Government Data* and the *Implementation Opinions of People's Government of Guiyang City on Accelerating the Sharing and Opening of Government Data in Guiyang*. Regarding OGD, different regions offer conflicting interpretations, lacking a systematic and authoritative definition with high legal force. There is an urgent need to unify and establish important concepts such as intension, extension, subject, scope, method, and so forth of OGD nationwide.

The urgency of remedies: The practical dilemma of administrative remedies

In current practice in China, for administrative disputes arising from OGD, if natural persons, legal persons, and other organizations do not protect their rights though litigation related to GID, their main recourse is through internal orders from higher and lower levels within administrative bodies. However, such orders are highly uncertain and lack effective supervision procedure. Because there are no laws or regulations explicitly stipulating whether administrative bodies should bear responsibility for OGD, or what kind of responsibility they should bear. The timing and type of orders issued to lower-level administrative bodies are usually entirely determined by higher-level administrative bodies. Such administrative remedy faces the predicament of being ineffective in practice, lacking guidance and guarantees for operation within the framework of the rule of law.

China's Responses to the Dilemmas of OGD

The dilemmas of OGD exists in practice and also profoundly reveals the shortcomings in the legal system. Such shortcomings can be summarized by three parts including chaotic handling, unreasonable data open system, and the lack of remedies. To address these shortcomings, the governance paradigm for OGD can be improved by formulating specific regulations, reconstructing data open system, and exploring remedies.

Regulating OGD through separate law

As of 2021, local governments in more than 50 provinces and municipalities in China have issued nearly 80 regulations related to OGD, forming three major normative systems, **1)** regarding OGD as an extension of GID; **2)** formulating regulations of OGD in accordance with GID; and **3)** formulating independent regulations of OGD [9]. However, such regulations have the weaknesses of conceptual vagueness, chaotic handling, low level of efficiency, insufficient legally binding force, and so forth. In this regard, the legal logic and legal requirements of OGD should be reviewed in order to formulate specific regulations for it.

Given the potential drawbacks of including OGD in GID, such as increasing the accountability orientation, hindering openness, creating disorder, and the framework of GID failing to meet the requirements of OGD [9], the framework of the *Regulations on the Disclosure of Government Information* should not be directly adapted as the normative framework for OGD. Instead, separate law should be enacted for OGD. To establish high-level and more unified guidelines, it may be possible to summarize the regulatory experiences of Qingdao, Weihai, Jinan, Harbin and other places regarding the OGD (Like the *Notice from the General Office of the People's Government of Qingdao City on Accelerating the Opening of Public Information Resources to the Society*, the *Implementation Opinions of the People's Government Office of Weihai City on Promoting the Opening and Utilization of Government Data Resources to the Society*, the *Notice of Jinan City on Promoting the Opening of Public Information Resources to the Society*, the *Implementation Plan of Harbin City on Promoting the Opening of Government Data to the Society*, and so forth), and for the State Council to promulgate the "Regulations on the Open Government Data".

Regarding the specific legislative framework, the first focus is on general requirements. This article suggests that the intension, extension, subject, scope, method, and other important concepts of OGD should be clearly defined. For example, in terms of the scope of OGD, the "list-based" management model can be adopted, requiring managers to regularly publish the types of open data to ensure predictability and up-to-dateness. As for the content of the types, opening up non-legally prohibited data and conditionally open data are generally included in the content. Simultaneously, as stipulated in the *Administrative Measures of Shanghai City on the Public Data and One-Netcom Office*, high-value data related to people's livelihoods and urgently needed by society should be given priority for opening up. Applying this as a guideline in OGD, especially in the Authorize-Operate model (AO model), has its value. For another example, in terms of the method of OGD, in addition to the direct opening model initiated by administrative bodies, the rules for AO model should also be a key focus in separate law. Furthermore, regarding basic principles, it is not advisable to continue adopting the principle that "public disclosure is the rule rather than the exception" in GID. The core purposes of OGD are to integrate and innovate data elements, achieve efficient flow of data resources, and effectively leverage the value of data. Therefore, OGD should take social needs as a key consideration, and the basic principle should be described as "oriented towards social needs".

The second focus is on execution requirements. Currently, OGD platforms across China generally suffer from low data quality, limited capacity, poor timeliness, and insufficient public interaction. Therefore, the proposed "Regulations on the Open Government Data" should stipulate the construction and operation rules for a unified national OGD platform in detail, establishing unified management. Simultaneously, it should also stipulate the appropriate retention of certain content from existing local regulations of OGD, thereby efficiently building a new system on the existing foundation.

Undoubtedly, the more specific content in various chapters of the "Regulations on the Open Government Data" can draw on the regulatory experience left by foreign laws and existing domestic laws, and then be discussed and considered in detail. Furthermore, the focus should be on the ever-changing nature of administrative law, with particular attention paid to the interpretation of such regulation.

Replacing the three-tier system with the three-tier & dual system

The data open system in various provinces of China generally form a three-tier system of "unconditional openness—limited openness—prohibited openness", as can be seen from the *Interim Procedures of Zhejiang Province on the Opening and Security Management of Public Data*, and so forth. Its criteria are based on risk standard, with the underlying principle being "openness with no impact—openness with some impact—openness with significant impact" [10]. However, this system will undoubtedly fall back into the old trap of GID. Risk is a vague concept, and using it as the standard to define the types of open data will undoubtedly turn the vague concept into the vague application.

In response, this article suggests that the three-tier system can be turned into a more specific institutional arrangements to better align with the requirements of OGD. Therefore, adopting the three-tier & dual system of "fully openness—limited openness—AO model" as the improved system is a reasonable approach.

The meaning of "three-tier" in improved system

Specifically, what does "three-tier" mean in improved system? Under the system of "full openness—limited openness—AO model," the three tiers are not only includes the number of system levels, but also an explanation of the classification criteria.

At the first tier, for government data other than personal privacy, trade secret, and other legally prohibited information, full openness should be encouraged, such as environmental data, market supervision data, and so forth. However, the scope of such openness should still be limited to the domestic area to prevent interference and improper use by foreign entities.

At the second tier, limited openness can be further divided into total prohibition and conditional openness. The former refers to legally prohibited data. The latter includes data that has undergone special processing of the former data type. When legally prohibited data undergoes data cleaning, data masking, and other necessary processes, resulting in data content that is unidentifiable, unretrievable, and has no risk of leakage, it is considered to have met the attached conditions and is included in the opening process.

At the third tier, AO model places more emphasis on public-private collaboration, and should consider the extent and scope of promoting the assetization and development of government data by the public and private sectors [11]. Therefore, not all government data can be authorized for operation by others. Data at full openness with asset value should be included in this tier. This will elevate the value of data from government decision-making reference to shared social welfare, such as the environmental data and market supervision data mentioned in the first tier that contribute to environmental optimization and market stability.

In addition, each tier can make appropriate adjustments based on the actual situation during the more specific judgment process. For example, the conditions can be included based on the level of technological development, or social needs and processing capabilities can be included in the authorization considerations, so as to achieve dynamic transformation within static tiers.

The meaning of "dual" in improved system

Specifically, what does "dual" mean in improved system? It concerns the subject and the content related to the subject. The subjects of full openness and limited openness are the government and its departments, which are types of direct openness. But the subject of AO model is the public and private sectors authorized by the government, which is an indirect type of openness, intending to release the value of government data through the collaboration and interaction among various entities. Indeed, the former has its own relatively operation logic, so it will not be elaborated upon further. Therefore, the following section will expand on the latter.

In terms of subject qualification, the purpose of AO model is to realize added value of data through the reuse of government data, thereby realizing its economic and social value as a form of public infrastructure [12]. Therefore, both enterprises driven by commercial value and social organizations aiming at social value may serve as authorized entities, the focus should not rest solely on their organizational type, but rather on their capabilities. To ensure the efficient utilization of government data, while simultaneously preventing leakage and misuse, authorized entities must possess the capacity to process data, as well as the determination to leverage the value of data for the benefit of society. The assessment indicators of such capability and determination include objective factors such as the scale of the enterprise/organization, economic and technical capabilities, social reputation, and so forth. By establishing these assessment indicators as the threshold for eligibility, government data can be efficiently leveraged to create data products, thereby realizing the ultimate purpose of "data availability and invisibility."

In terms of implementation logic, the "promoting the mechanism for the rights to government data ownership and authorization, and safeguarding the public interest in the supply and use of government data" stipulated in the *Twenty Data Measures* highlights that public interest is the underlying foundation of the AO model. Furthermore, the evaluation of public interest includes whether the exercise of administrative power meets the dual standards including the standard of publicness and the standard of interest. According to the standard of publicness, there should be no excessive investor mentality. Abundant data and information should be provided to the public, assisting disadvantaged groups excluded during OGD to gain a dominant position [13], and preventing the monopoly of data resources. According to the standard of interest, the processing of government data should not only satisfy the target interest, namely the consideration of increasing the value of data during its flow, but also pay attention to the legitimate rights and interests of data subjects. For example, in the process of processing data involving personal data, commercial data, and so forth, it is essential to prevent data infringement. Essentially, this is to prevent the foundation of social trust from being shaken and to guard against the impact of questioning on the institutional construction of OGD.

In terms of operational mechanism, the incentive mechanism and evaluation mechanism should be established to maintain a dynamic balance between public interest and private interest. Simultaneously, the interaction mechanism between public and private sectors should also be designed to establish the communication and feedback model throughout the entire data processing cycle, form an effective linkage between real-time supervision and guidance, and improve the effective path for AO model of OGD.

Establishing collaborative rights protection system

As mentioned above, OGD is entirely a proactive action. In practice, to achieve efficient opening and flow of government data, local regulations often stipulate that the government is not responsible for data opening, or even intervene in data opening through incentive mechanisms to enhance enthusiasm for opening. However, such relatively liberal regulations will make it difficult for the public to obtain effective remedies when they encounter infringements on their right to information during the process of OGD. Therefore, establishing the collaborative rights protection system involving both the administrative and judicial bodies will provide effective remedies for disputes.

Remedies within the administrative system

Firstly, the complaint-feedback system of OGD platform should be improved. OGD primarily relies on OGD platform as

the medium for publicizing data information or displaying data products. In this regard, the platform's complaint-feedback mechanism is perhaps the most direct and feasible channel for redressing infringements on the public's rights. Specifically, OGD platforms and their authorized operating platforms should have dedicated entrance for receiving complaints. When natural persons, legal persons and other organizations believe that the acts during OGD violate their legitimate rights (mainly including information rights and derivative rights), they can file complaints through this entrance. Simultaneously, dedicated personnel should be assigned to feedback these complaints, and detailed operational rules regarding handling methods, procedures, and results should be stipulated to ensure efficient complaint processing. Furthermore, stable and systematic rules can prevent secondary infringements and achieve effective accountability. From a broader perspective, improving the platform's complaint-feedback system can fully collect public opinions and suggestions during the process of promoting OGD, providing a foundation for subsequent administrative legislation.

Secondly, the disputes related to OGD should be included in the scope of administrative reconsideration. The newly amended *Administrative Reconsideration Law* of September 1, 2023, establishes administrative reconsideration as the primary channel for resolving administrative disputes [14], and expands its scope to include administrative compensation and GID (Article 11 of the *Administrative Reconsideration Law*). OGD is a system designed with administrative bodies as operators and managers, and direct supervision of other administrative bodies by administrative bodies is an effective solution. At the same time, OGD and GID have an inherent connection in terms of remedies, and infringements during OGD inevitably involve the issue of administrative compensation. In this regard, including disputes related to OGD within the scope of administrative reconsideration has a basis for reference and feasibility. Specifically, this article suggests that disputes concerning OGD, including infringements of right to privacy, right to trade secret, data correction right, and so forth, should be included in the scope of administrative reconsideration. In this way, illegal acts can be corrected within the administrative system, making the process efficient and timely. Moreover, although disputes over OGD will inevitably be dealt with through litigation, it can learn from Article 23 of the *Administrative Reconsideration Law*, which stipulates that GID should be subject to mandatory administrative reconsideration prior to litigation, and introduce such system into OGD. This would reduce the waste of judicial resources and resolve disputes more effectively.

Remedies outside the administrative system

Firstly, the scope of administrative public interest litigation should be expanded to include disputes related to OGD. OGD constitutes a pivotal system designed to facilitate the expansion of the data factor market. The various stages in OGD, including the openness, flow, utilization of data, and so forth, will impact a wide array of stakeholders. In this context, opened government data, as well as the derivative generated through its close integration, will frequently intersect with personal data, commercial data, and data regarding social welfare and public livelihood, thereby becoming closely linked with broader social interests. Meanwhile, although data subjects may generally resort to judicial remedies as a universal and cushioning measure whenever OGD results in an infringement of their rights, if such infringement affects thousands of data subjects simultaneously, litigation filed by each affected data subject would result in a severe waste of judicial resources and would fail to fundamentally resolve infringement. Therefore, this article suggests that the scope of administrative public interest litigation should be expanded to include disputes related to OGD. Such relief should be guided by the criterion of impact and encompass a diverse range of causes of action, including the protection of personal data, commercial data, and so forth. Concurrently, by establishing the complementary safeguard system that structured around a framework comprising pre-litigation prosecutorial recommendations, mid-litigation prosecutorial investigation and verification [15], and post-litigation prosecutorial trial supervision, it becomes possible to effectively resolve disputes regarding OGD that are closely linked to the social interest.

Secondly, the scope of administrative contracts litigation should be expanded to include disputes related to OGD. In practice, the Chinese license for open government data (CLOD) is employed to regulate data utilization activities. Under this type of license, the data-opening body and the user explicitly define their legal relationship, thereby clarifying the rights and obligations of both parties [16]. It is thus evident that CLOD is a kind of typical administrative contracts. And according to the *Administrative Procedure Law*, it is also within the scope of administrative litigation. However, CLOD has not been fully emphasized in China's local practice of OGD, and only the Zhejiang Province leverages the similar license to set out the use of rights, obligations, and liability for breach of contract, whereas other OGD platforms in other provinces do not accord this matter adequate importance [16]. Consequently, given the current reality where administrative contracts constitute an essential component of OGD yet their practical application often receives insufficient attention, resorting to administrative contracts litigation as a remedy for disputes arising from such agreements (including disputes caused by unclear agreements, infringements upon rights, and so forth) may well become the norm in the future. Therefore, based on the broad flow and high mobility characteristics of government data, it is necessary to explore and improve the operational mechanisms for administrative contracts litigation concerning CLOD. In particular, the operational mechanisms should be established to review the standardization of contractual language, review the neutrality of standard terms, facilitate provisional evidence deposit and online litigation. The aim is to standardize administrative contracts, resolve disputes efficiently, safeguard data openness, and facilitate flow of data.

Author Contributions: L.R. is responsible for framework design, writing all parts of this manuscript, review, and editing. F.Y is re-

sponsible for writing the parts 3 & 4 of this manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This article is supported by "the 2024 Research Project of Sichuan Medical Law Research Center-China Health Law Society" (YF24-Q20).

Informed Consent Statement: Not applicable.

Data Availability Statement: All of the data used in this article can be found in public database or document.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Shi, Y. (2023). Dilemma of Open Utilization of Government Data and Its Legal Guarantee Path in Era of Big Data. Journal of Beijing University of Posts and Telecommunications (Social Sciences Edition), 25(06), 70-78. https://doi.org/10.19722/j.cnki.1008-7729.2023.0113
2. Zhao, X., Ji, X. & Fan, Z. (2023). Evolution from Open Government Data to Open Public Data. Information Studies: Theory & Application, 47(04), 50-58+83. https://doi.org/10.16353/j.cnki.1000-7490.2024.04.008
3. Peng, Y. (2023). The Value Creation Role of Open Government Data: The Perspective of Firm's Total Factor Productivity. Journal of Quantitative & Technological Economics, 40(09), 50-70. https://doi.org/10.13653/j.cnki.jqte.20230725.001
4. Zheng, L. (2015). Study on Open Government Data: Definitions, Factors and Interactions. Chinese Public Administration, (11), 13-18. https://doi.org/10.3782/j.issn.1006-0863.2015.11.02
5. Huang, H., Zhao, Q. & Zhang, R. (2016). A Study of Government Information Freedom and Data Open: Reflection and Reconstruction of Existing Views. Chinese Public Administration, (11), 13-18. https://doi.org/10.3782/j.issn.1006-0863.2016.11.01
6. Wang, W. (2020). The Relationship Between Open Government Data and Government Information Disclosure. Law and Economy, (01), 13-24. https://doi.org/10.16823/j.cnki.10-1281/d.2020.01.002
7. DMG Lab Fudan University (2025). Report on the Opening and Utilization of Local Public Data in China 2025. http://ifopendata.fudan.edu.cn/report.
8. Fu, X. & Zheng, L. (2013). A Review of Domestic Research on Open Government Data. E-government, (06), 8-15. https://doi.org/10.16582/j.cnki.dzzw.2013.06.008
9. Song, S. (2021). Open Government Data should Adopt A Legislative Approach that is Different from Information Disclosure. Law Science, 470 (01), 91-104.
10. Gao, L. & Han, B. (2023). Legislative Construction of the Authorization Operation System from the Perspective of Typological Openness of Public Data. Shanghai Legal Studies, 6, 15-25. https://doi.org/10.26914/c.cnkihy.2023.016790
11. Wu, L. (2023). The Trend of Authorization Operation of Government Data through Public-Private Collaboration and the Improvement of Law. Oriental Law, (06), 43-52. https://doi.org/10.19404/j.cnki.dffx.20231107.001
12. Wu, L. (2023). The Improvement of Law in the Governance of the Government Data Authorization Operation. Legal Forum, 38(01), 111-121.
13. Henninger, M. (2013). The Value and Challenges of Public Sector Information. Cosmopolitan Civil Societies: An Interdisciplinary Journal, 5(3), 75-95. https://doi.org/10.5130/ccs.v5i3.3429
14. Jing, L. (2023). Provide Institutional Guarantee for Administrative Reconsideration to Become the Main Channel for Resolving Administrative Disputes: Interpretation of the Newly Amended Administrative Reconsideration Law. https://www.chinacourt.org/article/detail/2023/09/id/7509751.shtml.
15. Wang, M. & Wang, X. (2023). Research on Administrative Public Interest Litigation on Open Government Data. Beijing College of Politics and Law Journal, (03), 85-93.
16. Song, S. (2021). Is Open Government Data An Upgraded Version of Government Information Disclosure?: Based on the Comparison of Institutional Frameworks. Global Law Review, 43(05), 52-66.

JOURNAL OF GLOBAL TRENDS IN SOCIAL SCIENCE

J G T S S

Research article

https://doi.org/10.70731/8zydzz10

Exploring Key Dimensions of AI-powered Digital Human Live Streaming: A Qualitative Study Based on In-Depth Interviews with Multiple Stakeholders

Dan Hua[1,*], **Vincent Wee Eng Kima**[1]

Received 2 February 2026

Accepted 24 March 2026

Published 31 March 2026

Abstract: AI-powered digital human live streaming is emerging as a significant marketing format in the e-commerce sector. however, academic understanding of its core characteristics remains confined to conceptual transposition and theoretical deduction, lacking empirical evidence of actual perceptions from consumers and multiple stakeholders. Through in-depth interviews with 19 multiple stakeholders from China (Generation Z consumers, e-commerce practitioners, and academics), this study employed thematic analysis with manual coding throughout the process to systematically identify and define three core dimensions of AI-powered digital human live streaming: anthropomorphism, intelligent interactivity and personalized recommendation, while further revealing the multi-layered internal structures of each dimension. The findings reveal that anthropomorphism comprises three levels: visual anthropomorphism, Behavioral anthropomorphism and emotional anthropomorphism; intelligent interactivity encompasses three elements: response immediacy, response relevance and conversational coherence; and personalized recommendation consists of three dimensions: content relevance, perceived personal attention and interaction customizability. This study provides an empirically grounded, user-language-based feature dimension framework for the field of AI-powered digital human live streaming, bridging the conceptual gap between macro-theoretical concepts and micro-level user perceptions, and laying a solid conceptual foundation for subsequent scale development and quantitative model testing.

Keywords: AI-powered digital human live streaming; Feature dimensions; Anthropomorphism; Intelligent interactivity; Personalized recommendation

ISSN 2759-7830 (Online)
ISSN 2760-2508 (Print)

Introduction

AI-powered digital human live streaming is reshaping China's e-commerce ecosystem at a remarkable pace. According to the *Report on the High-Quality Development of Livestream E-commerce* (China International E-commerce Center Research Institute, 2025), as a technological form integrating computer vision, large language models and multimodal interaction, AI-powered digital human live streaming has been widely deployed on platforms such as Taobao, JD.com and Douyin, demonstrating significant advantages such as 24/7 service, controllable costs and consistency in information delivery. However, in stark contrast to the fervor of industry practice, academic research on the theoretical framework of AI-powered digital human live streaming has lagged significantly. A fundamental question remains unresolved: how do consumers perceive and describe AI-powered digital human live streaming? What are the key dimensions that constitute this unique experience?

This theoretical ambiguity severely hampers knowledge accumulation in this field. Existing research on the operationalization of AI-powered digital human live streaming characteristics is fragmented. Some studies directly borrow the concepts of perceived ease of use and perceived usefulness from traditional technology acceptance models, reducing AI anchors to mere technological tools (Davis, 1989; Hwang & Lim, 2025). Others transplant interaction and professionalism scales designed for human anchors, overlooking the funda-

[1]Universiti Tun Abdul Razak (UNIRAZAK), Kuala Lumpur 50400, Malaysia.
*Corresponding author. Email: hu.dan@ur2.unirazak.edu.my

mental differences in the interactive agency of AI anchors (McMillan & Hwang, 2002; Zhang et al., 2024). Yet others measure anthropomorphism as a single-dimensional construct without further differentiation, failing to capture the multi-layered connotations of appearance, behavior and emotion (Waytz et al., 2014; Kühne & Peter, 2023). This top-down theoretical approach results in a significant conceptual gap between measurement tools and consumers' actual experience. Current research often treats characteristics such as personification, intelligence and interactivity as isolated variables in a fragmented manner, and has yet to establish a classification system that is both consistent and theoretically grounded. Therefore, this study aimed to return to the authentic perceptions of consumers and multiple stakeholders to systematically identify and define the core characteristic dimensions of AI-powered digital human live streaming.

Given the fragmentation and conceptual gap in existing research on theoretical conceptualization, this study adopts an exploratory qualitative research design. Through in-depth interviews with 19 multiple stakeholders (Generation Z consumers, e-commerce practitioners and academics), it employs thematic analysis to systematically identify the core characteristic dimensions of AI-powered digital human live streaming and their underlying structure. This study aims to construct an integrative conceptual framework grounded in user language, reveal the multi-layered composition of consumer experience, and provide an empirical foundation for subsequent scale development and corporate practice.

Literature Review

Conceptual and contextual characteristics

The technological evolution of AI-powered digital human live streaming has undergone a hierarchical transition from virtual humans to digital humans and finally to AI-powered digital humans. Virtual humans generally refer to digital avatars based on computer-generated technology that possesses human appearance or behavioral characteristics, emphasizing virtual existence but lacking intelligent interactive capabilities (Griffor et al., 2017). Digital humans achieve high realism through high-precision 3D modeling and motion capture, but their Behavioral logic remains dependent on pre-set scripts (Hetherington & McRae, 2017). AI-powered digital humans, on the other hand, integrate natural language processing, emotion recognition and autonomous decision-making capabilities, enabling real-time, personalized interaction with users (Machidon et al., 2018; Xu et al., 2026). This technological evolution has transformed AI-powered digital humans from objects to be viewed as entities capable of dialogue.

In the context of livestream e-commerce, AI-powered digital humans exhibit three distinctive characteristics. First, they offer extended operational hours, enabling 24/7 uninterrupted service. Second, they demonstrate intelligent interaction, engaging in open-ended dialogue and context-aware responses based on large language models. Third, they provide personalized content, dynamically adjusting recommendation strategies through real-time data analysis. These characteristics create cognitive tension for consumers, who are aware that the entity is not human yet perceive it as human-like; that is, while they clearly recognize the algorithmic nature of the interactive entity, they also perceive cues of social interaction within its anthropomorphic presentation. This experience falls precisely within the specific range described by Mori's (1970) 'Uncanny Valley' theory: When an AI anchor closely resembles a real person yet reveals non-human characteristics, it may trigger cognitive dissonance and emotional discomfort; however, when its emotional expression capabilities exceed a specific threshold, it can establish a deep connection that transcends mere appearance (Mori et al., 2012). This unique experience is the focal point of this study.

Theoretical foundations: Social presence and technological trust

AI-powered digital human live streaming blurs the boundary between humans and machines, with consumers simultaneously perceiving the algorithmic nature of the experience while experiencing social interaction cues. To systematically capture this contradictory experience, this study employs social presence theory and technological trust theory as sensitizing concepts (Blumer, 1954) to guide data interpretation. This approach aligns with the methodological tradition of grounded theory, which maintains theoretical openness whilst utilizing existing concepts to enhance sensitivity to empirical phenomena.

The theory of social presence describes the degree to which individuals perceive the actual presence of others during mediated interaction (Short et al., 1976). In the context of AI-powered digital human live streaming, this theory helps identify how consumers perceive the social presence of digital avatars through the screen, including their perception of immediate responsiveness, expectations of emotional resonance, and the cognitive tension of 'knowing they are not human yet perceiving them as human-like'. This concept provides an initial analytical framework for exploring immersive experiences and the sense of being valued.

The theory of technological trust distinguishes three dimensions: competence trust (confidence in algorithmic efficacy), benevolence trust (skepticism regarding platform intentions), and honesty trust (judgement of information authenticity) (McKnight et al., 2002). Given that the recommendation behavior of AI anchors simultaneously exhibits intelligent professionalism and algorithmic uncertainty, this theory provides a conceptual framework for understanding the ambivalent psychological state of consumers when faced with algorithmic decisions (acknowledging their efficiency while questioning their motives).

Together, these theories constitute the conceptual framework for analyzing interview data. It is important to emphasize that this study does not presuppose causal pathways or dimensional structures for these theoretical variables, but rather treats them as perspectives for identifying emic constructs. During the data analysis process, the research maintained theoretical openness to emergent dimensions (such as potential algorithmic aversion or digital ethical anxieties), allowing user language to reconfigure or transcend existing theoretical frameworks.

Research gaps and theoretical opportunities

Existing research has three significant limitations. First, the research subjects exhibited fragmented characteristics. Existing literature dissects the appearance, interaction and recommendation functions of AI anchors into isolated variables for separate examination, lacking systematic attention to them as vehicles for an integrated experience; characteristics such as anthropomorphism and intelligence are treated in a piecemeal manner, and a consistent classification system has yet to be established. Second, the underlying mechanisms are insufficiently explained. Most existing models adopt a simplified 'features → psychological variables → behavior' pathway, which disconnects the complex psychological evolution of consumers and assumes that all consumers follow a homogeneous path, thereby overlooking the differences between groups. Third, research methods are relatively limited. Existing literature relies excessively on structured questionnaires; in an emerging field where 'it is not yet known what to measure', there is a conceptual gap between pre-defined scales and consumers' actual experiences, and qualitative exploration is severely lacking.

These limitations constitute a threefold theoretical gap in the field, creating opportunities for knowledge innovation. First, constructing an integrated conceptual framework for AI-powered digital human live streaming through qualitative induction can overcome the theoretical bottlenecks of the current fragmented research paradigm, establishing analytical dimensions that encompass both social presence and technological trust. Second, by deeply describing the heterogeneous evolutionary pathways of consumer psychological experiences, a process theory distinct from linear models can be developed to explain how different groups negotiate human-machine relationships in diverse ways. Finally, a grounded conceptual system built upon users' native language will provide an embodied theoretical foundation for subsequent scale development, bridging the theoretical gap between operationalized measurements and real-world experiences. Based on this, this study uses in-depth interviews to identify core perceptual dimensions, describing the complexity and group differences in psychological experiences, thereby providing a theoretical framework grounded in users' language for the field of AI anchors consumer behavior.

Research Design

This study adheres to the interpretivist research paradigm (Creswell & Poth, 2018) and adopts an exploratory qualitative research design. Data were collected through semi-structured in-depth interviews, and thematic analysis was employed for manual coding throughout the process (Braun & Clarke, 2022).

Data collection: Semi-structured in-depth interviews

Design of the interview guide

The interview guide was structured around four core modules: experience entry and situational recall; perception of characteristics and concrete descriptions; psychological mechanisms and behavioral decision-making; and boundary exploration and future prospects. First, situational priming was used to guide participants in recalling specific AI live-streaming experiences. Second, feature descriptions invited participants to characterize the performance of AI digital human anchors using everyday language. Third, mechanism exploration delved into psychological responses such as trust and a sense of presence. Finally, boundary exploration revealed unmet underlying needs.

This study was adapted to suit different respondent groups. For Generation Z consumers, everyday language was used, such as "What was your first visual and auditory impression of this AI anchor? In what ways did it feel 'like a real person' to you, and in what ways did it still feel 'like a machine' ?" For e-commerce practitioners, the focus was on commercial logic, technical implementation and industry insights, such as "What are the primary technical or cost bottlenecks currently limiting the ability of AI-powered digital human to achieve natural, in-depth interactions?" Academics focused on theoretical critique, such as "When applying traditional anthropomorphism theories to AI-powered digital human with interactive capabilities, do their conceptual dimensions need to be re-examined and expanded?"

Sampling strategy and sample characteristics

This study employed purposive sampling combined with the maximum diversity sampling method (Patton, 2015), aiming to achieve the widest possible conceptual coverage of the research phenomenon within a small sample by selecting cases that exhibited diversity across specific dimensions. Based on this strategy, this study established three key stakeholder groups as the sampling framework (**Table 1**).

The final sample comprised 19 participants. The sample size adhered to the principle of theoretical saturation (Saunders et al., 2018). No substantive new themes emerged following the coding of data from the first 16 respondents. Three subsequent interviews confirmed the stability of the thematic framework, leading to the conclusion that theoretical saturation had been achieved.

Data collection procedure

Data collection took place between September and December 2025, employing a combination of in-person and online video interviews, with each session lasting 40–60 minutes. All interviews were audio-recorded with the participants' informed consent, transcribed verbatim within 24 hours, and retained colloquial features and emotional nonverbal cues (e.g., [loud laughter], [pause]).

Ethical considerations

This study strictly adhered to the following academic ethical standards. First, regarding informed consent, an "Informed Consent Form" was sent to participants prior to the interview, detailing the research objectives and the methods of data anonymization. Second, regarding privacy protection, all identifiable information was thoroughly anonymized during the transcription stage (coding system: C01–C10, P01–

Table 1 | Sample Composition of Respondents

Group	Sampling Criteria	Sample Size	Key Contribution
Generation Z consumers	Aged 18–25, weekly active users on mainstream e-commerce platforms, with at least one distinct and memorable experience of watching an AI-powered digital human livestreaming within the past three months	10	Providing authenticity of experience: the ultimate source of construct validity
E-commerce Practitioners	Over 5 years of industry experience, working at leading livestream e-commerce platforms, well-known brands, or digital human technology firms, with direct involvement in decision-making or operations of AI-powered digital human live streaming projects	5	Ensuring practical applicability: guaranteeing the feasibility of the dimensions
Academics	Associate Professor or Professor, with research focusing on digital media, consumer behaviour or human-computer interaction	4	Providing theoretical rigour: ensuring alignment with academic literature

P05, A01–A04). Third, regarding data security, all raw files were stored in an encrypted form, and access was strictly restricted. Finally, regarding voluntary participation, interviewees were reminded throughout the study that they retained the right to withdraw from the interview at any time.

Data analysis: Thematic analysis

Integrative analysis strategy

This study adopted a cross-group integrative analysis strategy, merging data from the three groups of respondents for unified coding. This strategy is based on three theoretical considerations. First, the overarching nature of the research objectives requires a focus on constructing a system of characteristic dimensions rather than comparing group differences; the value of the three groups lies in their mutual complementarity and cross-validation (Chen, 2000). Second, the diverse pathways to theoretical saturation indicate that integrating the perspectives of consumers, practitioners and academics enables broader theoretical boundaries to be explored within a shorter timeframe (Wang, 2007). Finally, the prior assurance of construct validity ensured that the extracted dimensions possessed the triple attributes of experiential perceptibility, technical feasibility and theoretical dialogability.

Throughout the coding process, the researchers maintained sensitivity to the perspectives of different groups and conducted preliminary comparative analyses. The results revealed that, although the descriptions provided by the three groups differed in wording, they were highly consistent at the level of abstract constructs, pointing to three core themes: anthropomorphism, intelligent interactivity, and personalized recommendation. This provides robust cross-group evidence for the dimensional framework. By integrating these three groups, this study achieved qualitative assurance of construct validity: consumers provide experiential authenticity, practitioners ensure practical operability, and scholars contribute theoretical logic, thereby establishing a reproducible, multi-stakeholder triangulation paradigm for qualitative research in fields characterized by rapid technological iteration.

Thematic analysis procedure

This study employed the six-step thematic analysis method proposed by Braun and Clarke (2022) for manual coding throughout the process, emphasizing deep immersion in the data and iterative analysis.

Stage 1: Immersion in the data and preliminary interpretation. The researcher conducted active, repeated and critical readings of the interview transcripts, recording analysis notes to capture initial patterns and underlying concepts.

Stage 2: Generating open codes. A line-by-line analysis was employed, prioritizing the use of respondents' native concepts as code labels. For example, regarding appearance and emotional aspects: "dimples when smiling" (C03), "slight fine lines at the corners of the eyes" (C04), "micro-expressions can't keep up, so users still find it a bit artificial" (P03), "60 marks for appearance is enough, but 80 marks for emotion is essential" (A02); interaction quality-related comments such as: "a three-second reply but the information is incomplete" (C01), "picks up on the joke and stays on topic through multiple rounds" (C07), and "conversation retention is more important than responding to the first sentence" (P04); and at the recommendation service level: "I'd just mentioned wanting a lipstick shade to make my skin look fairer, and she immediately recommended that Rotten Tomatoes shade–it was exactly what I wanted" (C02), "It felt like it was prepared just for me" (C08), "Real-time recommendations based on behavior within the livestream" (P01). A total of 127 open codes were generated at this stage, covering the diverse perspectives of the three groups.

Stage 3: Formulation of focused codes. Logical connections were identified through repeated comparison and clustering of open codes. For example, physical details such as dimples and fine lines, together with vocal tone, point to the personification of appearance; the timing of nods and the rhythm of pauses point to the personification of behavior; and "feeling"-related vocabulary and expressions of empathy point to the personification of emotion. Similarly, codes relating to interactive efficacy (*instant replies*, *picking up on jokes*, *staying on topic*) were clustered into responsiveness, relevance of responses, and conversational coherence. Codes relating to recommendation services (*tailored preparation*, *real-time adjustment*) were clustered into content relevance, sense of personal attention, and interactive customization. This stage yielded nine candidate sub-dimensions, falling under the three core themes of anthropomorphism, intelligent interactivity, and personalized recommendation. Taking the anthro-

Table 2 | Example of thematic analysis coding evolution: Taking anthropomorphism as an example

Coding Level	Coding Type	Representative Content	Source Examples
Open Coding (Initial code)	Local Concept Labels	"She has dimples when she smiles, which I find very soothing"	C03
		"There are tiny fine lines at the corners of her eyes—the kind of natural creases you see when a real person smiles"	C04
		"Large language models can make digital human say 'I understand how you feel', but the micro-expressions don't keep up, so users still find it a bit artificial"	P03
		"A score of 60 for appearance is enough; emotions must score 80."	A02
		"Pause briefly before answering; it gives the impression they're thinking"	C05
		"The timing of the nods and pauses makes it feel like they're really listening to me, rather than following a pre-written script"	C06
		(Other codes relating to appearance, behaviour and emotional details omitted)	
Focus coding (Candidate themes)	Mid-level concept aggregation	Visual Anthropomorphism: dimples, wrinkles, tone of voice and other physical details	C03, CO4
		Behavioral Anthropomorphism: Timing of nods, rhythm of pauses, naturalness of body language	C05, C06
		Emotional anthropomorphism: "feeling" vocabulary, expressions of empathy, prioritisation of emotional capabilities	C03, P03, A02
Theoretical coding	Refining the Concept	Anthropomorphism: The extent to which AI-powered digital humans simulate human external characteristics and emotional expressions in terms of vision, speech, facial expressions and movements, comprising three levels: visual anthropomorphism, Behavioral anthropomorphism and emotional anthropomorphism.	Integration of all the above focused codings (consensus among consumers, practitioners and academics)

pomorphism dimension as an example, its coding evolution is presented in Table 2.

Stage 4: Review and refinement of themes. Based on the candidate themes identified above, this phase involved a systematic review of the nine candidate sub-dimensions. The internal consistency of the themes and the representativeness of the overall dataset were assessed; sub-dimensions with overlapping semantics were merged, such as combining *favorable physical appearance* and *human-like voice* into visual anthropomorphism. Candidate themes with low consensus were excluded, such as *entertainment value* and *curiosity*, which were provisionally classified as marginal themes due to their low frequency of occurrence and weak cross-group consensus. After several iterations, the independence and stability of the nine sub-dimensions were confirmed, with clear boundaries and distinct characteristics for each dimension.

Stage 5: Defining and naming themes. The connotations of the core themes and their respective sub-dimensions were precisely defined. Anthropomorphism comprises three levels: visual anthropomorphism (the degree of human resemblance in visual appearance and voice), Behavioral anthropomorphism (the naturalness of body language and interaction rhythm), and emotional anthropomorphism (emotional recognition and the ability to express empathy). Intelligent interactivity encompasses three elements: response immediacy (reaction speed), response relevance (content matching), and conversational coherence (ability to maintain context). Personalized recommendation consist of three levels: content relevance (accuracy of information matching), sense of personal attention (experience of being valued and understood), and interactivity customization (flexibility in real-time adjustment of recommendation strategies). The boundaries between each dimension are clear; while distinct from one another, they collectively form an integrated experiential framework.

Stage 6: Report writing. The findings from the preceding analyses were systematically integrated to construct a comprehensive narrative around each core theme, presenting empirical evidence and theoretical implications in detail to form the main body of Part 4 (Research Findings) of this study.

Through the analysis of the above six stages, this study distilled three core themes and nine sub-dimensions from 127 open-coded entries, systematically presenting the structural dimensions of AI-powered digital human live streaming.

Reflective strategies

This study ensured the rigour of the analysis through the following strategies. First, regarding reflective documentation, the researcher maintained analytical memos throughout the data collection and analysis process, meticulously recording theoretical assumptions, emotional responses, and conceptual uncertainties, thereby mitigating researcher bias through continuous self-reflection. Second, through counter-example analysis, the researcher proactively examined data fragments that contradicted the dominant themes (such as consumers' positive evaluations of the non-anthropomorphic characteristics of AI anchors), ensuring that contradictory evidence was not selectively ignored. Third, a comprehensive audit trail was established, systematically retaining raw transcripts, coding tables, analysis memos, iterative versions of themes, and decision logs to ensure the research process was auditable and reproducible. Finally, three participants were invited to review the preliminary findings through peer review to confirm that the extracted themes closely aligned with their original intentions.

Research Findings

Through systematic thematic analysis, this study distilled three core feature dimensions with a high degree of cross-group consensus from in-depth interview data with multiple stakeholders: anthropomorphism, intelligent interactivity, and personalized recommendation. The connotations and multi-layered internal structures of each dimension are elaborated below.

Anthropomorphism: A multi-layered evolution from physical resemblance to emotional resonance

In this study, anthropomorphism is defined as the degree to which AI-powered digital humans simulate human external characteristics and emotional expressions in terms of visual appearance, speech, facial expressions, and movements. This definition goes beyond the approach taken in traditional research, which treats anthropomorphism as a single dimension, and reveals its theoretical implications as a multi-layered construct.

Visual anthropomorphism

Visual Anthropomorphism constitutes the most fundamental level of anthropomorphism, encompassing both static and dynamic visual manifestations, such as the AI anchor's facial features, body proportions, texture of their clothing, lip-sync accuracy, and naturalness of their voice. Consumers are extremely sensitive to such characteristics, often forming overall judgements based on these details.

"Although the presenter looks artificial at first glance, you will notice that when she smiles, the corners of her mouth turn up, and there are slight fine lines at the corners of her eyes—the kind of natural creases you see when a real person smiles. Just this detail alone makes her feel far more comfortable to watch than those other presenters with stiff facial expressions." (Consumer C04)

In addition to visual characteristics, the naturalness of the voice is a key component of a human-like appearance, directly influencing the perception of how human it sounds.

"Voice is the biggest pitfall. Many AI voices sound instantly like TTS (text-to-speech), with a flat pitch for every syllable. But a good voice has variations in intonation and even emphasizes certain words when highlighting a product's selling points. That's the key to sounding human." (Practitioner P02)

Behavioral anthropomorphism

Behavioral anthropomorphism, in contrast to static appearance, refers to the degree to which an AI-powered digital human's movements, postures and reactions resemble human Behavioral patterns. Respondents generally believed that Behavioral anthropomorphism contributed even more to a sense of presence than visual anthropomorphism.

"The moment that really made me feel it was conscious was when, after I posted a comment, it paused briefly, gave a slight nod, and then said, 'That's a good question, friend.' The timing of that nod and pause made it feel as though it was genuinely listening to me, rather than following a pre-programmed script." (Consumer C06)

Achieving this requires subtle control of the timing and motion design, striking an optimal balance between zero latency and anthropomorphic delay.

"From a technical perspective, response latency is a double-edged sword. Zero latency makes users feel that it is a preset script; however a variable latency of 200–400 milliseconds, combined with appropriate head movements, creates the illusion that it is thinking. This is what we refer to as the anthropomorphism threshold." (Practitioner P03)

Emotional anthropomorphism

Emotional anthropomorphism represents the highest level of anthropomorphism, referring to an AI-powered digital human's ability to convey emotions, express empathy and establish emotional connections. Consumers' descriptions often carry strong emotional undertones, reflecting a deep-seated desire to be understood.

"I know it's just a programme, but once I asked about three different models of robot vacuum cleaners across various price points. It finally said, 'Based on your questions just now, I get the feeling you're quite focused on value for money; this model at two thousand three hundred might be the best fit for you.' When it used the word *feel*, I was genuinely taken aback—can AI really have feelings? But in that very moment, I felt it understood me." (Consumer C05)

Academics further have pointed out, from a theoretical perspective, that emotional anthropomorphism is the key to crossing the Uncanny Valley and establishing quasi-social relationships, and its importance surpasses that of physical realism.

"Anthropomorphism is not about making AI look more and more like a human, but about making people willing to treat it as a social partner. A 60 out of 100 for appearance is sufficient, but emotional expression must score over 80; otherwise, it will fall into the uncanny valley." (Academic A02)

Intelligent interactivity: Cognitive interaction beyond merely responding

In this study, intelligent interactivity is defined as the comprehensive ability of an AI-powered digital human to understand complex user intentions, maintain coherent dialogue with contextual relevance, and provide timely, accurate and contextually appropriate responses. The core of this dimension lies in intelligence; it goes beyond the emphasis on response speed found in traditional human-computer interaction research and points toward a deeper level of semantic understanding.

Response immediacy

Response immediacy is a fundamental element of intelligent interactivity. Although consumers expect interactions to occur instantly, they cannot tolerate a sacrifice in the quality of answers in the pursuit of speed.

"I asked it, 'What's the difference between this phone and the Pro version?' It replied within three seconds but only mentioned the screen size; it did not mention the differences

in the chip and camera, which are the very things I care about the most. I might as well have looked at the specs myself." (Consumer C01)

E-commerce practitioners, however, emphasize that true intelligence lies in the sustainability of the conversation rather than merely the speed of the initial response; conversation retention rates are a more critical operational metric.

"Many technical teams focus solely on optimising the response latency for the first sentence; however, true user satisfaction stems from conversation retention rates, that is whether users are willing to continue asking follow-up questions. This requires an understanding of context; it is not merely a matter of speed." (Practitioner P04)

Response relevance

Response relevance is a hallmark feature distinguishing intelligent interaction from traditional interaction. Respondents repeatedly mentioned local concepts such as "keeping up with the banter" and "staying on topic across multiple rounds", emphasizing the precise alignment of answers with the intent of the question and the ability to retain contextual memory.

"I chatted with it for five minutes, asking about everything from lipstick shades to foundation shades, and it actually remembered that I'd mentioned at the start that I have dry skin. At one point, I digressed to ask about delivery, and after answering, it proactively said, 'Let's get back to the foundation you were asking about earlier.' At that moment, I genuinely felt that I was being taken seriously." (Consumer C02)

This ability to retain contextual information relies on complex intent recognition and dialogue management technologies, requiring a balance between open-domain dialogue and task-oriented interactions.

"Currently, integrating large language models into live streaming is a growing trend, but the challenge lies in maintaining conversational coherence whilst preventing the bot from being sidetracked by users onto completely unrelated topics. This requires a highly refined design of intent recognition and dialogue strategies." (Practitioner P05)

Conversational coherence

Conversational Coherence is reflected in an AI-powered digital human's ability to perceive the conversational context and proactively adjust its interaction strategy, marking a transition from passive responses to active service.

"During one live stream, many people in the comment section asked if there were any discount vouchers for the same product. The AI anchor first explained the process for claiming the voucher, then proactively pinned the voucher link to the top of the screen, adding, 'I see many of you are still asking, so I've placed the link here for everyone to claim with a single click.' This was no longer merely answering a question, but solving a problem." (Consumer C04)

Academics regard this capability as the ultimate form of intelligent interaction, in which the system is not only capable of responding but also of anticipating and adapting, thereby truly possessing the value of a host rather than a mere tool.

"Adaptability is the ultimate form of intelligent interaction. It involves not only understanding what the user says but also what they do not say; for instance, sensing the collective sentiment in the comments section to determine whether the pace of the explanation needs to be adjusted. Only an AI capable of this truly possesses the value of a presenter rather than a mere tool." (Academic A03)

Personalised recommendation: A mechanism for a 'Thousand faces for one person' experience

In this study, personalized recommendation are defined as a service mechanism whereby AI-powered digital humans provide highly customized information content and product recommendations based on users' long-term historical behavior, real-time interactive context, or explicitly expressed preferences. The unique value of this dimension lies in the fact that it represents an area with the greatest potential for differentiated competitive advantage for AI anchors compared to human anchors.

Content relevance

Content relevance forms the basis of personalized recommendation, namely, the degree to which recommended results align with a user's explicit or implicit interests. Consumers are acutely aware of the fine line between precision and excessive intrusion, reflecting the complexities of privacy computing.

"Once it recommended a pair of Bluetooth headphones that I had searched for the previous week but had not bought. My first reaction wasn't that I was being watched, but rather, 'Oh, it knows I need this.' I don't mind recommendations like this; in fact, I feel it saves me the time of searching for them again." (Consumer C02)

However, e-commerce practitioners must carefully balance recommendation accuracy with user privacy protection to avoid a sense of surveillance triggered by cross-domain data retrieval.

"Personalisation must not cross the privacy line. Our current strategy is to make real-time recommendations 'based on behavior within the live stream,' while minimizing cross-domain access to user data from other apps. Although this reduces accuracy, it increases users' sense of security." (Practitioner P01)

Perceived personal attention

Perceived personal attention emphasizes the user's subjective experience of being treated as special, rather than merely the accuracy of algorithmic matching. This perception stems from the recognition of the social cue that something is "specifically for me."

"It says, 'These coffee beans have been specially prepared for you because you previously browsed pour-over equipment.' I know it's an algorithmic match, but when it says 'specially for you', it really makes me feel as though this streamer is my personal shopping assistant." (Consumer C08)

In operational practice, simple strategies such as repeating a user's nickname or remembering past preferences can

significantly enhance users' sense of being noticed, thereby improving user satisfaction.

"The essence of the perception of personalisation is making the user feel that 'this AI remembers me.' Our tests have shown that even simply repeating a user's nickname, or recalling preferences mentioned in their previous conversation, leads to a significant increase in user satisfaction." (Practitioner P02)

Interaction customizability

Interaction customization manifests as the ability to make dynamic adjustments based on real-time interactions. This represents the unique value of AI over static algorithmic recommendations, marking a leap from a *one-size-fits-all* approach to a *thousand faces for one person* approach.

"I told it that I did not like overly sweet snacks, and it immediately removed the chocolate biscuits from the recommendation list and replaced them with savoury nuts. This real-time adjustment convinced me more than any algorithmic claim that 'it understands me'." (Consumer C06)

Academics have pointed out from a theoretical perspective that this real-time interactive customization capability constitutes a competitive moat for AI anchors relative to human anchors, and represents the core competitive advantage of AI-powered digital human live streaming.

"Interactive customisation is the only way AI anchors can surpass human anchors. When broadcasting to an audience of tens of thousands, a human presenter cannot remember what each individual has said; however, an AI anchor can. If utilised effectively, this capability serves as the moat for AI-powered digital human live streaming." (Academic A01)

Discussion

Theoretical contributions

This study provides a feature-dimensional framework rooted in user language for the field of AI-powered digital human live streaming, bridging the conceptual gap between macro-theoretical concepts and micro-level user perceptions. Specifically, this study makes breakthroughs at the following three theoretical levels.

Multidimensional deconstruction of anthropomorphism theory. Using empirical data, this study reveals the existence of emotional anthropomorphism as an independent dimension, identifying it as the key mechanism for overcoming the uncanny valley and establishing emotional connections between humans and machines. Furthermore, the concept of Behavioral anthropomorphism proposed in this study (such as the anthropomorphism threshold for reaction time) provides a concrete operational pathway for Moriuchi's (2021) call for hierarchical measurement, thereby overcoming the theoretical limitations of traditional one-dimensional continua.

Cognitive shift in the interactivity construct. This study shifts theoretical focus from formal characteristics such as response speed to cognitive quality, proposing that intelligent interactivity comprises three elements: response immediacy, response relevance, and conversational coherence. This construct expands upon the dimensional classification proposed by Li et al. (2025), emphasizing that consumers expect proactive intelligence rather than passive responses, thereby extending the applicability of the technology acceptance model (TAM) in AI contexts.

Emotion-technology integration in personalization research. This study finds that consumers' perception of personalization transcends algorithmic accuracy; the core essence is constituted by a sense of being specially treated and the interactive process of real-time adjustment. Specifically, this hierarchical structure comprises three levels: content relevance, perceived personal attention, and interaction customization, thereby broadening the technical research perspective on personalized recommendation. More importantly, this study reveals the tension between privacy concerns and expectations of personalization: consumers anticipate the efficiency and sense of exclusivity brought by highly customized services, while remaining vigilant about the risk of data breaches. This ambivalence is particularly pronounced in contexts of uncertainty regarding privacy breaches (Lim & Kim, 2025).

Implications for practice

The findings of this study offer direct practical guidance for enterprises in the design, operation and iteration of AI-powered digital human live streaming. Based on the independence of these three dimensions, this study proposes the following differentiated resource allocation and optimization strategies.

Adopt a principle of moderation in anthropomorphic design. Consumers do not seek perfect human likeness but rather seek recognizable intelligent agents. Enterprises should adopt a resource-balancing strategy regarding appearance and emotion, prioritizing investment in emotional expression capabilities (such as empathetic responses and emotion recognition). In terms of technical implementation, it is recommended that Behavioral latency be controlled within a variable range of 200–400 milliseconds to create a natural sense of thought and avoid the robotic feel of instant responses. Concurrently, the AI's identity should be clearly labelled to prevent ethical controversies and trust crises arising from excessive anthropomorphism.

Prioritize conversational coherence in intelligent interactions. Industry feedback indicates that conversation retention rates are a better predictor of user satisfaction than response speed to the first sentence. It is recommended to prioritize optimizing multi-turn contextual memory and intent anchoring capabilities, rather than simply reducing latency; furthermore, collective emotional perception should be utilized to proactively trigger service actions, elevating interactions from mechanical question answering to companion-style service.

Balance privacy concerns with recommendation accuracy. It is recommended to establish a tiered customization strategy in which the foundational tier ensures privacy and security awareness based on real-time behavior within the live stream (non-cross-domain data); the emotional tier enhances the sense of being cared for through lightweight per-

sonalization methods such as repeating nicknames; the customization tier triggers real-time adjustments only when users explicitly express preferences, thereby avoiding the intrusive feeling caused by algorithmic predictions.

Research limitations and future directions

Although this study employed maximum difference sampling, the sample remained predominantly concentrated among young consumers in first- and second-tier cities, as well as specific types of practitioners and academics. Its generalizability requires careful validation when extrapolating to other groups (such as users in lower-tier markets and the elderly) and scenarios (such as B2B livestreaming and cross-border livestreaming). Furthermore, retrospective interviews may be subject to memory reconstruction bias; in particular, recollections of emotional experiences and interactive details may not be entirely accurate. Future research could employ the experience sampling method to validate these findings.

Future research could be expanded in three directions: first, developing standardized scales based on this framework to establish construct validity; second, conducting cross-cultural comparisons to examine the differential manifestations of the three-dimensional characteristics across different cultural contexts (such as high-context and low-context cultures); and third, exploring the moderating effects of product type (high-involvement or low-involvement goods) and platform environment (private-domain or public-domain livestreaming) on the perception of these characteristics.

Conclusion

Through in-depth interviews with 19 multi-stakeholders and the application of thematic analysis, this study systematically identified three core characteristic dimensions of AI-powered digital human live streaming: anthropomorphism (the progressive presentation of appearance, behavior and emotions), intelligent interactivity (the integration of immediacy, relevance and coherence), and personalized recommendation (the stratification of content precision, attention perception and interactive customization). This integrated framework transcends the fragmented examination of isolated technical variables found in existing literature, revealing the unique structure of AI anchors as carriers of experience with dual socio-technical attributes and bridging the conceptual gap between macro-level technical discourse and micro-level user perception. By providing operational definitions that combine local relevance with academic dialogue, this study not only lays the conceptual foundation for subsequent scale development and mechanism testing, but also offers a theoretical framework for explaining how AI anchors simultaneously trigger cognitive trust and emotional resonance. Future research should develop standardized measurement tools based on this framework and, through longitudinal tracking and cross-cultural comparisons, examine the dynamic evolution and situational specificity of the three-dimensional structure, thereby supporting the paradigm shift in this field from exploratory to confirmatory research.

References

1. Blumer, H. (1954). What's wrong with social theory? American Sociological Review, 19(1), 3–10. https://doi.org/10.2307/2088165
2. Braun, V., & Clarke, V. (2022). Thematic analysis: A practical guide. SAGE Publications.
3. Chen, X. (2000). Qualitative research in social sciences. Beijing: Educational Science Publishing House.
4. China International E-commerce Center Research Institute. (2025, May 14). High-quality development report of live streaming e-commerce. Beijing.
5. Creswell, J. W., & Poth, C. N. (2018). Qualitative inquiry and research design: Choosing among five approaches (4th ed.). SAGE Publications.
6. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340. https://doi.org/10.2307/249008
7. Griffor, E. R., Greer, C., Wollman, D. A., & Burns, M. J. (2017). Framework for cyber-physical systems: Volume 1, overview. National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.1500-201
8. Hetherington, R., & McRae, R. (2017). Make-believing animated films featuring digital human: A qualitative inquiry using online sources. Animation, 12(2), 166–180. https://doi.org/10.1177/1746847717710738
9. Hwang, J., & Lim, Y. (2025). Adoption of AI Services Based on the Technology Acceptance Model: A Meta-Research Approach. International Journal of Technology Diffusion, 15(1), 1–14. https://doi.org/10.4018/IJTD.394262
10. Kühne, R., & Peter, J. (2023). Anthropomorphism in human–robot interactions: A multidimensional conceptualization. Communication Theory, 33(1), 42–52. https://doi.org/10.1093/ct/qtac020
11. Li, R., Ma, B., Zhang, P., & Gao, X. (2025). Research on Mechanisms, Trends, and Impediments to Civil-military Technology Transfer in Emerging Fields. Journal of Beijing Institute of Technology (Social Sciences Edition). 27(1), 127–144, 187. https://doi.org/10.15918/j.jbitss1009-3370.2024.6959
12. Lim, S. E., & Kim, M. (2025). AI-powered personalized recommendations and pricing: Moderating effects of ethical AI and consumer empowerment. International Journal of Hospitality Management, 130, 104259. https://doi.org/10.1016/j.ijhm.2025.104259
13. Machidon, O. M., Duguleana, M., & Carrozzino, M. (2018). virtual human in cultural heritage ICT applications: A review. Journal of Cultural Heritage, 33, 249–260. https://doi.org/10.1016/j.culher.2018.01.007
14. McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. Information Systems Research, 13(3), 334–359. https://doi.org/10.1287/isre.13.3.334.81
15. McMillan, S. J., & Hwang, J. S. (2002). Measures of perceived interactivity: An exploration of the role of direction of communication, user control, and time in shaping perceptions of interactivity. Journal of Advertising, 31(3), 29–42. https://doi.org/10.1080/00913367.2002.10673674
16. Mori, M. (1970). Bukimi no tani [The uncanny valley]. Energy, 7(4), 33-35.
17. Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [From the field]. IEEE Robotics & Automation Magazine, 19(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811
18. Moriuchi, E. (2021). An empirical study on anthropomorphism and engagement with disembodied AIs and consumers' re-use behavior. Psychology & Marketing, 38(1), 21–34. https://doi.org/10.1002/mar.21407

19. Patton, M. Q. (2015). Qualitative research & evaluation methods: Integrating theory and practice (4th ed.). SAGE Publications.
20. Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., & Jinks, C. (2018). Saturation in qualitative research: Exploring its conceptualization and operationalization. Quality & Quantity, 52(4), 1893–1907. https://doi.org/10.1007/s11135-017-0574-8
21. Short, J., Williams, E., & Christie, B. (1976). The social psychology of telecommunications. London: John Wiley & Sons.
22. Xu, Y. (W.), Chi, C. G., Gursoy, D., & Cai, R. (R.). (2026). Rethinking AI anthropomorphism: A holistic conceptualization and scale across AI systems and service contexts. Technology in Society, 85, 103189. https://doi.org/10.1016/j.techsoc.2025.103189
23. Wang, N. (2007). The representativeness of case study and its sampling logic. Gansu Social Sciences. (5), 1–4. https://doi.org/10.15891/j.cnki.cn62-1093/c.2007.05.001
24. Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. Journal of Experimental Social Psychology, 52, 113–117. https://doi.org/10.1016/j.jesp.2014.01.005
25. Zhang, L., Zhang, J., Wang, D., & Mu, J. (2024). Development and validation of an AI virtual streamer scale for livestream e-commerce. International Journal of Human-Computer Interaction, 41(14), 8525–8538. https://doi.org/10.1080/10447318.2024.2411088

JOURNAL OF GLOBAL TRENDS IN SOCIAL SCIENCE
J G T S S

Review article
https://doi.org/10.70731/ptrjz986

Artificial Intelligence in Financial Decision-Making: Forecasting, Portfolio Optimization, and ESG-Related Corporate Finance Analysis

Adrian Lim [1], **Putri Rahayu** [2,*]

Received 16 January 2026

Accepted 23 March 2026

Published 31 March 2026

Abstract: Artificial intelligence has become a major methodological force in financial decision-making, but the literature remains fragmented across at least three partially connected domains: financial time-series forecasting, portfolio construction, and firm-level sustainability analysis. This review argues that these domains should be interpreted as parts of a broader decision architecture in which algorithms extract signals from noisy data, transform those signals into investment or financing choices, and then evaluate outcomes under multiple objectives that increasingly include environmental, social, and governance criteria. The review first synthesizes the evolution of forecasting methods from classical econometric models to recurrent neural networks, transformers, and hybrid architectures. It then examines how predictive outputs are translated into allocation rules, with emphasis on mean–variance optimization, shrinkage-based risk estimation, risk parity, hierarchical allocation, and reinforcement-learning-based dynamic rebalancing. The third substantive line concerns corporate finance and sustainable finance, where AI is used not only to predict ESG ratings and financial constraints but also to identify firm heterogeneity, financing frictions, and disclosure-based signals. Across these streams, the article compares predictive and explanatory models, clarifies the role of structured, textual, and alternative data, and evaluates major methodological risks including overfitting, regime instability, interpretability deficits, and institutional dependence. The central conclusion is that the next stage of research should not treat forecasting, allocation, and ESG-related corporate finance as separate literatures. Instead, future work should build integrated frameworks in which market prediction, portfolio design, and firm-level sustainable finance analysis are jointly modeled under explicit assumptions about data quality, decision frequency, and accountability.

Keywords: Artificial intelligence; Financial forecasting; Portfolio optimization; ESG; Financial constraints; Sustainable finance

ISSN 2759-7830 (Online)
ISSN 2760-2508 (Print)

Introduction

The contemporary AI literature in finance has expanded rapidly, but its conceptual organization is less clear than its technical growth. Bibliometric reviews show a sustained shift from early applications in credit scoring and bankruptcy prediction toward market prediction, portfolio construction, fraud detection, textual analytics, and, more recently, explainable finance and sustainable finance (Ahmed et al., 2022; Goodell et al., 2021; Chen et al., 2023). Even within this broad field, however, many studies still operate inside narrow technical silos. Forecasting papers focus on directional accuracy, regression loss, or benchmark beating. Portfolio studies emphasize Sharpe ratios or drawdown control. ESG and corporate finance studies examine ratings, financing frictions, or disclosure quality. What is often missing is a unifying interpretation of AI as a decision infrastructure rather than a collection of isolated algorithms.

This review starts from the proposition that AI in financial decision-making has two foundational research lines. The

[1] National University of Singapore, Singapore 117592; [2] Nanyang Technological University, Singapore 639798.
* Corresponding author. Email: p.rahayu@ntu.edu.sg

first is market prediction: identifying exploitable patterns in prices, returns, volatility, sentiment, and cross-sectional relations. The second is firm-level feature identification: using structured financial variables, disclosures, and sustainability indicators to infer latent firm characteristics such as financial constraints, governance quality, default risk, or ESG standing. The distinction matters because the first line usually targets tradable signals under short decision cycles, whereas the second line often targets slower-moving attributes that affect capital allocation, cost of capital, financing access, and strategic valuation (Sezer et al., 2020; Gupta et al., 2020; Mai et al., 2019; Lin & Hsu, 2023).

The two lines are analytically distinct but operationally connected. Asset pricing and portfolio construction depend not only on return forecasts but also on information about firms' balance sheets, financing capacity, exposure to sustainability risks, and disclosure credibility. Likewise, ESG analysis is no longer confined to normative screening. It increasingly enters expected-return models, downside-risk analysis, and strategic asset allocation because investors now treat sustainability information as a component of risk transmission, resilience, and regulatory exposure (Friede et al., 2015; Berg et al., 2022; Christensen et al., 2022; Amel-Zadeh & Serafeim, 2018). From this perspective, AI does not merely predict prices; it also transforms corporate attributes into investable signals.

The first problem that pushed finance toward AI was the persistent mismatch between the statistical complexity of financial data and the simplifying assumptions of classical models. Financial time series are noisy, weakly nonstationary, heteroskedastic, heavy-tailed, and exposed to regime shifts. Linear models remain useful because of their transparency and parsimonious structure, but they often struggle when signal generation depends on nonlinear interactions, temporal memory, or multimodal data. Deep learning entered finance largely by promising richer function approximation for such environments, especially in tasks where historical prices, market microstructure information, or textual sentiment interact in nonlinear ways (Fischer & Krauss, 2018; Krauss et al., 2017; Bao et al., 2017). Yet the record is mixed. AI models can improve benchmark performance in specific samples, but they also amplify model risk, overfitting, and sensitivity to regime change. The implication is that better prediction is not synonymous with better decision-making.

The second problem is that finance is not only about predicting the next return. Most important decisions are constrained optimization problems. Investors must translate beliefs into weights under transaction costs, turnover limits, factor exposures, risk budgets, liquidity constraints, and governance requirements. Corporate finance researchers face a parallel challenge: they must translate noisy indicators into inferences about financing constraints, sustainability performance, and capital market access. In both contexts, AI sits upstream of the final decision rule. A forecasting model is therefore only one component of a longer chain that includes target definition, state representation, optimization, rebalancing, and ex post evaluation (Markowitz, 1952; Kolm et al., 2014; Ban et al., 2018).

A third reason to reconsider the field at a higher level is the rise of sustainable finance. The empirical literature now shows that ESG information affects investor demand, access to finance, cost of capital, bank lending conditions, firm risk, and crisis-period resilience, although the magnitude and direction of effects depend heavily on materiality, rating methodology, and institutional setting (Cheng et al., 2014; El Ghoul et al., 2011; Goss & Roberts, 2011; Albuquerque et al., 2019; Lins et al., 2017; Albuquerque et al., 2020). This shift has created a new role for AI. Algorithms are being used to predict ESG ratings, extract sustainability signals from disclosures, reconcile rating divergence, and identify heterogeneity across firms and markets. The practical significance is clear: once ESG moves from peripheral screening to core financial analysis, AI applications in finance must connect return prediction with firm-level sustainability inference.

The objective of this review is therefore not to juxtapose two unrelated papers on LSTM-based portfolio optimization and financing constraints with ESG ratings. Instead, it uses them as entry points into a larger analytical structure. The paper by Li and Liu illustrates a common AI-finance workflow: generate forecasts with a recurrent network, then feed predictions into portfolio optimization (Li & Liu, 2023). The paper by Liu illustrates the firm-side line: construct a financing-constraint indicator and estimate its association with ESG ratings in the Chinese stock market (Liu, 2022). Taken together, these papers point to a more general question: how should AI be used when financial decision-making requires both short-horizon market prediction and slower-moving inference about firm quality, sustainability, and financing frictions? These two studies are retained throughout the review as focal anchor papers: Li and Liu (2023) on LSTM-based portfolio optimization, and Liu (2022) on financing constraints and ESG ratings in the Chinese stock market.

The review proceeds in eight sections. Section 2 synthesizes forecasting methods, from ARIMA-style baselines to recurrent networks, transformers, and hybrid models. Section 3 discusses how predictions are translated into portfolio decisions, emphasizing the gap between predictive accuracy and out-of-sample allocation performance. Section 4 reviews the AI literature on corporate finance and sustainable finance, focusing on financial constraints, ESG ratings, and firm heterogeneity. Section 5 examines data sources and measurement issues, including market data, financial statements, ESG scores, texts, and alternative data. Section 6 compares predictive and explanatory models, especially under high-stakes financial decisions. Section 7 assesses limitations and risks, including overfitting, sample dependence, interpretability, and institutional context. Section 8 concludes by proposing a more integrated research agenda in which forecasting, allocation, and ESG-related firm analysis are jointly modeled.

Financial Time-Series Forecasting Methods: from ARIMA to LSTM, Transformers, and Hybrid Models

The forecasting literature in finance predates contemporary AI by decades, and any serious review must begin by recognizing why classical econometric models remain relevant. Linear autoregressive frameworks, ARIMA-type models, and volatility models endure because they encode explicit assumptions about persistence, differencing, and conditional variance. They provide interpretable baselines, strong diagnostics, and relatively low estimation variance. In noisy financial environments, these virtues are substantial. Many AI studies still benchmark against linear time-series models because the real question is not whether neural networks can fit complex functions in-sample, but whether they deliver decision-relevant gains after accounting for instability, transaction costs, and changing regimes (Sezer et al., 2020; Patel et al., 2015).

Nevertheless, the attraction of machine learning in forecasting is easy to understand. Financial variables often exhibit nonlinear relations, threshold effects, and interactions across prices, volume, sentiment, and macro signals. Early machine-learning forecasting studies therefore used support vector machines, feedforward neural networks, and tree-based methods to relax linear restrictions while preserving manageable training costs (Kara et al., 2011; Patel et al., 2015). These models often improved directional classification or short-horizon return prediction, especially when the problem was framed as movement classification rather than precise point forecasting. Yet their limitations were also clear. Static machine-learning models do not naturally represent long-range temporal dependence, and their performance deteriorates when hand-engineered features fail to capture evolving market structure.

Recurrent neural networks, and especially long short-term memory models, became influential because they addressed this temporal limitation. LSTM architectures introduced gated memory mechanisms that can, in principle, retain relevant information over longer sequences while filtering noise. In finance, this made them attractive for predicting prices, returns, and volatility from historical windows of market data (Fischer & Krauss, 2018; Bao et al., 2017). The main empirical appeal of LSTM is not that financial series become suddenly predictable, but that sequence models can absorb richer lag structures and nonlinear dependence than conventional regressions. The paper by Li and Liu fits squarely within this tradition: it uses LSTM-generated forecasts as inputs to portfolio choice, representing the widespread belief that better temporal representation can improve downstream allocation (Li & Liu, 2023).

Even here, however, caution is required. The strongest results in LSTM-based finance often arise in specific samples, universes, or forecast horizons. Sequence models can exploit artifacts of market microstructure, training-window selection, or benchmark construction. They also require careful choices about sequence length, normalization, retraining frequency, and label definition. The literature therefore shows both promise and fragility. Fischer and Krauss report that LSTM models can outperform traditional classifiers in cross-sectional stock selection, while Bao and colleagues show gains in time-series forecasting when LSTM is combined with stacked autoencoders (Fischer & Krauss, 2018; Bao et al., 2017). At the same time, review studies emphasize that performance heterogeneity across datasets is large, and apparent gains often shrink once costs, turnover, or different market conditions are considered (Sezer et al., 2020).

A second development concerns ensemble and hybrid models. Rather than treating a single algorithm as sufficient, many studies now combine predictive modules with optimization or feature-selection layers. Ma and colleagues integrate machine-learning and deep-learning return prediction with mean–variance and omega portfolio models, while Chen and colleagues combine an improved XGBoost prediction stage with downstream mean–variance selection (Ma et al., 2021; Chen et al., 2021). Ta and colleagues similarly use LSTM-based prediction as an upstream component in quantitative trading and portfolio formation (Ta et al., 2020). The logic behind such hybrids is straightforward. Prediction and decision are different tasks, and separating them may improve overall system performance. A dedicated prediction block can identify candidate assets or expected returns, while an optimization block imposes diversification and risk control.

Transformers introduced a new stage in this literature. Originally developed for natural language processing, they replaced recurrent computation with attention mechanisms that can model long-range dependence more efficiently and flexibly. Time-series variants such as Informer were designed specifically for long-sequence forecasting, which is especially appealing when financial prediction uses multivariate histories, irregular patterns, or multimodal inputs (Zhou et al., 2021). In theory, transformer models can identify relevant time points or features without compressing all temporal information into a fixed hidden state. In practice, they may outperform recurrent models when the series is sufficiently long and when cross-variable relations matter. Yet their financial advantage is still conditional. Transformers are data hungry, computationally costly, and harder to regularize in low signal-to-noise environments. Their theoretical flexibility may exceed the information content of the data they are given.

Another major extension of forecasting is the incorporation of text and sentiment. Financial prices do not move only because of lagged returns. News, earnings calls, annual reports, social media sentiment, and management disclosures all shape expectations. Tetlock's study on media content remains foundational because it shows that textual negativity helps explain market behavior beyond price history (Tetlock, 2007). Bollen and colleagues later extend this line to social media mood, while more recent reviews document a broad field of finance-oriented text mining (Bollen et al., 2011; Gupta et al., 2020). The methodological implication is that forecasting should be viewed as a representation problem. Market data provide one representation of state; text provides another. AI becomes valuable not merely through non-

linear fitting, but through the ability to combine heterogeneous data sources.

This multimodal logic is especially important for firm-level prediction tasks. Mai and colleagues demonstrate that textual disclosures improve bankruptcy prediction, suggesting that narrative information contains signals about corporate quality that structured ratios alone cannot capture (Mai et al., 2019). The same lesson applies to sustainability analysis. ESG controversies, governance practices, and financing frictions may be only partially observable in accounting variables but more fully reflected in disclosure patterns, tone, and unstructured corporate communication. Thus, the forecasting literature and the firm-feature literature are already converging at the level of data representation, even when they are still treated as separate subfields.

What, then, is the correct benchmark for forecasting success? A common mistake is to focus on statistical metrics in isolation. Lower root-mean-squared error or higher directional accuracy does not automatically improve decisions. Financial forecasts are useful only to the extent that they improve ranking, screening, timing, or hedging after costs and constraints. This is why the literature increasingly emphasizes economic evaluation rather than purely statistical evaluation. A model that slightly improves predictive loss but generates unstable weights and high turnover may be less useful than a simpler model with weaker raw accuracy but better decision stability (DeMiguel et al., 2009; Kolm et al., 2014). The same principle applies to corporate finance forecasting: better ESG-score prediction is valuable only if it improves inference about financing conditions, valuation, or sustainability risk.

A further issue is target definition. Many studies use next-day return prediction, but finance offers multiple possible targets: direction, abnormal return, volatility, drawdown, tail loss, distress probability, or ESG-score change. Different targets imply different loss functions and different notions of success. A model optimized for one-step return prediction may be irrelevant for long-horizon strategic allocation. Likewise, a model that predicts aggregate ESG scores may miss material subdimensions, while a financial-constraint model may work for manufacturing firms but fail for digital firms with intangible-heavy balance sheets. Target selection is therefore not a technical detail. It defines the decision problem the model is actually solving.

In summary, the evolution from ARIMA-style models to LSTM, transformers, and hybrids reflects a real broadening of representational capacity, but not a universal solution to financial forecasting. The strongest lesson of the literature is comparative rather than triumphalist. Classical econometric models remain indispensable as transparent baselines. Machine-learning models are useful when nonlinearities and feature interactions matter. Sequence models are useful when temporal structure is central. Transformer architectures are promising when long-range dependency and multimodal integration become important. Hybrid systems are attractive when the forecasting stage must be tightly linked to screening or optimization. The real frontier is therefore not one superior algorithm but the design of forecasting systems matched to the data, horizon, and decision rule under study (Sezer et al., 2020; Fischer & Krauss, 2018; Zhou et al., 2021).

From Prediction to Allocation: Portfolio Optimization, Risk–Return Trade-off, and Dynamic Rebalancing

Portfolio optimization is the point at which financial prediction becomes an actionable decision. This stage is theoretically older than AI, but AI has changed both its inputs and its structure. Modern portfolio theory, beginning with Markowitz, formalized asset allocation as a trade-off between expected return and variance (Markowitz, 1952). Sharpe's asset-pricing framework further linked expected returns to systematic risk, helping establish the canonical structure of risk-adjusted choice (Sharpe, 1964). Yet decades of empirical work have shown that the practical implementation of mean–variance optimization is plagued by estimation error. Small changes in expected returns or covariances can generate unstable, concentrated, and economically implausible weights. This is one reason why naive diversification can outperform optimized portfolios out of sample (DeMiguel et al., 2009).

AI enters this problem in two main ways. First, it can improve the estimates that feed an optimizer. Instead of using historical mean returns, one may use machine-learning forecasts. Instead of relying on sample covariances, one may use shrinkage estimators, latent factor models, or learned risk structures. Second, AI can modify the allocation rule itself. Rather than solving a static mean–variance problem, one can use reinforcement learning or online learning to choose sequential portfolio actions directly (Ban et al., 2018; Yang, 2023; Li & Hoi, 2014).

The literature on AI-assisted portfolio formation often begins by inserting forecasts into classical optimizers. Li and Liu represent a straightforward version of this approach: LSTM outputs are used within maximum-Sharpe and minimum-variance frameworks (Li & Liu, 2023). Ma and colleagues provide a broader comparison by combining return prediction from random forests, support vector regression, LSTM, and other models with mean–variance and omega optimization (Ma et al., 2021). Chen and colleagues build a two-stage system in which an XGBoost-based prediction block selects promising assets before a mean–variance allocator determines weights (Chen et al., 2021). Ta and colleagues similarly embed LSTM prediction into a portfolio construction workflow (Ta et al., 2020). These studies share a common assumption: that improved predictive information can be translated into superior portfolios if optimization controls risk.

This assumption is reasonable but incomplete. The central challenge is not only prediction quality but forecast usability. Portfolio construction is highly sensitive to forecast error. A small upward bias in expected returns can induce excessive concentration; a slight ranking error can push the optimizer toward high-turnover positions with weak economic value. For this reason, many practical systems use AI for ranking or preselection rather than exact return estimation. The machine-learning block filters the asset universe or produces ro-

bust relative scores, and the optimizer then imposes diversification constraints. Ban and colleagues formalize this connection by studying the interaction between machine learning and portfolio optimization more generally (Ban et al., 2018).

Risk estimation is equally important. Sample covariance matrices are notoriously unstable in high dimensions, especially when the number of assets is large relative to the length of history. Ledoit and Wolf's shrinkage estimators remain central because they improve conditioning and reduce estimation variance, which often matters more for realized performance than marginal improvements in expected-return forecasts (Ledoit & Wolf, 2004a; Ledoit & Wolf, 2004b). In practical terms, AI-based portfolio models that boast sophisticated return prediction but rely on naïve covariance estimation may still fail out of sample. This is one of the most underappreciated points in the recent literature: portfolio performance is a joint function of alpha estimation, risk estimation, and execution rules, not of predictive architecture alone (Kolm et al., 2014).

An alternative tradition shifts attention from mean–variance optimization to risk budgeting and diversification structures. Risk parity and equal risk contribution portfolios allocate capital by balancing risk contributions rather than maximizing forecast-adjusted utility. Maillard and colleagues analyze the properties of equally weighted risk contribution portfolios, while Roncalli and Weisang extend risk parity toward risk-factor formulations (Maillard et al., 2010; Roncalli & Weisang, 2016). López de Prado introduces hierarchical risk parity, which uses clustering and recursive bisection to avoid some of the instability associated with inverting noisy covariance matrices (López de Prado, 2016). These approaches are highly relevant to AI because they show that richer prediction is not the only route to better portfolios. Sometimes the key gain comes from more robust structural diversification rather than more aggressive expected-return estimation.

Online and reinforcement-learning approaches go further by replacing the static optimizer with a sequential decision process. In online portfolio selection, the algorithm updates weights iteratively as new information arrives. Li and Hoi survey this tradition and show how follow-the-winner, follow-the-loser, pattern-matching, and meta-learning strategies can all be interpreted within a unified sequential decision framework (Li & Hoi, 2014). Li and colleagues' moving average reversion strategy exemplifies the use of online learning ideas in allocation, where the core intuition is that mean reversion can be exploited over multiple periods without explicitly forecasting full return distributions (Li et al., 2015).

Deep reinforcement learning extends this logic by treating portfolio management as a dynamic control problem. The agent learns a policy that maps the current state to portfolio weights while optimizing long-run reward, usually some function of return and risk. Deng and colleagues provide an influential early example of deep direct reinforcement learning for trading signals and allocation (Deng et al., 2017). More recent work by Yang develops a task-context mutual actor–critic framework for portfolio management, and Guan and Liu focus on explaining the behavior of deep reinforcement-learning agents in the portfolio setting (Yang, 2023; Guan & Liu, 2021). These studies are important because they relax the separation between prediction and optimization. The agent does not first predict returns and then solve a portfolio problem; it learns allocation actions directly from reward feedback.

This direct approach has several attractions. It can incorporate transaction costs, delayed rewards, and path dependence more naturally than one-shot optimization. It also aligns well with dynamic rebalancing, where the value of a trade depends on future states, not merely on current expected return. Yet reinforcement learning also amplifies some of finance's hardest identification problems. Reward functions may be misspecified, exploration is difficult in nonstationary markets, and learned strategies may overfit historical dynamics that are not repeatable. Performance can look impressive in backtests while depending heavily on environment design, feature engineering, or training-window luck. For this reason, reinforcement learning in finance remains promising but methodologically demanding rather than mature.

A critical issue throughout this literature is the difference between statistical forecasting gains and realized utility gains. A model may improve prediction in a regression sense yet fail to improve investor welfare because of turnover, estimation error, concentration, or slippage. Conversely, a model with mediocre point forecasts may still improve portfolio decisions by ranking assets better, stabilizing exposures, or avoiding extreme losses. This is why evaluation should move beyond raw predictive metrics toward realized Sharpe ratios, drawdowns, turnover-adjusted returns, tail risk, and robustness across regimes (Kolm et al., 2014; DeMiguel et al., 2009). The most meaningful contribution of AI to portfolio optimization is therefore not simply better forecasting, but better integration of signal extraction with robust decision rules.

The LSTM-based portfolio paper included in the present review is useful precisely because it illustrates both the promise and the limits of the prediction-to-allocation pipeline (Li & Liu, 2023). The promise lies in combining nonlinear sequence learning with standard portfolio rules. The limit is that the framework still depends on assumptions about the stability of forecast quality, the appropriateness of maximum-Sharpe or minimum-variance objectives, and the transferability of sample-specific gains. Similar comments apply to many recent portfolio studies. The field is progressing, but the key bottleneck remains economic robustness rather than architectural novelty.

The strongest current direction is likely a layered design. In such a design, prediction models generate robust expected-return or ranking signals; risk models provide stabilized covariance or factor estimates; allocation modules impose diversification, turnover, and ESG constraints; and rebalancing policies are learned or calibrated under explicit cost assumptions. This view is more realistic than the search for a single end-to-end algorithm. It also connects naturally to sustainable finance, because ESG characteristics can enter the allocation layer as constraints, penalties, or state variables rather than as a separate normative overlay (Friede et al., 2015; Berg et al., 2022).

Corporate Finance and Sustainability Analysis

If market forecasting is the first main line of AI in finance, firm-level feature identification is the second. This line includes bankruptcy prediction, credit risk, disclosure analysis, financing constraints, governance assessment, and ESG-score prediction. Its central task is to infer economically meaningful firm characteristics from large, noisy, and heterogeneous data. In recent years, sustainable finance has moved this line to the center of the field because ESG performance is now linked to capital allocation, risk management, and valuation rather than only to ethical screening (Friede et al., 2015; Amel-Zadeh & Serafeim, 2018).

The empirical case for the financial relevance of ESG is now extensive but not uniform. Meta-evidence indicates that a large share of studies report a nonnegative association between ESG and financial performance, though effect sizes vary by design, period, and outcome measure (Friede et al., 2015). More specific studies show that corporate social responsibility is associated with improved access to finance, lower costs of capital, and lower loan spreads under some conditions (Cheng et al., 2014; El Ghoul et al., 2011; Goss & Roberts, 2011). ESG-related behavior also appears linked to lower firm risk and stronger resilience during crisis episodes (Albuquerque et al., 2019; Lins et al., 2017; Albuquerque et al., 2020). These findings have made ESG information financially consequential even for investors without explicit social mandates.

At the same time, the ESG literature also highlights deep measurement problems. ESG ratings diverge substantially across providers because they differ in scope, indicators, weights, and treatment of controversies (Berg et al., 2022; Christensen et al., 2022). This creates an immediate opportunity for AI but also a methodological warning. Machine-learning models can predict provider-specific scores with reasonable accuracy, as demonstrated in studies such as Lin and Hsu, but high predictive accuracy does not solve the underlying disagreement about what is being measured (Lin & Hsu, 2023). In other words, AI can reproduce rating systems more easily than it can resolve their conceptual inconsistency.

This tension is especially important when ESG analysis is linked to corporate finance. The paper by Liu examines whether financing constraints affect firms' ESG ratings in the Chinese stock market (Liu, 2022). Substantively, this is a plausible question. Firms facing tight financing conditions may underinvest in environmental compliance, social programs, governance improvement, or disclosure quality because they prioritize liquidity preservation and short-run survival. Alternatively, constrained firms may use ESG signaling strategically to reduce financing frictions. The relationship is therefore theoretically open rather than obvious. What matters for the present review is that the paper sits at the intersection of two literatures that are often studied separately: financial-constraint measurement and ESG outcome assessment.

The measurement of financing constraints has long been controversial. Kaplan and Zingales famously argue that investment-cash flow sensitivity is not a valid general proxy for financing constraints, opening a broader debate about construct validity (Kaplan & Zingales, 2000). Almeida and colleagues focus on the cash-flow sensitivity of cash, Foley and coauthors show how cash holdings are shaped by tax and internal-liquidity considerations, and Whited and Wu propose a structural index of financing constraints risk (Almeida et al., 2004; Foley et al., 2007; Whited & Wu, 2006). Hadlock and Pierce develop the SA index to move beyond the limitations of the KZ framework, and Farre-Mensa and Ljungqvist question whether commonly used measures actually capture financing constraints in the intended sense (Hadlock & Pierce, 2010; Farre-Mensa & Ljungqvist, 2016). This literature matters because AI studies that use financing-constraint labels or indices inherit these construct ambiguities. A high-performing model trained on a weak proxy does not produce a strong economic inference.

Once this point is recognized, the contribution of AI to corporate finance becomes clearer. Its value lies less in replacing theory-driven proxies than in combining multiple sources of information to improve latent-variable inference. Financial constraints, governance quality, and sustainability orientation are not directly observed. They are inferred from accounting variables, market outcomes, disclosures, financing structure, ownership patterns, and sometimes textual cues. AI can integrate these heterogeneous sources more flexibly than traditional linear models. But the inference remains only as credible as the conceptual mapping between the latent construct and the observed proxies.

Firm heterogeneity is another reason why AI has become important. ESG and financing dynamics are unlikely to be homogeneous across industries, ownership structures, regulatory environments, and life-cycle stages. Material ESG issues differ across sectors, which is precisely why materiality-based analyses often perform better than generic aggregate scoring (Khan et al., 2016). Constrained firms in asset-heavy industries may face different sustainability trade-offs from digital firms with intangible capital. Firms in bank-dominated financial systems may respond differently from firms in market-based systems. Emerging markets add further layers involving disclosure quality, state involvement, and policy intervention. AI methods are attractive in this setting because they can detect interactions and nonlinear heterogeneity that are cumbersome in conventional specification search.

Textual and disclosure analysis intensify this potential. Annual reports, management discussion sections, sustainability reports, and earnings calls contain information about governance quality, risk management, strategic orientation, and perhaps greenwashing. Earlier studies show that textual disclosure helps distress prediction (Mai et al., 2019), and broader reviews confirm that finance text mining has become a substantial field (Gupta et al., 2020). In sustainable finance, this means AI can be used not only to forecast a rating but also to identify the textual pathways through which a firm signals sustainability commitment or obscures risk. This is particularly important when rating disagreement is high, because textual analysis can uncover which dimensions drive a specific model's prediction.

The rise of ESG-score prediction studies illustrates both the opportunities and the limits of this approach. Lin and Hsu show that machine-learning models can predict ESG scores for Taiwanese nonfinancial companies using structured firm information (Lin & Hsu, 2023). Such work is useful for three reasons. First, it helps identify the variables most associated with provider-specific ESG assessments. Second, it can extend coverage to firms or markets with sparse ratings. Third, it can be used as a screening or monitoring tool when ratings are missing or delayed. But its limitation is equally clear: a predicted ESG score is still conditional on the training label. If the original label is noisy, divergent, or conceptually narrow, prediction quality does not guarantee substantive validity.

A related issue concerns causality. Much of the literature, including work on CSR and access to finance or ESG and cost of capital, relies on observational designs (Cheng et al., 2014; El Ghoul et al., 2011; Goss & Roberts, 2011). AI can improve prediction within such data, but it does not by itself identify causal effects. This distinction is especially important in corporate finance. If constrained firms have lower ESG ratings, is that because financing pressure reduces sustainability investment, because poor ESG raises financing costs, or because both reflect deeper firm quality? Prediction-focused AI models often blur these alternatives. For research aimed at explanation rather than screening, model design must therefore be tied to identification strategy, institutional context, and temporal ordering.

Despite these cautions, AI has real advantages in ESG-related corporate finance analysis. It can capture multidimensional firm states, integrate text with accounting data, identify heterogeneous effects, and improve early-warning systems. It is particularly valuable when the research question is classification or ranking under large-scale, noisy information. It is less decisive when the question is structural explanation under contested constructs. The most defensible position is that AI should augment, not replace, theory-based corporate finance analysis. In the case of financing constraints and ESG, this means using AI to discover patterns, improve measurement, and test heterogeneity, while retaining careful attention to what the chosen labels and proxies actually mean (Hadlock & Pierce, 2010; Farre-Mensa & Ljungqvist, 2016; Berg et al., 2022).

In this sense, the corporate finance line and the portfolio line are converging. Investors increasingly use firm-level sustainability and financing information in cross-sectional allocation, risk control, and engagement strategies. Corporate finance researchers increasingly use machine learning to model disclosure-based or rating-based firm characteristics. The next step is to connect these streams more explicitly. A firm's financing constraints can influence its ESG investments and disclosure quality; these, in turn, can affect its cost of capital, market risk, and inclusion in investor portfolios. AI is well placed to model such linked systems, but only if it moves beyond narrow single-task prediction.

Data and Indicators

Any evaluation of AI in financial decision-making must take data architecture as seriously as model architecture. Many published comparisons between algorithms are in fact comparisons between data pipelines, label definitions, and preprocessing choices. Finance is a domain in which small differences in frequency, timing, survivorship treatment, or universe selection can materially alter conclusions. This is true for both market forecasting and firm-level sustainability analysis.

The most conventional input set remains market data: prices, returns, volume, volatility measures, order-flow proxies, and sometimes factor returns. These variables are attractive because they are high frequency, standardized, and directly tradable. They are also limited because they are close to the outcome being predicted, which raises the danger of learning transient autocorrelation or microstructure artifacts rather than stable economic relationships. For this reason, classical baselines remain important, and sequence models must be evaluated under carefully defined rolling or expanding windows rather than randomly shuffled samples (Sezer et al., 2020; Fischer & Krauss, 2018).

Financial statements and accounting ratios form the second major data family. These are central in corporate finance, bankruptcy prediction, ESG estimation, and cross-sectional stock selection. They are slower moving than prices but often more interpretable. Variables related to leverage, profitability, liquidity, asset growth, cash holdings, and investment are widely used in both firm-level and asset-pricing studies. The financing-constraint literature, however, shows that even apparently straightforward accounting variables can encode contested constructs (Kaplan & Zingales, 2000; Hadlock & Pierce, 2010). AI models trained on such data therefore require conceptual discipline: users must know whether the model is predicting a realized outcome, a provider score, or a theoretical latent attribute.

ESG data constitute a third family and bring distinct problems. ESG ratings are multidimensional, provider specific, and often incomplete outside large firms or developed markets. They may be available as aggregate scores, pillar scores, or issue-level indicators. Their main practical attraction is that they operationalize sustainability in investable form. Their main analytical weakness is cross-provider divergence (Berg et al., 2022; Christensen et al., 2022). Researchers using ESG data must therefore specify whether they are interested in predicting a given provider's score, identifying material sustainability exposure, or inferring an underlying latent construct that transcends rating systems. These are not equivalent tasks.

Textual data have become indispensable because many financially relevant signals are disclosed narratively rather than numerically. News articles, earnings-call transcripts, annual reports, sustainability reports, analyst commentary, and social media all supply information about expectations, risk perception, and organizational quality (Tetlock, 2007; Bollen et al., 2011; Gupta et al., 2020). Text is particularly important when structured data are lagged or sparse. For example, ESG controversies may surface in news before they enter provider

ratings. Governance problems may appear in tone, emphasis, or omission before they are visible in standard ratios. Likewise, distress risk can often be inferred from managerial language and disclosure complexity (Mai et al., 2019). AI has comparative advantage here because modern language models and document embeddings can convert such unstructured information into tractable features.

Alternative data broaden the field further. Satellite imagery, web traffic, geolocation data, job postings, patent text, supply-chain records, and search intensity have all been used in parts of the broader finance literature, though not always in the core papers reviewed here. The relevance of alternative data lies not only in novelty but in temporal lead. These sources can capture operational or reputational changes before they appear in statements or prices. In ESG-related analysis, alternative data may help monitor physical risk exposure, environmental incidents, labor conditions, or public scrutiny. But they also raise acute concerns about access inequality, reproducibility, and legal or ethical boundaries.

Data frequency and alignment are another underappreciated issue. Market prediction often uses daily or intraday observations, whereas ESG and corporate finance analysis typically use quarterly or annual data. Integrating these two levels requires careful temporal design. A portfolio model that uses annual ESG scores for daily rebalancing may implicitly assume that sustainability signals change more often than they do. Conversely, a firm-level study that matches financing data to contemporaneous ESG scores may obscure lag structures. The integrated perspective proposed in this review therefore requires multi-frequency modeling. Short-horizon market states and long-horizon firm attributes should not be forced into the same temporal resolution without justification.

Label construction can be even more consequential than the raw inputs. In forecasting, researchers must decide whether the label is raw return, excess return, sign, volatility, abnormal performance, or a portfolio-relative ranking. In corporate finance, labels may be distress events, rating categories, financing-constraint indices, or sustainability scores. These labels are not neutral. They reflect theoretical choices about what matters. The example of financial constraints is especially instructive: the KZ, WW, and SA frameworks are not interchangeable, and studies have shown that common measures can disagree materially (Whited & Wu, 2006; Hadlock & Pierce, 2010; Farre-Mensa & Ljungqvist, 2016). A model's apparent performance can therefore change not because the algorithm is better, but because the label is easier to predict.

Preprocessing is similarly decisive. Normalization, winsorization, lagging rules, handling of missing values, treatment of delisted firms, and prevention of look-ahead bias all matter. In text analysis, tokenization, vocabulary restriction, document truncation, and the distinction between domain-adapted and general language models affect results. In ESG work, whether controversies are treated as level variables or event shocks can alter interpretation. Thus, one of the persistent problems in AI-finance comparisons is that published studies often compare end results without making preprocessing decisions sufficiently explicit.

The policy implication is that data governance should be treated as part of the research design, not as a preliminary technical step. This is especially important when AI systems move from academic experimentation to real decision support. A forecasting system for portfolio management needs clear rules about data vintage, revision handling, universe construction, and cost assumptions. An ESG-prediction system needs clear rules about provider selection, missingness, coverage bias, and update frequency. Without such governance, model performance is difficult to interpret and almost impossible to reproduce.

For integrated AI-based financial decision-making, the most promising data architecture is layered. High-frequency market data should capture near-term state changes; firm-level structured data should capture slower-moving financial conditions; text should capture narrative and disclosure-based signals; ESG data should represent sustainability-related exposures and assessments; and alternative data should be added only when their timing, legality, and economic meaning are clear. This layered architecture is what allows the two main lines of finance AI—market prediction and firm-level identification—to be analyzed jointly rather than separately.

Method Comparison

A recurring source of confusion in AI-finance research is the failure to distinguish between predictive and explanatory aims. Predictive models are judged by out-of-sample accuracy, ranking power, economic utility, or decision performance. Explanatory models are judged by interpretability, identification, construct validity, and the plausibility of causal or structural claims. These goals overlap, but they are not identical. In finance, the difference is especially consequential because many tasks involve high-stakes decisions under legal, fiduciary, or regulatory constraints.

Predictive models dominate the market-forecasting literature. Their objective is straightforward: improve forecasts that can be translated into trading, hedging, or allocation. In this setting, model complexity is acceptable if it improves stable out-of-sample performance. LSTM, transformer, and reinforcement-learning systems are therefore often justified even when they are opaque, because the immediate benchmark is economic performance rather than interpretive clarity (Fischer & Krauss, 2018; Zhou et al., 2021; Yang, 2023). But the justification weakens when the model is deployed in institutional contexts that require explanation, such as fiduciary asset management, regulated advice, or risk governance.

Explanatory models dominate much of corporate finance and sustainable finance. When the research question concerns whether financing constraints affect ESG outcomes, whether CSR reduces the cost of capital, or why ESG ratings diverge, the aim is not merely prediction. It is to infer a mechanism, a channel, or a structural relation (Cheng et al., 2014; El Ghoul et al., 2011; Berg et al., 2022). In these contexts, black-box predictive success is insufficient. A model may predict ESG scores well while obscuring whether it is relying on size, industry, disclosure volume, or genuinely material sus-

tainability information. It may identify correlations between financing variables and ESG scores without clarifying temporal ordering or omitted variables.

This is why explainability has become central in recent discussions of finance AI. Rudin argues that in high-stakes settings, the field should prefer inherently interpretable models where possible rather than rely on post hoc explanations for black boxes (Rudin, 2019). Arrieta and colleagues provide a broader framework for explainable AI, distinguishing among concepts, audiences, and explanatory techniques (Arrieta et al., 2020). In finance specifically, bibliometric evidence shows rapid growth in explainable AI research, indicating that the field increasingly recognizes opacity as a practical constraint rather than a tolerable side effect (Chen et al., 2023).

The trade-off, however, should not be oversimplified. Interpretable models are not always superior, and black-box models are not always irresponsible. The relevant question is whether the model's complexity is proportionate to the task and whether the decision environment requires local, global, or procedural explanation. For example, a high-frequency portfolio signal used internally within a diversified multi-signal platform may tolerate lower interpretability than a model used to deny credit, assign a sustainability score, or justify fiduciary stewardship actions. In other words, the appropriate level of explanation depends on the financial context.

Portfolio management provides a useful middle case. Investors often care less about the internal semantics of a signal than about stable realized performance and controlled exposures. Yet as reinforcement learning and end-to-end systems become more common, explainability becomes necessary for understanding why the strategy loads on particular market states, factors, or asset clusters. Guan and Liu's work on explainable deep reinforcement learning is important precisely because it tries to connect policy behavior to interpretable reference weights and prediction power (Guan & Liu, 2021). This kind of work suggests that finance does not have to choose between prediction and explanation in absolute terms. Instead, it can design systems in which opaque components are surrounded by interpretable diagnostic layers.

In ESG-related corporate finance, the need for explanation is even stronger. ESG ratings affect investor perception, capital allocation, and in some settings regulatory classification. If a machine-learning model predicts ESG scores, users must know whether it is learning material sustainability characteristics or merely proxying for firm size, international visibility, or disclosure capacity. This is not only a methodological problem; it is also a fairness and accountability problem. Divergence across rating providers already reveals that scoring systems embody implicit judgments (Berg et al., 2022; Christensen et al., 2022). Adding AI to this environment without interpretive discipline risks amplifying rather than reducing opacity.

A second comparison concerns causal versus associational modeling. Predictive models are usually associational. They learn patterns that improve forecasts, regardless of whether those patterns are causal. Explanatory finance research often seeks something stronger. Yet AI methods are frequently imported into explanatory settings without corresponding identification strategies. The result can be technically impressive but substantively ambiguous studies. For example, a model may show that certain accounting and market variables predict financing-constraint categories or ESG scores, but this does not establish the causal role of those variables. Therefore, when the objective is explanation, AI should be paired with research designs that address endogeneity, timing, and heterogeneity rather than treated as a substitute for them.

A third comparison concerns evaluation metrics. Predictive models are often judged by error, AUC, directional accuracy, or portfolio return statistics. Explanatory models require additional criteria: variable stability, sign consistency, theoretical coherence, and robustness across specifications. In some cases, a simpler model with slightly weaker predictive power may be preferable because it yields a more credible substantive interpretation. This is especially true in sustainable finance, where the policy and governance implications of model outputs can be large.

The practical conclusion is that finance researchers should specify model purpose before selecting model class. If the goal is high-frequency ranking under dense data and low explanation requirements, complex predictive models may be appropriate. If the goal is firm-level inference under contested constructs and governance relevance, interpretable or hybrid models are often more defensible. The broad theme of this review is not that one class should dominate the other, but that AI in finance should be evaluated as part of a decision system with explicit epistemic goals.

Limitations and Risks

The literature reviewed above shows substantial progress, but it also reveals recurring limitations that are too often treated as technical inconveniences rather than central research problems. Four risks are especially important: overfitting, out-of-sample instability, interpretability deficits, and institutional dependence.

Overfitting remains the most obvious challenge. Financial data have low signal-to-noise ratios, regime shifts, structural breaks, and adaptive agents. These characteristics make flexible models vulnerable to learning patterns that do not survive outside the training sample. Deep architectures can fit nonlinear relations extremely well, but finance does not reward in-sample fit. It rewards robust, repeatable economic value after costs and under changing conditions (Sezer et al., 2020; Krauss et al., 2017). The risk is particularly acute when researchers tune architectures extensively on a fixed historical window or report the best-performing specification without fully accounting for search. Apparent innovation may then reflect selection bias rather than genuine predictive improvement.

Out-of-sample failure is closely related but deserves separate emphasis. Even when a model is not overfit in a narrow statistical sense, it may still fail because the environment changes. Financial time series are non-stationary not only in mean and variance but also in institutional structure, liquidity, regulation, and investor behavior. A model trained in a

low-rate environment may break in a tightening cycle. A sentiment model trained on one media ecosystem may weaken after platform changes. An ESG model trained on a stable rating methodology may lose validity after provider revisions. Thus, genuine robustness requires temporal, cross-market, and cross-regime testing rather than a single train-test split (DeMiguel et al., 2009; Kolm et al., 2014).

Interpretability deficits are the third risk. In portfolio applications, opacity can conceal unstable exposures, hidden leverage to macro conditions, or spurious dependence on a narrow set of features. In corporate finance and ESG analysis, opacity is more serious because model outputs may influence financing decisions, stewardship, screening, or public claims about sustainability quality. Post hoc explanation methods can help, but they do not always recover the actual reasoning process of a complex model (Rudin, 2019; Arrieta et al., 2020). This is why interpretable design should be considered at the start of model development rather than appended at the end as a compliance layer.

Institutional dependence is the fourth and perhaps least discussed limitation. Financial models are embedded in specific market structures, accounting systems, disclosure regimes, investor bases, and regulatory frameworks. A result obtained in the U.S. equity market may not generalize to China's stock market, and a financing-constraint proxy validated in one setting may behave differently in another. ESG is especially sensitive to institutional context because reporting standards, enforcement, ownership structures, and policy priorities vary widely. The paper on financing constraints and ESG in the Chinese stock market is therefore not just a finance application; it is an example of how institutional setting shapes both the data and the substantive interpretation (Liu, 2022).

These limitations also interact. A model can be overfit partly because the institutional regime represented in the training data is unusually stable. A model can appear interpretable but still lack validity because its features proxy for country-specific disclosure habits rather than economically meaningful behavior. A model can show out-of-sample success in one market and fail in another because the target variable itself is constructed differently. Therefore, robustness should be treated multidimensionally: across time, across datasets, across institutions, and across target definitions.

A related risk is benchmark misselection. Many finance AI papers benchmark against weak baselines, such as poorly specified linear models or naïve optimizers, making gains look larger than they are. Strong baselines should include transparent econometric models, robust covariance estimators, shrinkage techniques, naïve diversification, and where relevant simple textual models. Without such baselines, the incremental value of AI is difficult to judge. This point is especially relevant in portfolio optimization, where historical experience shows that complicated expected-return estimates often underperform simpler diversification rules (DeMiguel et al., 2009; Ledoit & Wolf, 2004b).

Another limitation concerns reproducibility. Proprietary data, inconsistent preprocessing, and incomplete reporting of hyper parameters remain common. This problem is acute with alternative data and some ESG datasets. Reproducibility is not only a norm of good science; it is a practical necessity in finance because small implementation details can materially affect outcomes. The field would benefit from more standardized evaluation protocols, clearer reporting of data vintages, and stronger separation between model development, validation, and final testing.

Finally, there is a conceptual risk in treating AI as neutral infrastructure. Models always embed objectives. In forecasting, the objective may prioritize short-horizon excess return. In portfolio management, it may prioritize Sharpe ratio, downside risk, or turnover efficiency. In ESG analysis, it may prioritize prediction of existing scores rather than independent assessment of material sustainability performance. These objective choices shape results. They determine which signals are learned and which trade-offs are ignored. Therefore, one of the most important future tasks is to make objective functions explicit and align them with the intended decision context.

Conclusion and Future Directions

This review has argued that AI in financial decision-making should be understood through a broader architecture rather than through isolated technical applications. The first core line is market prediction, where models attempt to extract tradable signals from prices, returns, volatility, and text. The second core line is firm-level feature identification, where models infer financing constraints, ESG characteristics, distress risk, and other latent corporate attributes. These lines have evolved separately, but they increasingly interact in real financial decisions.

The literature on forecasting shows that the movement from classical econometric models to LSTM, transformers, and hybrid systems has expanded representational power, but not eliminated the basic problems of weak signals and unstable regimes (Sezer et al., 2020; Fischer & Krauss, 2018; Zhou et al., 2021). The portfolio literature shows that predictive gains matter only when they survive the translation into robust allocation under noisy covariance estimates, transaction costs, and rebalancing constraints (Markowitz, 1952; Kolm et al., 2014; Ban et al., 2018). The sustainable finance literature shows that ESG and financing information now affect access to capital, cost of capital, risk, and resilience, but also suffer from substantial measurement disagreement and institutional heterogeneity (Cheng et al., 2014; Berg et al., 2022; Christensen et al., 2022).

The two focal papers included by the user are therefore best interpreted not as isolated cases, but as representative nodes in this larger architecture. The LSTM-based portfolio paper captures the prediction-to-allocation logic that defines much of quantitative asset management (Li & Liu, 2023). The financing-constraint-and-ESG paper captures the firm-side logic in which financial frictions and sustainability outcomes are jointly analyzed (Liu, 2022). The real research opportunity lies in connecting these domains. Future studies should model how firm-level sustainability and financing signals enter cross-sectional expected returns, downside risk, portfolio

constraints, and dynamic rebalancing. They should also examine how investor allocation mechanisms feed back into firms' financing conditions and ESG incentives.

Methodologically, future work should pursue three directions. First, it should build multi-layer models that combine forecasting, risk estimation, and optimization rather than overemphasizing any single stage. Second, it should incorporate explainability more deeply in corporate finance and sustainable finance settings where model outputs influence high-stakes decisions (Rudin, 2019; Arrieta et al., 2020). Third, it should move toward multi-frequency and multi-modal designs that integrate market data, financial statements, ESG indicators, and text under clear temporal logic.

In short, the next stage of AI in finance is not simply more complex forecasting. It is the construction of integrated, accountable, and decision-relevant systems in which market prediction, portfolio design, and ESG-related corporate finance analysis are treated as interdependent components of the same financial decision process.

References

1. Ahmed, S., Alshater, M. M., El Ammari, A., & Hammami, H. (2022). Artificial intelligence and machine learning in finance: A bibliometric review. Research in International Business and Finance, 61, 101646. https://doi.org/10.1016/j.ribaf.2022.101646
2. Albuquerque, R., Koskinen, Y., & Zhang, C. (2019). Corporate social responsibility and firm risk: Theory and empirical evidence. Management Science, 65(10), 4451–4469. https://doi.org/10.1287/mnsc.2018.3043
3. Albuquerque, R., Koskinen, Y., Yang, S., & Zhang, C. (2020). Resiliency of environmental and social stocks: An analysis of the exogenous COVID-19 market crash. The Review of Corporate Finance Studies, 9(3), 593–621. https://doi.org/10.1093/rcfs/cfaa011
4. Almeida, H., Campello, M., & Weisbach, M. S. (2004). The cash flow sensitivity of cash. The Journal of Finance, 59(4), 1777–1804. https://doi.org/10.1111/j.1540-6261.2004.00679.x
5. Amel-Zadeh, A., & Serafeim, G. (2018). Why and how investors use ESG information: Evidence from a global survey. Financial Analysts Journal, 74(3), 87–103. https://doi.org/10.2469/faj.v74.n3.2
6. Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012
7. Ban, G.-Y., El Karoui, N., & Lim, A. E. B. (2018). Machine learning and portfolio optimization. Management Science, 64(3), 1136–1154. https://doi.org/10.1287/mnsc.2016.2644
8. Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. PLOS ONE, 12(7), e0180944. https://doi.org/10.1371/journal.pone.0180944
9. Berg, F., Kölbel, J. F., & Rigobon, R. (2022). Aggregate confusion: The divergence of ESG ratings. Review of Finance, 26(6), 1315–1344. https://doi.org/10.1093/rof/rfac033
10. Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007
11. Chen, W., Zhang, H., Mehlawat, M. K., & Jia, L. (2021). Mean–variance portfolio optimization using machine learning-based stock price prediction. Applied Soft Computing, 100, 106943. https://doi.org/10.1016/j.asoc.2020.106943
12. Chen, X.-Q., Ma, C.-Q., Ren, Y.-S., Lei, Y.-T., Huynh, N. Q. A., & Narayan, S. (2023). Explainable artificial intelligence in finance: A bibliometric review. Finance Research Letters, 56, 104145. https://doi.org/10.1016/j.frl.2023.104145
13. Cheng, B., Ioannou, I., & Serafeim, G. (2014). Corporate social responsibility and access to finance. Strategic Management Journal, 35(1), 1–23. https://doi.org/10.1002/smj.2131
14. Christensen, D. M., Serafeim, G., & Sikochi, A. (2022). Why is corporate virtue in the eye of the beholder? The case of ESG ratings. The Accounting Review, 97(1), 147–175. https://doi.org/10.2308/TAR-2019-0506
15. DeMiguel, V., Garlappi, L., & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? The Review of Financial Studies, 22(5), 1915–1953. https://doi.org/10.1093/rfs/hhm075
16. Deng, Y., Bao, F., Kong, Y., Ren, Z., & Dai, Q. (2017). Deep direct reinforcement learning for financial signal representation and trading. IEEE Transactions on Neural Networks and Learning Systems, 28(3), 653–664. https://doi.org/10.1109/TNNLS.2016.2522401
17. El Ghoul, S., Guedhami, O., Kwok, C. C. Y., & Mishra, D. R. (2011). Does corporate social responsibility affect the cost of capital? Journal of Banking & Finance, 35(9), 2388–2406. https://doi.org/10.1016/j.jbankfin.2011.02.007
18. Farre-Mensa, J., & Ljungqvist, A. (2016). Do measures of financial constraints measure financial constraints? The Review of Financial Studies, 29(2), 271–308. https://doi.org/10.1093/rfs/hhv052
19. Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654–669. https://doi.org/10.1016/j.ejor.2017.11.054
20. Foley, C. F., Hartzell, J. C., Titman, S., & Twite, G. (2007). Why do firms hold so much cash? A tax-based explanation. Journal of Financial Economics, 86(3), 579–607. https://doi.org/10.1016/j.jfineco.2006.11.006
21. Friede, G., Busch, T., & Bassen, A. (2015). ESG and financial performance: Aggregated evidence from more than 2000 empirical studies. Journal of Sustainable Finance & Investment, 5(4), 210–233. https://doi.org/10.1080/20430795.2015.1118917
22. Goodell, J. W., Kumar, S., Lim, W. M., & Pattnaik, D. (2021). Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis. Journal of Behavioral and Experimental Finance, 32, 100577. https://doi.org/10.1016/j.jbef.2021.100577
23. Goss, A., & Roberts, G. S. (2011). The impact of corporate social responsibility on the cost of bank loans. Journal of Banking & Finance, 35(7), 1794–1810. https://doi.org/10.1016/j.jbankfin.2010.12.002
24. Guan, M., & Liu, X.-Y. (2021). Explainable deep reinforcement learning for portfolio management: An empirical approach. In Proceedings of the Second ACM International Conference on AI in Finance (pp. 50:1–50:9). ACM. https://doi.org/10.1145/3490354.3494415
25. Gupta, R., Chen, M. Y., Hardle, W. K., Lee, T. M., & Mamaysky, H. (2020). A comprehensive review of text-mining applications in finance. Financial Innovation, 6, 49. https://doi.org/10.1186/s40854-020-00205-1
26. Hadlock, C. J., & Pierce, J. R. (2010). New evidence on measuring financial constraints: Moving beyond the KZ index. The Review of Financial Studies, 23(5), 1909–1940. https://doi.org/10.1093/rfs/hhq009

27. Kaplan, S. N., & Zingales, L. (2000). Investment-cash flow sensitivities are not valid measures of financing constraints. The Quarterly Journal of Economics, 115(2), 707–712. https://doi.org/10.1162/003355300554782
28. Kara, Y., Boyacioglu, M. A., & Baykan, Ö. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. Expert Systems with Applications, 38(5), 5311–5319. https://doi.org/10.1016/j.eswa.2010.10.027
29. Khan, M., Serafeim, G., & Yoon, A. (2016). Corporate sustainability: First evidence on materiality. The Accounting Review, 91(6), 1697–1724. https://doi.org/10.2308/accr-51383
30. Kolm, P. N., Tütüncü, R., & Fabozzi, F. J. (2014). 60 years of portfolio optimization: Practical challenges and current trends. European Journal of Operational Research, 234(2), 356–371. https://doi.org/10.1016/j.ejor.2013.10.060
31. Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. European Journal of Operational Research, 259(2), 689–702. https://doi.org/10.1016/j.ejor.2016.10.031
32. Ledoit, O., & Wolf, M. (2004a). A well-conditioned estimator for large-dimensional covariance matrices. Journal of Multivariate Analysis, 88(2), 365–411. https://doi.org/10.1016/S0047-259X(03)00096-4
33. Ledoit, O., & Wolf, M. (2004b). Honey, I shrunk the sample covariance matrix. The Journal of Portfolio Management, 30(4), 110–119. https://doi.org/10.3905/jpm.2004.110
34. Li, B., & Hoi, S. C. H. (2014). Online portfolio selection: A survey. ACM Computing Surveys, 46(3), Article 35. https://doi.org/10.1145/2512962
35. Li, B., Hoi, S. C. H., Sahoo, D., & Liu, Z.-Y. (2015). Moving average reversion strategy for on-line portfolio selection. Artificial Intelligence, 222, 104–123. https://doi.org/10.1016/j.artint.2015.01.006
36. Li, H., & Liu, T. (2023). Portfolio optimization based on the LSTM forecasting model. In Proceedings of the 2nd International Conference on Financial Technology and Business Analysis (Vol. 48, pp. 97–106). Advances in Economics, Management and Political Sciences. https://doi.org/10.54254/2754-1169/48/20230431
37. Lin, H.-Y., & Hsu, B.-W. (2023). Empirical study of ESG score prediction through machine learning: A case of non-financial companies in Taiwan. Sustainability, 15(19), 14106. https://doi.org/10.3390/su151914106
38. Lins, K. V., Servaes, H., & Tamayo, A. (2017). Social capital, trust, and firm performance: The value of corporate social responsibility during the financial crisis. The Journal of Finance, 72(4), 1785–1824. https://doi.org/10.1111/jofi.12505
39. Liu, T. (2022, December). Financial constraint' impact on firms' ESG rating based on Chinese stock market. In Proceedings of the 2022 4th International Conference on Economic Management and Cultural Industry (ICEMCI 2022) (pp. 1085–1095). Atlantis Press. https://doi.org/10.2991/978-94-6463-098-5_122
40. López de Prado, M. (2016). Building diversified portfolios that outperform out-of-sample. The Journal of Portfolio Management, 42(4), 59–69. https://doi.org/10.3905/jpm.2016.42.4.059
41. Ma, Y., Han, R., & Wang, W. (2021). Portfolio optimization with return prediction using deep learning and machine learning. Expert Systems with Applications, 165, 113973. https://doi.org/10.1016/j.eswa.2020.113973
42. Mai, F., Tian, S., Lee, C., & Ma, L. (2019). Deep learning models for bankruptcy prediction using textual disclosures. European Journal of Operational Research, 274(2), 743–758. https://doi.org/10.1016/j.ejor.2018.10.024
43. Maillard, S., Roncalli, T., & Teïletche, J. (2010). The properties of equally weighted risk contribution portfolios. The Journal of Portfolio Management, 36(4), 60–70. https://doi.org/10.3905/jpm.2010.36.4.060
44. Markowitz, H. (1952). Portfolio selection. The Journal of Finance, 7(1), 77–91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x
45. Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. Expert Systems with Applications, 42(1), 259–268. https://doi.org/10.1016/j.eswa.2014.07.040
46. Roncalli, T., & Weisang, G. (2016). Risk parity portfolios with risk factors. Quantitative Finance, 16(3), 377–388. https://doi.org/10.1080/14697688.2015.1046907
47. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x
48. Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. Applied Soft Computing, 90, 106181. https://doi.org/10.1016/j.asoc.2020.106181
49. Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. The Journal of Finance, 19(3), 425–442. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x
50. Ta, V. D., Liu, C.-M., & Tadesse, D. A. (2020). Portfolio optimization-based stock prediction using long-short term memory network in quantitative trading. Applied Sciences, 10(2), 437. https://doi.org/10.3390/app10020437
51. Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. The Journal of Finance, 62(3), 1139–1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x
52. Whited, T. M., & Wu, G. (2006). Financial constraints risk. The Review of Financial Studies, 19(2), 531–559. https://doi.org/10.1093/rfs/hhj012
53. Yang, S. (2023). Deep reinforcement learning for portfolio management. Knowledge-Based Systems, 278, 110905. https://doi.org/10.1016/j.knosys.2023.110905
54. Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 35(12), 11106–11115. https://doi.org/10.1609/aaai.v35i12.17325

JOURNAL OF GLOBAL TRENDS IN SOCIAL SCIENCE

J G T S S

Review article

https://doi.org/10.70731/1178nq90

Privacy-Preserving and Trustworthy AI Infrastructures for Digital Commerce: Federated Learning, Cross-Channel Measurement, and Social Advertising

Ahmad Zulkifli Bin Idris [1], **Weiling Tan** [2,*], **Kavitha Rajendran** [3]

Received 19 January 2026

Accepted 23 March 2026

Published 31 March 2026

Abstract: Digital commerce is shifting from unconstrained data accumulation to a context in which privacy regulation, platform restrictions, and public distrust shape how artificial intelligence is designed and deployed. This review argues that privacy-preserving and trustworthy commerce AI should be understood as an infrastructure problem rather than a set of isolated model-level improvements. It synthesizes research on privacy-preserving machine learning, digital advertising measurement, recommender systems, social commerce, creator monetization, platform architecture, and AI governance, and proposes an analytical framework structured around four layers: data topology, learning protocol, measurement logic, and governance architecture. At the data layer, consumer traces are fragmented across merchants, platforms, creators, and devices, making centralized modeling increasingly costly, risky, and legally fragile. At the learning layer, federated learning, differential privacy, secure aggregation, and related techniques enable distributed training and protected analytics, but introduce trade-offs in accuracy, communication, personalization, and security. At the measurement layer, the erosion of third-party identifiers makes privacy-preserving attribution and audience matching central to campaign optimization. At the governance layer, multi-tenant architecture, API standardization, auditability, and risk management determine whether such systems can operate at scale, especially for SMBs. The review concludes that the future of commerce AI will depend less on prediction alone than on the ability to institutionalize privacy, transparency, and accountability within commercially viable infrastructures.

Keywords: Digital commerce; Federated learning; Differential privacy; Secure aggregation; Social advertising; Creator economy; Recommender systems; AI governance

ISSN 2759-7830 (Online)
ISSN 2760-2508 (Print)

Introduction

The technical and economic logic of digital commerce has changed. This transition is also part of a broader restructuring of digital and social media marketing research, where data access, platform governance, and AI-mediated personalization have become central rather than peripheral concerns (Dwivedi et al., 2021). For roughly two decades, the dominant paradigm assumed that more data centralization would reliably produce better personalization, more precise advertising, stronger attribution, and lower customer acquisition cost. Recommendation engines, bidding systems, and marketing analytics were therefore designed around the continuous extraction and fusion of behavioral traces across sites, devices, and application contexts. That model is now under pressure from multiple directions. Privacy regulation has expanded the compliance burden surrounding personal data processing; platform policies have restricted tracking and identifier sharing; consumers have become more aware of surveillance-oriented advertising; and firms increasingly face the reputational and operational costs of building AI systems on data practices that are difficult to justify or audit (Boerman et al., 2017; Dinev et al., 2013; Morimoto, 2021; Truong et al., 2021). As a result, digital commerce has moved into a stage in which the central challenge is no longer only how to optimize

[1] University of Malaya, Kuala Lumpur 50603, Malaya; [2] Monash University Malaysia, Selangor 47500, Malaysia; [3] University of Technology Malaysia, Johor 81310, Malaysia.
* Corresponding author. Email: weiling.tan@monash.edu.

AI models, but how to construct AI infrastructures that remain effective under privacy, governance, and interoperability constraints.

This shift is especially consequential because digital commerce is structurally heterogeneous. Data relevant to a single commercial decision can be distributed across retailers, marketplaces, social platforms, creators, payment intermediaries, analytics providers, and cloud services. A recommendation model may depend on on-site browsing, purchase history, inventory states, content metadata, and social signals. An advertising system may combine audience estimation, creative selection, attribution, fraud detection, and budget allocation. A creator monetization workflow may require audience matching, engagement prediction, content moderation, revenue allocation, and contractual auditability. In each case, the technical object is not a single model but a socio-technical pipeline embedded in a wider platform ecosystem (Hein et al., 2020; Mukhopadhyay & Bouwman, 2019). This is why privacy-preserving AI in commerce cannot be understood merely as the addition of a privacy mechanism to an otherwise unchanged system. It entails redesigning the infrastructure through which data, incentives, models, and accountability are coordinated.

Research on trustworthy AI has reinforced this broader view. Surveys and policy-oriented syntheses increasingly argue that trustworthy AI is not exhausted by fairness or explainability at the model level; it also includes robustness, traceability, responsibility allocation, and risk management across the system life cycle (Corrêa et al., 2023; Díaz-Rodríguez et al., 2023; Laux et al., 2024; Lewis & Moorkens, 2020; Tabassi, 2023). In digital commerce, these issues are intensified by commercial imperatives. Firms want more granular measurement, faster experimentation, and scalable personalization, while regulators and users demand minimization, transparency, and control. The resulting tension cannot be resolved by normative statements alone. It requires architectures capable of combining privacy preservation with commercially usable outputs.

Federated learning has become a prominent candidate because it allows multiple parties or devices to train shared models without directly pooling raw data (Cheng et al., 2020; Kairouz et al., 2021). Yet federated learning by itself does not solve the underlying commerce problem. Distributed training still leaves questions about leakage from gradients, participant heterogeneity, personalization quality, verification, content governance, and downstream measurement. Differential privacy can limit leakage, but often at the cost of utility (Abadi et al., 2016; Wei et al., 2020). Secure aggregation can reduce exposure of participant updates, but it complicates orchestration and operational resilience (Bonawitz et al., 2017). Zero-knowledge proofs and related cryptographic mechanisms can improve verification and accountability, but they introduce complexity and latency that many commerce systems are not designed to absorb (Sun et al., 2021). In short, privacy-preserving AI for commerce is modular, not monolithic.

The importance of this problem is amplified by two business trends. First, commerce journeys are now inherently cross-channel. Consumers move between search, social feeds, creator content, marketplaces, merchant websites, loyalty applications, and physical channels. Measurement and attribution have therefore become more difficult at exactly the moment when identifiers are less stable. The classical attribution problem, already difficult in pay-per-conversion settings, has become more severe in privacy-constrained environments (Jordan et al., 2011). Second, small and medium-sized businesses increasingly rely on external AI infrastructures rather than internal data science teams. This expands the relevance of multi-tenant architectures, standardized recommendation and measurement APIs, and platform-level governance mechanisms that can lower adoption barriers while embedding compliance and auditability into default workflows (Arnold et al., 2022; Gieß & Hutterer, 2025; Xue et al., 2019).

Against this background, the present review treats privacy-preserving and trustworthy commerce AI as an integrated infrastructure domain. It does not approach recommendation, advertising, attribution, creator monetization, and SMB enablement as separate literatures that only happen to share some methods. Instead, it examines how these functions become interdependent once privacy constraints limit frictionless data pooling. The review makes three contributions. First, it synthesizes the technical stack of federated learning, differential privacy, secure aggregation, and zero-knowledge verification in terms of their relevance for commerce rather than in purely generic machine learning terms. Second, it links privacy-preserving learning to cross-channel measurement and social advertising, two domains where commercial value depends on the controlled recovery of signal from partially observable environments. Third, it shows why trustworthy AI in commerce must be operationalized at the platform architecture and governance level, especially for SMB-focused ecosystems.

The rest of the paper follows this logic. Section 2 defines the main commerce scenarios and explains why they generate distinct data topologies and privacy risks. Section 3 reviews the core privacy-preserving technical stack. Section 4 addresses cross-channel measurement and attribution in a post-cookie environment. Section 5 examines social commerce advertising and the creator economy. Section 6 analyzes trustworthy AI platforms for SMBs. Section 7 discusses governance, standards, and accountability. Section 8 concludes with a research agenda centered on the balance between privacy, utility, and commercial implementability.

Scenario Decomposition

Digital commerce is frequently described as a unified domain, but the infrastructural requirements of its main AI tasks differ sharply. A useful review must therefore begin by distinguishing scenarios according to their decision objects, data topology, latency requirements, and regulatory exposure. Five scenarios are especially central: recommendation, advertising delivery and targeting, cross-channel measurement and attribution, creator monetization, and platformized AI services for SMBs.

Recommendation remains the most mature commerce AI application. It includes product ranking, bundle suggestion,

personalized search, media selection, and retention-oriented recommendation. The modern literature emphasizes increasingly complex representations, including graph neural networks, cross-domain transfer, and self-supervised learning (Chen et al., 2025; Gao et al., 2023; Sharma et al., 2024). However, the privacy problem in recommendation is unusually acute because consumer preference data are both economically valuable and behaviorally intimate. Historical work on privacy-preserving recommender systems focused on anonymization, obfuscation, or cryptographic computation (Checco et al., 2017; Sweeney, 2002; Wang et al., 2018). More recent work shifts toward federated recommendation, where interaction data remain local to users, enterprises, or domains while models are coordinated centrally or semi-centrally (Asad et al., 2023; Kalloori & Klingler, 2021; Luo et al., 2024; Qin et al., 2021). The central trade-off is that preserving locality often worsens sparsity, non-IID data problems, and personalization difficulty.

Advertising delivery and targeting form a second scenario. Here the key task is not simply ranking items but matching audiences, contexts, and creatives under auction-like or budget-constrained conditions. Behavioral advertising research has long shown that performance depends on the availability of reliable behavioral features and identity linkage, but these same features trigger privacy concern, persuasion resistance, and trust deterioration when consumers perceive tracking as intrusive (Boerman et al., 2017; Carlson et al., 2022; Jiang et al., 2013). In social media settings, personalization also interacts with platform-specific motivations and trust perceptions, which means that advertising effectiveness depends not only on the accuracy of targeting but on the legitimacy of the targeting process itself (Carlson et al., 2022; Morimoto, 2021). Privacy-preserving advertising therefore requires both a technical solution to audience estimation and an institutional solution to perceived manipulation.

Cross-channel measurement and attribution constitute a third scenario and should not be reduced to a reporting problem. In digital commerce, firms optimize campaigns, recommendations, and incentives on the basis of measured downstream outcomes. When user journeys span multiple channels, devices, and actors, the absence of stable identifiers creates measurement loss: some conversions cannot be linked, some touchpoints cannot be sequenced, and some causal contributions cannot be isolated. The attribution literature already identified the difficulty of assigning value across multiple ad exposures and interactions even before current privacy restrictions (Jordan et al., 2011). The contemporary challenge is more severe because signal loss is now built into the environment. Privacy-preserving conversion measurement, secure aggregation of event counts, and differentially private reporting are increasingly central to campaign optimization rather than supplementary safeguards (Delaney et al., 2024; Farahat et al., 2009; Zhong et al., 2022).

Creator monetization forms a fourth scenario, especially within social commerce and influencer-led retail. In this setting, the economic unit is often neither the platform alone nor the merchant alone, but a triangular relationship among creators, audiences, and commercial sponsors. The relevant AI tasks include audience matching, engagement prediction, conversion estimation, content recommendation, fraud or manipulation detection, and revenue allocation. Research on influencer and live-streaming commerce shows that consumer response is mediated by credibility, parasocial relationships, platform affordances, and the perceived authenticity of content (Bargoni et al., 2023; Bi & Zhang, 2023; Libai et al., 2025; Teng et al., 2022; Xue & Liu, 2023). This complicates privacy-preserving design because the system must often infer high-value audience segments from interactional signals that are socially embedded and partly creator-specific. The result is a need for architectures that can support monetization without forcing creators to surrender comprehensive audience-level data to centralized intermediaries.

The fifth scenario is AI-as-infrastructure for SMBs. Most small firms cannot build proprietary recommender systems, privacy engineering pipelines, or governance functions. They rely on platforms that expose standardized APIs for recommendation, campaign management, analytics, catalog enrichment, or content moderation. From a systems perspective, SMB adoption depends on whether these services can be offered in a multi-tenant form that is sufficiently modular, interoperable, and compliant. Research on digital platform ecosystems shows that value creation depends on orchestrating complementors through shared interfaces and governance rules rather than merely providing software functionality (Hein et al., 2020; Mukhopadhyay & Bouwman, 2019; Xue et al., 2019). Recent platform research also highlights the importance of data architecture, modularity, and governance in shaping who can participate and under what conditions (Arnold et al., 2022; Gieß & Hutterer, 2025). Trustworthy AI for SMBs is therefore not only a matter of offering "responsible" models; it is a matter of packaging compliance, auditability, and privacy guarantees into accessible infrastructure.

These scenarios share some common technological primitives, but their data topologies differ. Recommendation often involves horizontally partitioned user-interaction data or cross-domain preference transfer. Advertising may combine platform-side audience estimates with merchant-side conversion signals. Attribution connects event logs generated across multiple systems. Creator monetization adds relational data between creators and audiences, as well as platform-specific content signals. SMB services require multi-tenant separation, tenant-specific policy controls, and standard interfaces that can bridge heterogeneous tools. Vertical federated learning is relevant when distinct organizations hold different features about overlapping user sets, while horizontal federated learning is more natural when participants hold similar features over disjoint populations (Khan et al., 2025; Kalloori & Klingler, 2021). Cross-silo and cross-device settings therefore have different security and orchestration implications.

A scenario-based view also clarifies why the privacy question cannot be resolved by abstract "compliance" language. The same privacy mechanism can be appropriate in one scenario and inadequate in another. For example, local differential privacy may be tolerable for aggregate measurement but too destructive for personalized recommendation. Secure aggregation may protect participant updates in cross-silo learn-

Table 1 | Core digital commerce scenarios and their infrastructural implications

Scenario	Main AI task	Data topology	Primary privacy risk	Main infrastructure implication
Recommendation	Ranking, retrieval, personalization	User- or domain-local interaction data	Preference leakage, re-identification, profiling	Federated recommendation, privacy-aware personalization
Social advertising	Audience matching, bidding, creative optimization	Platform-side behavioral traces plus merchant outcomes	Opaque targeting, cross-context tracking	Privacy-preserving targeting and reporting
Cross-channel measurement	Attribution, lift analysis, conversion counting	Event logs across channels and devices	Linkage risk, unverifiable conversion claims	Secure aggregation, DP reporting, auditable measurement
Creator monetization	Matching sponsors, audiences, and content	Creator-specific audiences plus platform interaction graphs	Centralized exposure of audience value	Verifiable revenue sharing and protected audience analytics
SMB platform services	Model hosting, recommendation APIs, content governance	Multi-tenant enterprise data and policy settings	Tenant leakage, inconsistent compliance	Trustworthy multi-tenant architecture and standardized APIs

ing but still leave unresolved questions about malicious clients or biased participation. Zero-knowledge proofs may be particularly valuable where revenue sharing or creator compensation requires verifiable accounting, but less necessary in low-stakes catalog ranking. A useful commerce architecture must therefore map privacy tools to scenario-specific functional requirements. **Table 1** summarizes the scenario logic used in this review.

The remainder of the review builds on this scenario differentiation by examining how privacy-preserving technologies can be assembled into commerce-specific infrastructures.

Privacy-Preserving Technical Stack

The privacy-preserving technology stack relevant to digital commerce is best understood as layered. Federated learning determines where model training occurs and how updates are coordinated. Differential privacy constrains what can be inferred from outputs or intermediate updates. Secure aggregation protects the visibility of participant-level updates during coordination. Zero-knowledge proofs and related cryptographic techniques support verification and accountability. Each layer addresses a different failure mode, and none is sufficient on its own.

Federated learning as an infrastructural rather than purely algorithmic choice

Federated learning emerged as a response to the concentration of data in centralized machine learning pipelines. Instead of moving raw data to a single repository, participants compute local updates that are then aggregated into a shared model (Cheng et al., 2020; Kairouz et al., 2021). This basic idea is attractive for commerce because merchants, platforms, creators, and user devices often have incentives not to share raw data. Federated learning can preserve local data residency while still enabling collective model improvement.

Yet the relevance of federated learning to commerce depends on deployment form. Cross-device federated learning is suitable when the primary participants are consumer devices, for example in on-device ranking or personalized content selection. Cross-silo federated learning fits settings where participants are institutions, such as merchants collaborating with an advertising network or multiple brands joining a retail media platform. Horizontal and vertical federated learning further distinguish whether parties hold similar feature spaces over different users or different feature spaces over overlapping users (Kalloori & Klingler, 2021; Khan et al., 2025). These distinctions matter because the operational problems differ. Cross-device settings face high churn, weak trust assumptions, and device heterogeneity. Cross-silo settings face stronger governance and contractual issues, but usually enjoy more stable participation and richer local computation.

For recommender systems, federated learning is appealing because interaction histories are especially privacy-sensitive. Surveys show rapid growth in federated recommendation, including matrix factorization, neural recommendation, graph-based models, and personalized federated strategies (Asad et al., 2023; Kalloori & Klingler, 2021; Luo et al., 2024; Qin et al., 2021). At the same time, several technical obstacles remain persistent. First, user behavior data in commerce are highly non-IID. Consumers differ across language, category preference, price sensitivity, and channel use, which makes global models unstable. Second, sparse and long-tail item spaces create difficulties when local clients only observe tiny slices of the inventory. Third, recommendation quality often depends on rich side information, some of which may be distributed across merchants, platforms, and content systems rather than concentrated on a single device. These problems mean that federated learning must often be combined with personalization layers, cross-domain transfer, or graph-aware representations (Chen et al., 2025; Gao et al., 2023; Sharma et al., 2024).

Federated learning also creates new attack surfaces. Even when raw data stay local, model updates can leak information, and malicious participants can poison training or infer sensitive attributes. Surveys consistently identify gradient leakage, inference attacks, poisoning, backdoors, and free-riding as major risks (Gosselin et al., 2022; Jere et al., 2021; Mothukuri et al., 2021; Papadopoulos et al., 2021). For commerce applications, these risks are not only technical. An adversarial merchant could manipulate shared recommendation training; a platform participant could attempt to infer the strategic value of another tenant's audience; or a creator-side

application could introduce distorted engagement signals. Consequently, federated learning should be seen not as a privacy guarantee but as a controlled redistribution of where risk appears.

Differential privacy and the pricing of privacy loss

Differential privacy provides a mathematically explicit way to limit the impact of any single record on released outputs or model parameters (Dwork, 2006). In practice, it is relevant to commerce for three broad reasons. First, it can reduce the likelihood that recommendation, advertising, or measurement outputs expose identifiable behavior. Second, it provides a language of privacy budgets that can be documented and governed. Third, it allows organizations to communicate formal guarantees in settings where informal assurances about "anonymization" have become unreliable.

The introduction of differential privacy into deep learning made these ideas operational for modern models, although often with substantial utility trade-offs (Abadi et al., 2016). In federated settings, the utility-privacy trade-off is further complicated by client heterogeneity, communication limits, and the fact that noise may be inserted at local or aggregate stages (Wei et al., 2020). For commerce, the key question is not whether differential privacy can be added, but where in the pipeline it is most valuable. Local differential privacy offers stronger participant-side protection but often destroys fine-grained signal needed for personalization. Central differential privacy, applied after secure aggregation, can preserve more utility but requires stronger trust in the aggregation server or protocol. Aggregate reporting tasks such as campaign measurement may tolerate higher noise than item-level ranking or dynamic pricing decisions.

The privacy budget perspective is particularly useful for commerce governance. Recommendation, targeting, and measurement are not one-time computations; they are recurring processes that consume privacy budget across repeated releases, experiments, and reporting intervals. A firm that runs continuous campaign attribution, creator analytics, and user segmentation needs an accounting framework for cumulative privacy loss. This links privacy engineering to platform governance: privacy budgets are not only mathematical objects but resource-allocation decisions shaped by business priorities. Surveys on privacy-preserving federated learning show that practical deployments must decide which outputs are worth preserving at higher fidelity and which can absorb more noise (Gu et al., 2023; Truong et al., 2021). In that sense, differential privacy is also a managerial instrument.

Secure aggregation and the hidden middle layer of trust

Secure aggregation protects the confidentiality of individual participant updates during federated coordination. The canonical design goal is that the server should learn only the aggregate sum of client updates, not the update of any single client (Bonawitz et al., 2017). In commerce, this is critical when multiple organizations jointly train models or contribute outcome events. Without secure aggregation, federated learning may offer only superficial privacy because the coordinating party can inspect participant-level gradients.

Secure aggregation is especially relevant to cross-channel measurement. If advertisers, platforms, merchants, and analytics providers each contribute event signals, a naïve aggregation system can expose partner-level conversion patterns or strategic performance information. Privacy-preserving event reporting and frequency management were already recognized in earlier online advertising work (Farahat et al., 2009). The current environment strengthens the case for secure aggregation because direct identifier-level matching is both harder and more controversial. Measurement systems increasingly need to recover aggregate signal while hiding participant-specific detail.

However, secure aggregation is often treated as a neutral technical layer when it is actually a design choice with governance implications. The protocol determines fault tolerance, dropout handling, communication burden, and the point at which trust is centralized. In cross-silo commerce systems, these operational details affect participation. Small merchants or creators may not tolerate protocols that are too costly or brittle. Thus, secure aggregation should be evaluated not only in cryptographic terms but in infrastructural terms: who can realistically join, who controls orchestration, and what evidence of correct execution is available.

Zero-knowledge proof, verification, and accountable commerce AI

Zero-knowledge proofs (ZKPs) are usually discussed in blockchain contexts, but their relevance to commerce AI lies more broadly in verifiable computation and accountability. A zero-knowledge protocol allows one party to prove that a statement is true without revealing the underlying secret information (Sun et al., 2021). In commerce infrastructures, this can support claims such as: a conversion count was computed according to an agreed protocol; a revenue share was allocated using an approved formula; a participant satisfied eligibility conditions without disclosing raw data; or a model-serving process complied with a defined policy rule.

This is particularly salient in creator monetization and federated advertising. Creators often depend on platform-provided analytics to assess sponsorship value and payout fairness. Merchants depend on platforms to report audience quality and conversions. When raw logs cannot be disclosed for privacy or proprietary reasons, verification becomes difficult. ZKPs do not solve all of these problems, but they provide a route toward auditable claims without full data exposure. Their main limitation is practical: proof generation and verification impose computational costs, and integration into real-time or near-real-time systems is still complex.

Architectural composition and trade-offs

The main analytical point is that commerce systems should not select privacy tools one by one in isolation. The functional unit is a composed stack. Federated learning addresses data locality; differential privacy constrains inferential exposure; secure aggregation protects intermediate coordination; ZKPs support verification; and API-level governance

Table 2 | Privacy-preserving building blocks for commerce AI

Building block	Primary function	Commerce use cases	Main strengths	Main limitations
Federated learning	Distributed model training	Recommendation, audience modeling, shared fraud detection	Preserves data locality; enables multi-party learning	Non-IID data, communication overhead, attack surface
Differential privacy	Formal privacy guarantee through controlled noise	Reporting, analytics, recommendation training, measurement	Quantifiable privacy budget; useful for governance	Utility loss; difficult budget allocation across repeated tasks
Secure aggregation	Hides participant-level updates during coordination	Federated training, conversion counting, partner reporting	Reduces visibility of local updates	Protocol complexity; orchestration burden; dropout handling
Zero-knowledge proof	Verifiable computation without revealing raw data	Payout verification, conversion claims, compliance checks	Supports auditability under data minimization	Computation cost; integration complexity
Standardized APIs and policy layers	Controlled exposure of model and analytics functions	SMB enablement, platform orchestration, tenant governance	Scalability, interoperability, embedded compliance	Requires mature governance and version control

structures determine how these components are exposed to tenants and partners. Research on privacy-preserving recommendation-as-a-service and privacy-aware commercial AI points toward this compositional logic, even when implementation details vary (Wang et al., 2018; Papadopoulos et al., 2021). **Table 2** summarizes the main roles and trade-offs of these primitives.

The literature suggests that utility-preserving privacy in commerce will not come from maximizing any single technique. Instead, it will come from designing a stack whose components are matched to the information requirements and trust assumptions of specific scenarios. This point becomes clearer once we turn from training infrastructure to measurement infrastructure.

Cross-Channel Measurement and Attribution in A Post-Cookie Environment

Cross-channel measurement is the point at which privacy preservation most visibly collides with commercial decision making. Recommendation and targeting models are valuable because they influence outcomes, but firms cannot justify continued investment without measurement. Once identifiers become unstable and event-level linkage becomes constrained, campaign optimization, budget allocation, and channel evaluation all degrade. The core question is therefore not simply how to measure less invasively, but how to reconstruct enough signal for action while respecting data minimization.

The attribution literature identified the difficulty of assigning conversion credit long before today's privacy restrictions. In pay-per-conversion advertising, multiple exposures may contribute jointly to a conversion, making straightforward last-touch accounting strategically misleading (Jordan et al., 2011). The contemporary setting adds two new complications. First, observable user paths are incomplete because tracking is fragmented across browsers, applications, platforms, and walled gardens. Second, even when data exist somewhere in the system, they may not be legally or contractually combinable. This transforms attribution from an econometric challenge into an infrastructural one.

One response is to shift attention from user-level traceability to privacy-preserving aggregate measurement. Recent work on ad conversion measurement develops differentially private mechanisms that release aggregate campaign statistics while bounding privacy leakage (Delaney et al., 2024). Similar efforts in advertising security propose privacy-preserving conversion tracking and bidding protocols that allow parties to optimize campaigns without fully exposing conversion logs (Zhong et al., 2022). Historically, even comparatively narrow tasks such as frequency capping generated privacy concerns because they required maintaining exposure histories (Farahat et al., 2009). The current research frontier generalizes this insight: almost every useful advertising metric depends on linking behavior over time, which means privacy-preserving measurement must selectively reconstruct just enough linkage to support action.

This requirement produces measurement loss, a concept that is useful even when different papers use different terminology. Measurement loss refers to the gap between true causal or transactional activity and what the system can legitimately and technically observe. In commerce, this loss has four dimensions. First, identity loss occurs when the system cannot reliably connect touchpoints to the same user. Second, path loss occurs when intermediate interactions are not visible. Third, outcome loss occurs when conversions happen in contexts not observable to the measurement partner. Fourth, semantic loss occurs when privacy-preserving aggregation removes detail needed to interpret heterogeneous conversions. Differential privacy can protect against leakage, but it also enlarges semantic loss if the released outputs are too coarse. Secure aggregation can hide local contributions, but it does not guarantee causal interpretability. This is why privacy-preserving measurement must be paired with careful metric design.

A second response is to redesign incentives around partial observability. If the system cannot perfectly observe every conversion path, then contracts and optimization rules must be robust to incomplete measurement. This insight has practical relevance in creator commerce and affiliate-like settings, where payout formulas may depend on noisy or delayed attribution. Platform ecosystems can reduce conflict by using

standardized reporting protocols and clearly documented confidence intervals or eligibility rules, rather than pretending that privacy-preserving measurement is exact. The problem is not only technical uncertainty but contestability. When merchants, creators, and platforms do not share raw logs, they need rules for how approximate measurement enters billing, budgeting, and compensation.

A third response is to exploit structured forms of distributed learning and matching. Vertical federated learning is potentially relevant when different organizations hold complementary features about overlapping user sets, such as platform-side engagement features and merchant-side purchase outcomes (Khan et al., 2025). Federated cross-domain recommendation and federated rating prediction also illustrate how signal can be transferred across domains without unrestricted data pooling (Wu et al., 2022). These approaches are not measurement systems in the narrow sense, but they show how cross-channel information can be leveraged under partitioned data conditions. The limitation is that overlap resolution and identity correspondence remain difficult, especially when privacy constraints forbid explicit linkage.

The measurement problem is therefore best understood as a trade space among fidelity, privacy, and verifiability. High-fidelity user-level attribution maximizes optimization value but creates the greatest exposure. Strong privacy with heavy noise or coarse aggregation minimizes exposure but may be too weak for commercial action. Verifiable protocols improve trust but can slow deployment. A privacy-preserving commerce system must choose a point in this trade space based on decision purpose. Strategic budget allocation across channels may tolerate aggregate reporting with uncertainty bounds. Real-time bidding and creative optimization may require more granular proxies. Creator compensation may require auditable but delayed settlement rather than instantaneous perfect attribution.

Behavioral research further indicates that measurement design shapes trust. Consumers' privacy concerns are not driven solely by formal data collection volume, but by perceived loss of control, opacity, and manipulative intent (Chaudhuri et al., 2023; Dinev et al., 2013; Jiang et al., 2013). Thus, privacy-preserving measurement can have indirect commercial benefits if it reduces the perception that advertising relies on hidden surveillance. At the same time, if reporting becomes too opaque or technically obscure, merchants and creators may distrust the platform instead. The design task is therefore dual-facing: measurement must be privacy-legible to users and accountability-legible to business participants.

The literature suggests several research priorities. First, more work is needed on the interaction between differential privacy parameters and business decision quality in realistic campaign settings. Second, attribution models should explicitly incorporate observability constraints rather than treating missingness as a nuisance. Third, privacy-preserving measurement systems need stronger audit layers, potentially including cryptographic proofs of correct aggregation or rule execution. Fourth, incentive-compatible contracts are needed for settings where measurement is approximate by design. Privacy-preserving measurement is therefore not a residual technical adjustment to post-cookie advertising; it is becoming the central coordination mechanism through which digital commerce decides what counts as performance.

A recent preprint by Yi (2026a) is illustrative here. It proposes a federated and differentially private framework for cross-channel measurement that explicitly integrates Topics and Protected Audience on the web side with Attribution Reporting and SKAdNetwork on the app side, while linking measurement to incentive allocation for SMB advertisers. Because the work is currently a preprint, it should be interpreted cautiously. Even so, it is directly relevant to this review because it treats post-cookie measurement as a joint problem of privacy budgeting, channel harmonization, and decision support rather than as a narrow reporting task.

Social Commerce Advertising and The Creator Economy

Social commerce and the creator economy have transformed the structure of advertising by embedding persuasion, discovery, and transaction inside relationship-rich media environments. In conventional display advertising, targeting and measurement are often discussed as relatively separate layers. In creator commerce, they are intertwined. The value of an impression depends not only on who sees it but on who delivers it, how the content is framed, what platform norms govern disclosure, and how conversion is attributed across social and transactional touchpoints. This makes privacy-preserving design especially difficult.

Research in marketing and interactive media shows that creator influence depends on credibility, perceived similarity, parasocial interaction, and the alignment between message form and platform culture (Bi & Zhang, 2023; Carlson et al., 2022; Teng et al., 2022). Consumers often respond to creators not as interchangeable ad inventory but as trusted or quasi-relational intermediaries. In live-streaming and ecosystem-based analyses of social selling, value creation flows through interactions among platforms, anchors, brands, and audiences rather than through one-directional promotion alone (Xue & Liu, 2023). Recent synthesis work on the creator economy similarly emphasizes that value chains are multi-actor systems involving platforms, creators, advertisers, agencies, and analytics providers (Libai et al., 2025). This means that the data generated in creator commerce are relational, contextual, and partially co-produced.

From a privacy perspective, this relationality matters in two ways. First, creator audiences are strategic assets. Platforms and brands want to infer which audiences are likely to convert, but creators may resist architectures that fully expose their audience data because those data underpin bargaining power. Second, consumers may accept creator recommendations partly because they perceive them as socially situated rather than purely algorithmic. Excessively invasive targeting can undermine that perception and erode trust. The problem is therefore not simply how to protect user privacy,

but how to preserve the autonomy and informational position of intermediaries within the commerce ecosystem.

Federated and privacy-enhancing methods are relevant because they can decouple shared model improvement from unrestricted data access. Audience modeling or recommendation for creator-brand matching could, in principle, be trained across creators, merchants, or platforms without centralizing all raw interaction data. Surveys of federated recommendation and privacy enhancement suggest that this is technically plausible, particularly where participants have partially aligned objectives but do not want to share raw histories (Asad et al., 2023; Gosselin et al., 2022; Papadopoulos et al., 2021). Yet social commerce introduces additional complications. Engagement signals can be noisy, strategic, and vulnerable to manipulation. Creator-side optimization may also encourage gaming behaviors if payout formulas are visible but measurement is imperfect. Thus, privacy-preserving social advertising requires both learning protection and integrity control.

The creator economy also raises the question of revenue sharing and compensation verification. Influencer campaigns, affiliate arrangements, and platform-mediated creator funds all depend on claims about reach, engagement quality, conversions, or downstream sales. Bargoni et al. (2023) show that endorsement services can affect campaign outcomes, but commercial value is contingent on how those services are operationalized. If the creator, platform, and sponsor do not share raw data, disputes can arise over whether audience delivery and conversions were measured correctly. Here zero-knowledge proofs or other verifiable reporting methods may become useful because they allow a party to demonstrate compliance with an agreed formula without disclosing raw logs. Even when such mechanisms are not fully implemented, the design logic is clear: creator monetization requires privacy-preserving visibility rather than either total secrecy or total transparency.

Another relevant strand of literature concerns virtual influencers and platform-mediated identity. The emergence of virtual or AI-driven influencers intensifies privacy tensions because the distinction between content generation, user data exploitation, and synthetic persuasion becomes blurred. Recent work conceptualizes this through a multi-privacy paradox in which consumers may disclose or accept more than they normatively endorse under conditions of convenience, entertainment, or social immersion (Liyanaarachchi et al., 2024). This observation has broader relevance for creator commerce. A privacy-preserving infrastructure cannot assume that disclosed preference is a reliable measure of informed consent. Trustworthy design must therefore include procedural safeguards and governance constraints, not only predictive optimization.

There is also an asymmetry between large platforms and small creators or merchants. Major platforms can absorb privacy engineering costs and may internalize large-scale behavioral data regardless of external restrictions. Smaller actors depend on whatever analytics and APIs the platform exposes. This suggests that trustworthy creator commerce cannot be evaluated solely at the firm level; it must be assessed at the ecosystem level. Standardized, privacy-preserving reporting interfaces can improve access for smaller participants, but only if the platform's governance rules also address auditability, explainability of payout logic, and content moderation consistency. Otherwise, privacy discourse may simply mask further asymmetry in informational control.

A useful design principle is to treat audience matching, conversion estimation, and revenue allocation as distinct yet connected layers. Audience matching can often tolerate partial decentralization and privacy-preserving representation learning. Conversion estimation may rely on differentially private or aggregated outcome reporting. Revenue allocation may require explicit verification and contract rules. Collapsing all three into one opaque platform metric produces efficiency in the short term but undermines long-term trust. Social advertising under privacy constraints therefore benefits from modularity.

The literature also points toward several open questions. First, little is known about how privacy-preserving advertising architectures affect creator bargaining power and market concentration. Second, more research is needed on fairness in creator recommendation and monetization when audience data are unevenly observable. Third, current work still underspecifies how content governance and privacy protection interact. A platform may privacy-protect user data while still recommending harmful or misleading commercial content. Finally, the boundary between recommendation and advertising becomes increasingly blurred in creator ecosystems, suggesting that governance categories inherited from earlier digital advertising may be analytically insufficient. Trustworthy commerce AI in this domain must therefore integrate privacy, content accountability, and economic verification rather than addressing them sequentially.

A closely related 2026 article by Yi (2026c) extends this line of thinking to social e-commerce advertising and creator monetization. Its proposed combination of federated learning and zero-knowledge verification is useful for this review not because it resolves the empirical question of platform fairness, but because it makes explicit a crucial architectural distinction: audience modeling, ad interaction verification, and creator payout accountability can be separated and then recomposed. That formulation reinforces the argument advanced here that trustworthy creator commerce depends on privacy-preserving visibility rather than either unrestricted surveillance or opaque platform reporting.

Trustworthy AI platforms for SMBs

The commercial relevance of privacy-preserving AI depends heavily on whether it can be operationalized for small and medium-sized businesses. Large technology firms can build proprietary data infrastructure, privacy engineering teams, and internal audit capabilities. Most SMBs cannot. They adopt AI through platforms, software-as-a-service vendors, marketplaces, and ecosystem intermediaries. The critical question is therefore how trustworthy AI can be packaged into multi-tenant infrastructures that are technically scalable and organizationally usable.

Table 3 | Design blueprint for trustworthy SMB-oriented commerce AI infrastructure

Layer	Design objective	Key components
Data layer	Protect tenant and user data while enabling useful learning	Local storage controls, data minimization, retention policies, tenant isolation
Learning layer	Support shared improvement without unrestricted pooling	Federated learning, secure aggregation, personalization modules, attack monitoring
Measurement layer	Provide actionable analytics under privacy constraints	Differentially private reporting, aggregate conversion measurement, confidence disclosure
Governance layer	Make the system auditable and controllable	Risk management workflows, logging, incident response, human oversight
Interface layer	Lower adoption barriers for SMBs	Standardized APIs, service cards, policy-aware configuration, interoperable documentation

Multi-tenant architecture is central because SMB-facing systems must serve many organizations with limited customization cost. Yet multi-tenancy creates its own privacy and governance problems. Tenant data must be logically or cryptographically separated; model improvements may need to be shared without leaking competitive information; policy controls must vary across sectors and jurisdictions; and audit logs must remain intelligible to clients who are not AI specialists. Architectural work on platforms and industrial data infrastructures shows that modularity, orchestration logic, and interface design strongly shape adoption and control (Arnold et al., 2022; Gieß & Hutterer, 2025; Hein et al., 2020). In commerce AI, the same principle applies. A platform that merely offers an API endpoint for "recommendation" without governance metadata, logging, and privacy options is not providing trustworthy infrastructure; it is externalizing governance burdens to the least capable actors.

Standardized APIs occupy a strategic position in this architecture. They translate complex learning, ranking, and measurement processes into callable services for merchants, creators, and app developers. Research on APIs and complementor innovation shows that interface standardization can stimulate external innovation, but it can also reproduce dependency and copying risks when governance is weak (Xue et al., 2019). For privacy-preserving commerce AI, API design should do more than expose functionality. It should define the permissible scope of data use, the granularity of outputs, the retention and deletion logic, and the audit trails associated with each invocation. Put differently, trustworthy AI for SMBs requires policy-aware APIs rather than bare prediction endpoints.

Recommendation-as-a-service provides a concrete example. Traditional centralized services ask clients to upload user and item data to the provider. Privacy-preserving recommendation-as-a-service aims to protect data while still enabling shared computation, often through distributed learning or cryptographic protocols (Wang et al., 2018). For SMBs, the attraction is clear: they can access advanced recommendation without building internal pipelines. But unless the service also supports tenant-level governance, documentation of privacy guarantees, and clear liability boundaries, adoption may be superficial. The platform may be "privacy-enhancing" in a technical sense while remaining opaque in contractual or operational terms.

Content governance is another essential component. Digital commerce increasingly depends on AI systems that rank or generate product descriptions, moderate user-generated reviews, screen promotional content, and route creator materials. Trustworthy infrastructure therefore requires alignment between privacy design and content governance. A platform that protects training data but fails to control harmful or noncompliant commercial content is not trustworthy in the commerce sense. Conversely, aggressive moderation without due process or auditability can damage SMBs that depend on the platform for visibility. The governance challenge is to embed policy constraints into workflow design rather than bolting them onto downstream review.

This is where risk management frameworks become important. NIST's AI RMF frames trustworthy AI through functions such as govern, map, measure, and manage, emphasizing life-cycle controls rather than one-off ethical commitments (Tabassi, 2023). The European trustworthy AI guidance and associated assessment tools similarly move from principles toward operational checklists and self-assessment structures (High-Level Expert Group on Artificial Intelligence [HLEG], 2019, 2020). For SMB-oriented commerce platforms, these frameworks suggest that providers should supply governance scaffolding as part of the service. Examples include configurable privacy budgets, model cards or service cards, incident reporting procedures, escalation channels, and human-review triggers for sensitive content or decisions.

A trustworthy SMB platform must also reconcile global shared models with tenant-specific context. A generic recommender or targeting model may not reflect local catalog structure, legal obligations, or sector-specific sensitivities. Federated approaches offer one route by enabling shared learning across tenants while keeping raw data local, but this is only part of the solution. Platforms also need tenant-aware policy layers and configurable governance modules. The most promising direction is therefore not "one model for all" but a layered system in which shared representation learning, tenant-specific fine-tuning, privacy accounting, and rule enforcement are separated but interoperable. **Table 3** summarizes a design blueprint for trustworthy SMB AI platforms.

The SMB context highlights a broader point: trustworthy AI is not simply a quality of algorithms but a service design principle. It concerns how capabilities are exposed, constrained, documented, and monitored. Privacy-preserving infrastructures become economically meaningful only when

these properties are operationalized in ways that non-expert organizations can actually use.

On the provider side, Yi (2026b) offers a complementary architecture paper on trusted AI commercialization infrastructure for SMBs. The paper emphasizes multi-tenant governance controls, standardized recommendation APIs, incentive systems, and content-governance workflows. Its importance for the present review lies in showing that privacy preservation alone is insufficient for SMB adoption unless it is packaged together with auditability, reusable interfaces, and policy-aware service design. As with other recent architecture papers, the contribution is best read as a design blueprint rather than as settled evidence of market-wide effectiveness.

Governance and Standards

The governance debate surrounding commerce AI has matured beyond general ethical exhortation, but operational gaps remain significant. Trustworthy AI is now commonly associated with transparency, accountability, human oversight, robustness, fairness, and privacy. The difficulty is that these principles are often stated at a high level while commerce systems require concrete allocations of decision rights, reporting duties, and technical responsibilities across multiple actors.

Large-scale reviews of AI governance guidelines show convergence around a limited set of principles, especially transparency, justice or fairness, non-maleficence, responsibility, and privacy (Corrêa et al., 2023). However, the same reviews also show fragmentation in interpretation and implementation. Commerce systems intensify this problem because many relevant decisions are distributed across platforms, advertisers, creators, vendors, and analytics intermediaries. A social advertising platform may control ranking and moderation; a merchant may control pricing and campaign objectives; a creator may control message framing; and a measurement vendor may control attribution outputs. Governance therefore has to assign responsibility across a chain rather than inside a single organization.

Trustworthy AI scholarship increasingly argues that this chain perspective must be made explicit. Díaz-Rodríguez et al. (2023) connect AI principles to system requirements and regulation, emphasizing that trustworthy AI involves the translation of abstract norms into design controls, governance processes, and compliance mechanisms. Laux et al. (2024) further caution that trustworthiness should not be conflated with mere acceptability of risk. This distinction matters in commerce because a platform may technically comply with risk categories while still cultivating opaque incentive structures or asymmetrical information control. Governance should therefore ask not only whether risk is bounded, but also who defines acceptable risk and who bears the residual cost when the system fails.

Transparency is often the first principle invoked, yet it is especially difficult to operationalize in commerce. Consumers do not need the same disclosure as merchants, regulators, or creators. Overly general notices can satisfy formal disclosure requirements while communicating little about actual system behavior. Conversely, highly technical descriptions may be unusable for most stakeholders. The practical lesson is that transparency must be role-specific. Users need intelligible explanations of data use and targeting logic at an appropriate level. Business participants need service-level information about measurement error, payout rules, and moderation triggers. Regulators and auditors need traceable logs, documentation, and evidence of policy enforcement. Trustworthy commerce AI therefore requires layered transparency rather than a single disclosure artifact.

Auditability is the next key requirement. AI audits in commerce should extend beyond model performance to include data lineage, privacy accounting, policy conformance, and dispute resolution. Existing governance work and emerging audit discussions indicate that standards boards, risk frameworks, and formalized assessment procedures are increasingly necessary if organizations are to move from principle statements to repeatable oversight (Tabassi, 2023). In digital commerce, audits should be able to address questions such as: Were conversion counts generated according to the documented privacy-preserving protocol? Were creators paid according to the disclosed revenue-sharing rule? Were moderation decisions applied consistently across comparable commercial content? Was model retraining triggered by valid and logged inputs? These are infrastructural audit questions, not merely algorithmic ones.

Responsibility allocation is also changing because privacy-preserving architectures can blur agency. When a recommendation outcome results from federated updates contributed by many participants, or when attribution depends on securely aggregated event streams, it can become harder to determine who is accountable for a harmful or misleading result. Governance must therefore define accountability despite distributed computation. One approach is to separate responsibility by function: participants remain responsible for data lawfulness at source; orchestrators are responsible for protocol integrity and model governance; service providers are responsible for API behavior and documentation; and deploying firms remain responsible for downstream business use. While such separation is imperfect, it is more realistic than treating "the AI system" as a singular responsible entity.

A rights-based perspective strengthens this analysis. Research on trustworthy AI in social media argues that rights language can clarify what is at stake when commercial platforms use AI to shape expression, exposure, and behavior (Lewis & Moorkens, 2020). In commerce, rights-based thinking helps connect privacy to due process, contestability, and freedom from manipulative opacity. This is especially important in creator ecosystems and SMB platform dependence, where weaker actors may be constrained by platform-defined rules they cannot meaningfully negotiate. Governance should thus include appeal mechanisms, explanation pathways, and contractual clarity for participants, not only consumer-facing privacy controls.

The practical standards landscape is becoming denser. NIST's AI RMF provides a flexible but structured reference for organizational risk management (Tabassi, 2023). The EU's trustworthy AI guidelines and assessment tools provide lifecycle-oriented self-assessment logic (HLEG, 2019, 2020). Reg-

ulatory developments such as the EU AI Act intensify the pressure to specify risk classification, technical documentation, and accountability arrangements, even if debate continues regarding the relation between trustworthiness and legal acceptability (Laux et al., 2024). For commerce platforms, the implication is clear: governance cannot remain an informal compliance appendix. It has to be embedded in architecture, documentation, vendor management, and interface design.

Several unresolved issues deserve emphasis. First, privacy and transparency can conflict. Strong privacy-preserving computation may reduce the amount of raw evidence available for audit or explanation. This increases the importance of protocol-level verification and carefully designed logs. Second, governance frameworks still under-address platform power asymmetries. A dominant platform may comply procedurally while imposing opaque economic dependencies on merchants and creators. Third, the connection between AI governance and content governance remains insufficiently theorized in commerce research. Commercial content ranking, sponsored recommendations, and creator monetization all blur the line between economic infrastructure and informational governance. Finally, most current governance frameworks are organization-centric, whereas digital commerce is ecosystemic. Future governance work must therefore address interoperability of audit artifacts, shared incident taxonomies, and cross-organizational accountability standards.

In summary, governance is not an external constraint on privacy-preserving commerce AI. It is the mechanism that determines whether privacy-preserving methods become trustworthy infrastructures or merely technical shields around opaque systems.

Conclusion

This review has argued that privacy-preserving and trustworthy AI in digital commerce should be understood as an infrastructure problem. Recommendation, advertising, cross-channel measurement, creator monetization, and SMB enablement all depend on the controlled circulation of data, model updates, metrics, and governance signals across organizational boundaries. Privacy constraints do not eliminate the need for these flows; they force them to be redesigned.

The literature shows that federated learning, differential privacy, secure aggregation, and zero-knowledge verification offer a meaningful technical toolkit, but their value depends on how they are composed. Federated learning without governance remains vulnerable. Differential privacy without scenario-sensitive metric design can produce unusable outputs. Secure aggregation without verifiability may protect data while undermining trust among business participants. Zero-knowledge methods without operational integration remain aspirational. The most promising direction is therefore modular composition anchored in scenario-specific requirements.

Three conclusions follow. First, cross-channel measurement has become the strategic bottleneck of privacy-preserving commerce AI. Without credible and privacy-aware measurement, firms cannot allocate budgets, evaluate creators, or refine recommendation and advertising systems. Second, social commerce and creator monetization reveal that trustworthiness is both consumer-facing and partner-facing. Privacy protection must coexist with auditable revenue logic, content governance, and role-specific transparency. Third, SMB adoption depends on whether trustworthy AI is embedded in multi-tenant platforms, standardized APIs, and governance-by-design rather than offered as a collection of advanced but inaccessible technical options.

Future research should therefore move in four directions. One direction is quantitative evaluation of privacy-utility trade-offs in realistic commerce workflows rather than in abstract benchmarks alone. A second is the development of privacy-preserving measurement systems that include uncertainty modeling, audit support, and incentive compatibility. A third is the study of ecosystem power under privacy-preserving architectures, especially whether such architectures decentralize control or simply repackage centralization. A fourth is the design of interoperable governance artifacts for platform ecosystems, including service cards, privacy ledgers, and standardized audit evidence. The next generation of digital commerce AI will be judged not only by whether it predicts well, but by whether it can produce commercially useful intelligence without undermining privacy, contestability, and accountability.

References

1. Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC conference on computer and communications security (pp. 308–318). https://doi.org/10.1145/2976749.2978318
2. Arnold, L., Jöhnk, J., Vogt, F., & Urbach, N. (2022). IIoT platforms' architectural features: A taxonomy and five prevalent archetypes. Electronic Markets, 32(2), 927–944. https://doi.org/10.1007/s12525-021-00520-0
3. Asad, M., Shaukat, S., Javanmardi, E., Nakazato, J., & Tsukada, M. (2023). A comprehensive survey on privacy-preserving techniques in federated recommendation systems. Applied Sciences, 13(10), Article 6201. https://doi.org/10.3390/app13106201
4. Bargoni, A., Giachino, C., Battisti, E., & Iaia, L. (2023). The effects of influencer endorsement services on crowdfunding campaigns. Journal of Services Marketing, 37(1), 40–52. https://doi.org/10.1108/JSM-12-2021-0444
5. Bi, N. C., & Zhang, R. (2023). "I will buy what my 'friend' recommends": The effects of parasocial relationships, influencer credibility and self-esteem on purchase intentions. Journal of Research in Interactive Marketing, 17(2), 157–175. https://doi.org/10.1108/JRIM-08-2021-0214
6. Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, B., Patel, S., Ramage, D., Segal, A., & Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. In Proceedings of the 2017 ACM SIGSAC conference on computer and communications security (pp. 1175–1191). https://doi.org/10.1145/3133956.3133982
7. Boerman, S. C., Kruikemeier, S., & Zuiderveen Borgesius, F. J. (2017). Online behavioral advertising: A literature review and research agenda. Journal of Advertising, 46(3), 363–376. https://doi.org/10.1080/00913367.2017.1339368
8. Carlson, J. R., Hanson, S., Pancras, J., Ross, W. T., & Rousseau-Anderson, J. (2022). Social media advertising: How online motivations and congruency influence perceptions of trust. Jour-

nal of Consumer Behaviour, 21(2), 197–213. https://doi.org/10.1002/cb.1989

9. Chaudhuri, R., Chatterjee, S., & Vrontis, D. (2023). Antecedents of privacy concerns and online information disclosure: Moderating role of government regulation. EuroMed Journal of Business, 18(3), 467–486. https://doi.org/10.1108/EMJB-11-2021-0181
10. Chen, Z., Gan, W., Wu, J., Hu, K., & Lin, H. (2025). Data scarcity in recommendation systems: A survey. ACM Transactions on Recommender Systems, 3(3), Article 27. https://doi.org/10.1145/3639063
11. Cheng, Y., Liu, Y., Chen, T., & Yang, Q. (2020). Federated learning for privacy-preserving AI. Communications of the ACM, 63(12), 33–36. https://doi.org/10.1145/3387107
12. Checco, A., Bianchi, G., & Leith, D. J. (2017). BLC: Private matrix factorization recommenders via automatic group learning. ACM Transactions on Privacy and Security, 20(2), Article 4. https://doi.org/10.1145/3041760
13. Corrêa, N. K., Galvão, C., Santos, J. W., Del Pino, C., Pontes Pinto, E., Barbosa, C., Massmann, D., Mambrini, R., Galvão, L., Terem, E., & de Oliveira, N. (2023). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. Patterns, 4(10), Article 100857. https://doi.org/10.1016/j.patter.2023.100857
14. Delaney, J., Ghazi, B., Harrison, C., Ilvento, C., Kumar, R., Manurangsi, P., Pál, M., Prabhakar, K., & Raykova, M. (2024). Differentially private ad conversion measurement. Proceedings on Privacy Enhancing Technologies, 2024(2), 124–140. https://doi.org/10.56553/popets-2024-0044
15. Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., López de Prado, M., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy artificial intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. Information Fusion, 99, Article 101896. https://doi.org/10.1016/j.inffus.2023.101896
16. Dinev, T., Xu, H., Smith, J. H., & Hart, P. (2013). Information privacy and correlates: An empirical attempt to bridge and distinguish privacy-related concepts. European Journal of Information Systems, 22(3), 295–316. https://doi.org/10.1057/ejis.2012.23
17. Dwivedi, Y. K., Ismagilova, E., Hughes, L., Carlson, J., Filieri, R., Jacobson, J., Jain, V., Karjaluoto, H., Kefi, H., Krishen, A. S., Kumar, V., Rahman, M. M., Raman, R., Rauschnabel, P. A., Rowley, J., Salo, J., Tran, G. A., & Wang, Y. (2021). Setting the future of digital and social media marketing research: Perspectives and research propositions. International Journal of Information Management, 59, Article 102168. https://doi.org/10.1016/j.ijinfomgt.2020.102168
18. Dwork, C. (2006). Differential privacy. In M. Bugliesi, B. Preneel, V. Sassone, & I. Wegener (Eds.), Automata, languages and programming (pp. 1–12). Springer. https://doi.org/10.1007/11787006_1
19. Farahat, A., Duda, C., Doshi, A., & Muthukrishnan, S. (2009). Privacy preserving frequency capping in internet banner advertising. In Proceedings of the 18th international conference on World Wide Web (pp. 1147–1148). https://doi.org/10.1145/1526709.1526900
20. Gao, C., Zheng, Y., Li, N., Li, Y., Qin, Y., Piao, J., Quan, Y., Chang, J., Jin, D., He, X., & Li, Y. (2023). A survey of graph neural networks for recommender systems: Challenges, methods, and directions. ACM Transactions on Recommender Systems, 1(1), 1–51. https://doi.org/10.1145/3568022
21. Gieß, A., & Hutterer, A. (2025). The future of data management: A delimitation of data platforms, data spaces, data meshes, and data fabrics. Information Systems and e-Business Management, 23(4), 971–997. https://doi.org/10.1007/s10257-025-00707-4
22. Gosselin, R., Vieu, L., Loukil, F., & Benoit, A. (2022). Privacy and security in federated learning: A survey. Applied Sciences, 12(19), Article 9901. https://doi.org/10.3390/app12199901
23. Gu, X., Sabrina, F., Fan, Z., & Sohail, S. (2023). A review of privacy enhancement methods for federated learning in healthcare systems. International Journal of Environmental Research and Public Health, 20(15), Article 6539. https://doi.org/10.3390/ijerph20156539
24. Hein, A., Schreieck, M., Riasanow, T., Setzke, D. S., Wiesche, M., Böhm, M., & Krcmar, H. (2020). Digital platform ecosystems. Electronic Markets, 30(1), 87–98. https://doi.org/10.1007/s12525-019-00377-4
25. High-Level Expert Group on Artificial Intelligence. (2019). Ethics guidelines for trustworthy AI. Publications Office of the European Union. https://doi.org/10.2759/346720
26. High-Level Expert Group on Artificial Intelligence. (2020). Assessment list for trustworthy artificial intelligence (ALTAI) for self-assessment. Publications Office of the European Union. https://doi.org/10.2759/002360
27. Jiang, Z., Heng, C. S., & Choi, B. C. F. (2013). Research note—Privacy concerns and privacy-protective behavior in synchronous online social interactions. Information Systems Research, 24(3), 579–595. https://doi.org/10.1287/isre.1120.0441
28. Jere, M., Farnan, T., & Koushanfar, F. (2021). A taxonomy of attacks on federated learning. IEEE Security & Privacy, 19(2), 20–28. https://doi.org/10.1109/MSEC.2020.3039941
29. Jordan, P., Mahdian, M., Vassilvitskii, S., & Vee, E. (2011). The multiple attribution problem in pay-per-conversion advertising. In M. Mavronicolas & V. V. Vazirani (Eds.), Algorithmic game theory (pp. 31–43). Springer. https://doi.org/10.1007/978-3-642-24829-0_5
30. Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Nitin Bhagoji, A., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G. L., Eichner, H., El Rouayheb, S., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P. B., … Zhao, S. (2021). Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 14(1–2), 1–210. https://doi.org/10.1561/2200000083
31. Kalloori, S., & Klingler, S. (2021). Horizontal cross-silo federated recommender systems. In Proceedings of the 15th ACM conference on recommender systems (pp. 680–684). https://doi.org/10.1145/3460231.3478863
32. Khan, A., ten Thij, M., & Wilbik, A. (2025). Vertical federated learning: A structured literature review. Knowledge and Information Systems, 67, 3205–3243. https://doi.org/10.1007/s10115-025-02356-y
33. Laux, J., Wachter, S., & Mittelstadt, B. (2024). Trustworthy artificial intelligence and the European Union AI act: On the conflation of trustworthiness and acceptability of risk. Regulation & Governance, 18(1), 3–32. https://doi.org/10.1111/rego.12512
34. Lewis, D., & Moorkens, J. (2020). A rights-based approach to trustworthy AI in social media. Social Media + Society, 6(3). https://doi.org/10.1177/2056305120954672
35. Libai, B., Rosario, A. B., Beichert, M., Donkers, B., Haenlein, M., Hofstetter, R., van der Lans, R., Lanz, A., Li, H. A., Mayzlin, D., Muller, E., Shapira, D., Yang, J., & Zhang, L. (2025). Influencer marketing unlocked: Understanding the value chains driving the creator economy. Journal of the Academy of Marketing Science, 53, 4–28. https://doi.org/10.1007/s11747-024-01073-2
36. Liyanaarachchi, G. P., Mifsud, M., & Viglia, G. (2024). Virtual influencers and data privacy: Introducing the multi-privacy paradox. Journal of Business Research, 176, Article 114584. https://doi.org/10.1016/j.jbusres.2024.114584
37. Luo, S., Xiao, Y., Zhang, X., Liu, Y., Ding, W., & Song, L. (2024). PerFedRec++: Enhancing personalized federated recommendation with self-supervised pre-training. ACM Transactions on

Intelligent Systems and Technology, 15(5), Article 98. https://doi.org/10.1145/3664927

38. Morimoto, M. (2021). Privacy concerns about personalized advertising across multiple social media platforms in Japan: The relationship with information control and persuasion knowledge. International Journal of Advertising, 40(3), 431–451. https://doi.org/10.1080/02650487.2020.1796322
39. Mothukuri, V., Parizi, R. M., Pouriyeh, S., Huang, Y., Dehghantanha, A., & Srivastava, G. (2021). A survey on security and privacy of federated learning. Future Generation Computer Systems, 115, 619–640. https://doi.org/10.1016/j.future.2020.10.007
40. Mukhopadhyay, S., & Bouwman, H. (2019). Orchestration and governance in digital platform ecosystems: A literature review and trends. Digital Policy, Regulation and Governance, 21(4), 329–351. https://doi.org/10.1108/DPRG-11-2018-0067
41. Papadopoulos, P., Abramson, W., Hall, A. J., Pitropakis, N., & Buchanan, W. J. (2021). Privacy and trust redefined in federated machine learning. Machine Learning and Knowledge Extraction, 3(2), 333–356. https://doi.org/10.3390/make3020017
42. Qin, J., Liu, B., & Qian, J. (2021). A novel privacy-preserved recommender system framework based on federated learning. In Proceedings of the 4th international conference on simulation, modeling and intelligent computing systems (pp. 82–88). https://doi.org/10.1145/3451471.3451485
43. Sharma, K., Lee, Y.-C., Nambi, S., Salian, A., Shah, S., Kim, S.-W., & Kumar, S. (2024). A survey of graph neural networks for social recommender systems. ACM Computing Surveys, 56(10), Article 265. https://doi.org/10.1145/3661821
44. Sun, X., Yu, F. R., Zhang, P., Sun, Z., Xie, W., & Peng, X. (2021). A survey on zero-knowledge proof in blockchain. IEEE Network, 35(4), 198–205. https://doi.org/10.1109/MNET.011.2000473
45. Sweeney, L. (2002). k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10(5), 557–570. https://doi.org/10.1142/S0218488502001648
46. Tabassi, E. (Ed.). (2023). Artificial intelligence risk management framework (AI RMF 1.0). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1
47. Teng, T., Li, H., Fang, Y., & Shen, L. (2022). Understanding the differential effectiveness of marketer versus user-generated advertisements in closed social networking sites: An empirical study of WeChat. Internet Research, 32(6), 1910–1929. https://doi.org/10.1108/INTR-04-2021-0268
48. Truong, N., Sun, K., Wang, S., Guitton, F., & Guo, Y. (2021). Privacy preservation in federated learning: An insightful survey from the GDPR perspective. Computers & Security, 110, Article 102402. https://doi.org/10.1016/j.cose.2021.102402
49. Wang, J., Arriaga, A., Tang, Q., & Ryan, P. Y. A. (2018). Facilitating privacy-preserving recommendation-as-a-service with machine learning and cryptographic techniques. In Proceedings of the 2018 ACM SIGSAC conference on computer and communications security (pp. 1900–1917). https://doi.org/10.1145/3243734.3278504
50. Wei, K., Li, J., Ding, M., Ma, C., Yang, H. H., Farokhi, F., Jin, S., Quek, T. Q. S., & Poor, H. V. (2020). Federated learning with differential privacy: Algorithms and performance analysis. IEEE Transactions on Information Forensics and Security, 15, 3454–3469. https://doi.org/10.1109/TIFS.2020.2988575
51. Wu, M., Li, L., Tao, C., Rigall, E., Wang, X., & Xu, C.-Z. (2022). FedCDR: Federated cross-domain recommendation for privacy-preserving rating prediction. In Proceedings of the 31st ACM international conference on information & knowledge management (pp. 2179–2188). https://doi.org/10.1145/3511808.3557320
52. Xue, J., & Liu, M. T. (2023). Investigating the live streaming sales from the perspective of the ecosystem: The structures, processes and value flow. Asia Pacific Journal of Marketing and Logistics, 35(5), 1157–1186. https://doi.org/10.1108/APJML-11-2021-0822
53. Xue, L., Song, P., Rai, A., Zhang, C., & Zhao, X. (2019). Implications of application programming interfaces for third-party new app development and copycatting. Production and Operations Management, 28(8), 1887–1902. https://doi.org/10.1111/poms.13021
54. Zhong, K., Ma, Y., & Angel, S. (2022). Ibex: Privacy-preserving ad conversion tracking and bidding. In Proceedings of the 2022 ACM SIGSAC conference on computer and communications security (pp. 3223–3237). https://doi.org/10.1145/3548606.3560651
55. Yi, X. (2026a). A federated and differentially private incentive–marketing framework for privacy-preserving cross-channel measurement in AI-powered digital commerce. Preprints, 202602.1929. https://doi.org/10.20944/preprints202602.1929.v1
56. Yi, X. (2026b). Trusted AI commercialization infrastructure for SMBs: A unified multi-tenant architecture integrating incentive systems, content governance, and standardized recommendation APIs. Preprints, 202602.1885. https://doi.org/10.20944/preprints202602.1885.v1
57. Yi, X. (2026c). Privacy-enhanced ad targeting for social e-commerce: A federated learning framework with zero-knowledge verification for creator monetization. Frontiers in Business and Finance, 3(1), 102–113. https://doi.org/10.71465/fbf653

Call for Papers

Scope

The Journal of Global Trends in Social Sciences is dedicated to publishing high-caliber scholarship defined by global vision, interdisciplinary synergy, and theoretical innovation. We prioritize original research that addresses frontier issues through deep analytical lenses, with a strong preference for transnational collaborative studies. Our areas of interest include, but are not limited to:

Core Disciplines

- **Communication & Media:** International/intercultural communication, digital media dynamics, and platform governance.
- **Surrounding Communication:** Regional communication networks, civilizational dialogue, and cultural exchange patterns.
- **Computational Social Science:** Computational communication, data journalism, algorithmic society, and AI ethics.

Interdisciplinary Frontiers

- **Technology, Policy & Society:** Digital governance, S&T policy, IP law, and data privacy.
- **Global Governance:** International relations, public diplomacy, and social policy.
- **Economic & Cultural Transformation:** Digital economy, cultural industries, and sustainable social innovation.

Strategic Focus: JGTSS particularly welcomes innovative methodologies that integrate social sciences with technological or engineering perspectives. Submissions should demonstrate theoretical depth and prospective insight, offering meaningful responses to significant real-world challenges and contributing to policy discourse. Every manuscript must reflect the journal's commitment to driving theoretical and methodological evolution within the context of global technological upheaval.

Thematic Section: Surrounding Communications Research

Guest Editor: Di Lu

Professor School of Journalism & Communication, Peking University, Beijing, China; Director, Center for Surrounding Communication Studies, Peking University

This section serves as a strategic cooperation platform with the Center for Surrounding Communication Studies, Peking University. It is dedicated to advancing the systematic development of China's original "Surrounding Communication Theory" and fostering global dialogue. The theory was first proposed by Professor Di LU in 2013 and has since been recognized as one of the "Four Major Indigenous Theories in Journalism and Communication Studies in China" and one of the "Top Ten Original Theories in Chinese Philosophy and Social Sciences." It provides an innovative analytical framework for understanding geopolitical information flows and cross-cultural communication mechanisms. It seeks to:

- Paradigm innovation and empirical studies in surrounding communication
- Comparative analyses of regional communication ecosystems
- Digital media's evolving role in geopolitical discourse
- Mechanisms of cross-border cultural diffusion
- Policy evaluation in transnational communication governance

Thematic Section:
Frontier Research in Intellectual Property Law

Guest Editor: Yanmin Quan

Professor, School of Law, Xi'an Jiaotong University; Council Member, Intellectual Property Law Research Committee of China Law Society; Council Member, China Law Society on Science and Technology Law

This section invites scholarly inquiry into the frontiers of Intellectual Property (IP) law, entertainment law, and the legal governance of traditional culture. It emphasizes the evolutionary trajectory of legal systems amidst the dual pressures of globalization and digital transformation. We welcome research that bridges theoretical legal frameworks with practical applications in industrial innovation and social governance. Key themes include, but are not limited to:

- IP protection and emergent legal challenges in the digital economy.
- Legal mechanisms for IP licensing, financing, and fostering industrial innovation.
- The platform economy, Generative AI (AIGC), and associated legal liabilities.
- Regulatory strategies and countermeasures against malicious IP litigation in cyberspace.

Submission Guidelines

JGTSS accepts the following manuscript types (all word counts include references and notes):

Research Articles (4,000–8,000 words): Original empirical or theoretical studies presenting novel findings and rigorous conceptual arguments.

Review Articles (4,000–8,000 words): Comprehensive and critical evaluations of major academic debates, research traditions, or emerging thematic directions.

Case Studies (4,000–6,000 words): In-depth analyses of specific projects, communities, or cultural practices with broader scholarly and theoretical implications.

Book Reviews (1,500–3,000 words): Analytical critiques of recently published monographs relevant to the journal's scope.

Note: Only original, unpublished works are considered. Translations of previously published material will not be accepted.

Peer review process

JGTSS adheres to a double-anonymized peer review model to ensure the highest standards of academic integrity. Each submission undergoes an initial internal screening by the editorial office to evaluate its alignment with the journal's scope and adherence to formatting requirements. Manuscripts that pass this preliminary stage are subsequently assigned to a minimum of two independent subject-matter experts for rigorous external evaluation.

Open access policy

JGTSS is a fully Gold Open Access publication. Upon acceptance, all articles are made immediately and permanently accessible online to a global audience without financial barriers. Detailed information regarding Creative Commons licensing, copyright retention, and institutional repository policies can be found in our comprehensive Policy Section.

Publication frequency & Timeliness

Schedule: JGTSS is published bimonthly, with full issues released in January, March, May, July, September and November.

Online First: To facilitate the rapid dissemination of research, accepted manuscripts are published as "Articles in Press" (Online First) individually. These versions appear online following the completion of peer review, editorial revision, and production, prior to their inclusion in a formal paginated issue.

Submission logistics

Format & Channel: Authors are required to prepare manuscripts in Microsoft Word (.doc/.docx) format. All submissions must be managed through the journal's Electronic Submission Portal.

Compliance: Before initiating a submission, authors must ensure their manuscript strictly complies with the latest Author Guidelines.

Rolling Basis: The journal maintains a continuous submission cycle, accepting manuscripts year-round without fixed deadlines for general issues.

Article processing charges (APC)

To support the costs of open-access publishing, JGTSS levies a one-time Article Processing Charge (APC) of USD 80 for each manuscript successfully accepted for publication.

Journal metadata & Contact

Submission Portal: https://jandoopress.com/journal/jgtss

Editorial Office: For all inquiries regarding the submission process or editorial decisions, please direct correspondence to contact@press.jandoo.ac.

Journal of Global Trends in Social Science (JGTSS) is an international, peer-reviewed, open-access venue committed to identifying and analyzing transformative global trends within the social sciences. JGTSS distinguishes itself by championing interdisciplinary synthesis, specifically fostering the nexus between social scientific inquiry and technological advancement.

JGTSS is dedicated to publishing high-caliber scholarship defined by global vision, interdisciplinary synergy, and theoretical innovation. We prioritize original research that addresses frontier issues through deep analytical lenses, with a strong preference for transnational collaborative studies.

Published by Jandoo Press Co., Ltd.

ISSN 2759-8764 (Online)
ISSN 2760-330X (Print)