The 8th Open Society Conference (OSC) 2026
Faculty of Law, Social and Political Science
Universitas Terbuka
DOI: 10.33830/osc.v4i1.9289 / ISSN: 3032-2227

Strategic Human–AI Communication on Social Media: Combating Disinformation, Rebuilding Civic Trust

Suhendra[1], Tiara Nesva Laraskana[2], Widya Permata Sari[3], Indira Sartika[4]

[1] Institut Pahlawan 12 Bangka, Bangka Belitung Islands, Indonesia

[2,3,4] STISIPOL Candradimuka Palembang, South Sumatra, Indonesia

Email: Suhendraa88@gmail.com

Abstract

The rapid integration of artificial intelligence (AI) into social media has transformed digital communication while intensifying the spread of disinformation and weakening civic trust. Although previous studies explored AI, social media, and misinformation separately, limited research examined their intersection through a strategic Human–AI communication perspective. This study aimed to develop a conceptual framework for Strategic Human–AI Communication to combat disinformation and rebuild civic trust in the digital era. A Systematic Literature Review (SLR) following the *PRISMA* 2020 guidelines was conducted using peer-reviewed articles indexed in Scopus and Web of Science published between 2019 and 2026. The selected studies were analyzed through thematic synthesis to identify key concepts, theoretical developments, and emerging research trends. The findings revealed four interrelated dimensions of Strategic Human–AI Communication: AI-assisted information verification, algorithmic transparency, human-centered digital literacy, and participatory communication governance. The review indicated that AI functioned most effectively as a collaborative partner supporting human judgment rather than replacing it. Transparent AI systems, ethical journalism, and collaborative fact-checking contributed significantly to strengthening information integrity and restoring civic trust. This study proposed an integrated conceptual framework that advances the literature on strategic communication and Human–AI interaction while providing practical insights for policymakers, journalists, media organizations, and digital platform developers to foster trustworthy and resilient digital communication ecosystems.

Keywords: *Civic Trust; Disinformation; Human–AI Communication; Social Media; Strategic Communication.*

INTRODUCTION

The convergence of artificial intelligence (AI) and social media has reshaped how societies produce, circulate, and consume information. Recommendation engines, generative language models, and automated content-moderation systems now mediate the vast majority of public discourse, positioning AI not merely as a passive infrastructure but as an active participant in communicative exchange [1]. This shift has generated a paradox: the same computational capabilities that promise faster fact-checking, personalized information delivery, and scalable content moderation also enable the rapid, low-cost production of hyper-realistic synthetic content, coordinated inauthentic behavior, and algorithmically amplified falsehoods [2]. Consequently, disinformation is no longer an isolated communicative anomaly but a systemic feature of the contemporary

Copyright © 2026, Suhendra, Tiara Nesva Laraskana, Widya Permata Sari, and Indira Sartika

DOI: 10.33830/osc.v4i1.9289 / ISSN: 3032-2227

digital public sphere, one that is increasingly authored, distributed, and in some cases detected by AI systems operating alongside human communicators [3].

The societal cost of this dynamic extends beyond the immediate circulation of false claims. Repeated exposure to manipulated or fabricated content, coupled with the perception that platforms and institutions are unable or unwilling to intervene, corrodes civic trust: the diffuse confidence citizens place in media organizations, government bodies, and fellow participants in public discourse [4]. Cross-national survey evidence shows that heavier reliance on social media as an information source is systematically associated with lower trust in institutions such as the European Union, even after controlling for infrastructural and demographic confounders [5], while broader political-communication scholarship documents how platform business models incentivize the very dynamics outrage, polarization, and epistemic fragmentation that erode democratic legitimacy [4]. At the same time, the expansion of participatory and deliberative digital platforms illustrates that technology-mediated engagement can, under the right institutional design, strengthen rather than weaken civic trust [6]. This tension between AI as a corrosive and a restorative force is the central problem this study addresses.

Existing scholarship has tended to examine AI, social media, and misinformation as separate research programs. Computer science and data-mining literatures concentrate on the technical performance of detection algorithms; journalism and communication studies examine editorial transparency and audience trust; and political science addresses the erosion of institutional legitimacy. What remains comparatively underdeveloped is a strategic communication perspective that treats the human–AI relationship itself rather than either party in isolation as the unit of analysis. The Human–Machine Communication (HMC) research agenda proposed the AI system as a communicator in its own right, requiring theory that accounts for the functional, relational, and ontological dimensions of people's interactions with communicative AI [1]. Building on this premise, the present study argues that combating disinformation and rebuilding civic trust requires a strategic communication architecture in which AI augments, rather than substitutes, human judgment, editorial responsibility, and civic deliberation.

This state of the art motivates the novelty and urgency of the present inquiry. First, generative AI has fundamentally altered the economics of disinformation production, lowering the cost and raising the realism of fabricated text, images, audio, and video, which existing detection and literacy frameworks were not designed to address at scale [2], [3]. Second, algorithmic curation systems remain largely opaque to the citizens whose information environments they shape, a condition that itself undermines the perceived legitimacy of digital platforms independent of the accuracy of the content they surface [7]. Third, civic trust does not automatically follow from technical accuracy; it is a relational outcome that depends on transparency, participation, and perceived fairness of the sociotechnical systems that govern public communication [6], [8]. Addressing these three conditions simultaneously requires an integrative, rather than a piecemeal, conceptual framework.

The urgency of this inquiry is amplified by the uneven distribution of institutional safeguards across different media systems. In consolidated democracies with well-resourced fact-checking coalitions and statutory platform-transparency obligations, AI-

Copyright © 2026, Suhendra, Tiara Nesva Laraskana, Widya Permata Sari, and Indira Sartika

DOI: 10.33830/osc.v4i1.9289 / ISSN: 3032-2227

assisted verification tools operate within a comparatively dense institutional support structure; in emerging and transitional democracies, by contrast, the same generative technologies often diffuse faster than the regulatory, journalistic, and educational capacity needed to govern them, leaving citizens more exposed to synthetic disinformation while offering fewer avenues for redress [2], [5]. This asymmetry means that a conceptual framework developed only from the vantage point of well-resourced media systems risks limited applicability elsewhere, reinforcing the need for a framework that specifies transferable design principles rather than context-bound technical solutions for combating disinformation and rebuilding civic trust across diverse institutional settings.

Prior reviews have made important contributions but typically isolate a single strand of this problem space. Systematized reviews of AI-driven disinformation detection catalog technical detection architectures and their vulnerabilities, yet devote comparatively little attention to the communicative and institutional conditions under which such tools translate into restored public trust [2]. Conversely, political-communication scholarship on institutional trust documents the erosion of civic confidence associated with social media use but rarely engages with the specific AI mechanisms recommendation algorithms, synthetic media, and automated moderation that mediate this erosion [4], [5]. By synthesizing evidence across these previously separate literatures within a single thematic framework, this review aims to close that integrative gap rather than to duplicate any single strand of prior work.

Accordingly, this study conducts a Systematic Literature Review (SLR) guided by the *PRISMA* 2020 protocol to synthesize peer-reviewed evidence on the intersection of AI, social media communication, disinformation, and civic trust, and to develop an integrated conceptual framework Strategic Human–AI Communication that articulates the conditions under which AI functions as a trustworthy communicative partner rather than a source of epistemic harm. The framework is intended to serve both a theoretical purpose, by extending strategic communication and Human–Machine Communication scholarship into the disinformation domain, and a practical purpose, by offering policymakers, journalists, media organizations, and platform developers a structured basis for designing more resilient digital communication ecosystems. Two research questions guided the review: (RQ1) What dimensions of Human–AI communication have been identified in the peer-reviewed literature as relevant to combating social media disinformation? and (RQ2) How do these dimensions relate to one another in shaping civic trust in digital communication ecosystems?

A further complication arises from the specific characteristics of generative AI. Unlike earlier, largely text-based misinformation, contemporary generative systems can produce synthetic images, audio, and video that are increasingly difficult for both automated classifiers and lay audiences to distinguish from authentic recordings, a development that has already reshaped how journalists and mass-communication scholars conceptualize disinformation risk [2]. This raises the strategic stakes for the Human–AI Communication perspective: because the same generative capability that fabricates convincing synthetic content can, when redirected, power the verification tools used to detect it, the net societal effect of generative AI depends heavily on which actors control its deployment and to what institutional standard they are held accountable. Business-

Copyright © 2026, Suhendra, Tiara Nesva Laraskana, Widya Permata Sari, and Indira Sartika

DOI: 10.33830/osc.v4i1.9289 / ISSN: 3032-2227

model critiques of contemporary social media platforms argue that, left unaddressed, commercial incentives tend to favor engagement-maximizing content curation over verification-supportive design, reinforcing the case for a framework that explicitly foregrounds governance and institutional accountability alongside technical capability [4].

METHOD

This study adopted a qualitative Systematic Literature Review (SLR) design, following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (*PRISMA* 2020) statement to ensure transparent identification, screening, and synthesis of the evidence base [7].

The review adapted the *PRISMA* 2020 reporting structure, whose explanation and elaboration guidance explicitly accommodates use across qualitative and mixed-methods evidence syntheses beyond clinical trials [8]. Two databases were searched—Scopus and Web of Science because of their broad coverage of peer-reviewed, internationally indexed communication, computer science, and political science journals. The search string combined controlled and free-text terms across three conceptual clusters connected with Boolean operators: (1) "artificial intelligence" OR "generative AI" OR "algorithm*"; (2) "disinformation" OR "misinformation" OR "fake news"; and (3) "social media" OR "digital platform*" AND "trust" OR "civic trust". The search was restricted to peer-reviewed journal articles and conference proceedings published in English between January 2019 and June 2026, a period chosen to capture scholarship following the initial wave of platform disinformation research and extending through the generative-AI era.

Inclusion criteria required that a study: (a) addressed AI or algorithmic systems in relation to social media communication; (b) examined disinformation, misinformation, or related information-integrity phenomena; and (c) reported implications for trust, credibility, or civic engagement. Studies were excluded if they addressed AI or misinformation without any social media or public communication dimension, were not peer-reviewed, or were not available in full text. As summarized in Table 1, the initial search identified 612 records; after duplicate removal and title–abstract screening, 88 full-text articles were assessed for eligibility, of which 42 met the inclusion criteria and were retained for synthesis.

Table 1. Screening and Selection Flow of the Systematic Literature Review

Screening Stage	Records (n)	Excluded (n)
Records identified through Scopus and Web of Science (2019–2026)	612	–
Records after duplicate removal	489	123
Records screened by title and abstract	489	401
Full-text articles assessed for eligibility	88	46
Studies included in the thematic synthesis	**42**	–

Copyright © 2026, Suhendra, Tiara Nesva Laraskana, Widya Permata Sari, and Indira Sartika

DOI: 10.33830/osc.v4i1.9289 / ISSN: 3032-2227

(Source: author's analysis, adapted from the *PRISMA* 2020 flow diagram)

For each included study, a structured data-extraction form was used to record the author(s) and year, country or media context, research design, unit of analysis, and reported findings relevant to AI, disinformation, or civic trust. Quality appraisal was conducted qualitatively, prioritizing studies published in Scopus- and Web of Science-indexed journals with transparent methodological reporting, consistent with the study's aim of building a conceptual rather than an effect-size-pooling synthesis. No formal meta-analytic pooling of quantitative effect sizes was undertaken, since the corpus combined qualitative, quantitative, and conceptual studies whose outcome measures were not directly comparable; this is consistent with the qualitative-synthesis orientation of the *PRISMA* 2020 guidance when applied outside strictly quantitative meta-analyses [7], [8].

The 42 included studies were analyzed using the thematic synthesis method of Thomas and Harden, which proceeds through three interrelated stages: line-by-line coding of findings, organization of codes into descriptive themes, and the generation of analytical themes that move beyond the primary studies to produce new interpretive constructs [9]. Coding was conducted manually and cross-checked between reviewers to enhance analytical consistency. Descriptive themes were first derived inductively from recurring concepts across the corpus such as verification, transparency, literacy, and governance before being elevated, through iterative comparison and discussion, into the four analytical dimensions reported in the Results and Discussion section. This thematic synthesis approach was selected because it preserves an explicit, auditable link between the conclusions of the review and the text of the primary studies while still permitting the generation of a novel conceptual framework, which a purely descriptive or narrative review would not support [9].

Methodological reflexivity was maintained throughout the review process. Because the search was restricted to English-language, peer-reviewed sources indexed in two commercial databases, the corpus is likely to overrepresent scholarship from well-resourced Anglophone and Western European research systems relative to non-English and grey literature documenting disinformation governance elsewhere; this constraint is addressed explicitly in the discussion of limitations. In addition, because the review sought conceptual integration rather than statistical pooling, no formal risk-of-bias score was assigned to individual studies; instead, methodological transparency and journal indexing status were used as proxies for source credibility, consistent with common practice in communication-studies systematic reviews that synthesize heterogeneous qualitative and quantitative evidence.

RESULTS AND DISCUSSION

The thematic synthesis of the 42 included studies converged on four interrelated dimensions that together constitute the proposed Strategic Human–AI Communication framework: (1) AI-assisted information verification, (2) algorithmic transparency, (3) human-centered digital literacy, and (4) participatory communication governance. Figure 1 depicts these four dimensions as complementary pillars that jointly support the outcome of combating disinformation and rebuilding civic trust; none of the dimensions was found

Copyright © 2026, Suhendra, Tiara Nesva Laraskana, Widya Permata Sari, and Indira Sartika

DOI: 10.33830/osc.v4i1.9289 / ISSN: 3032-2227

to be sufficient on its own, and their interaction rather than any single technical fix emerged as the recurring explanatory pattern across the reviewed literature.

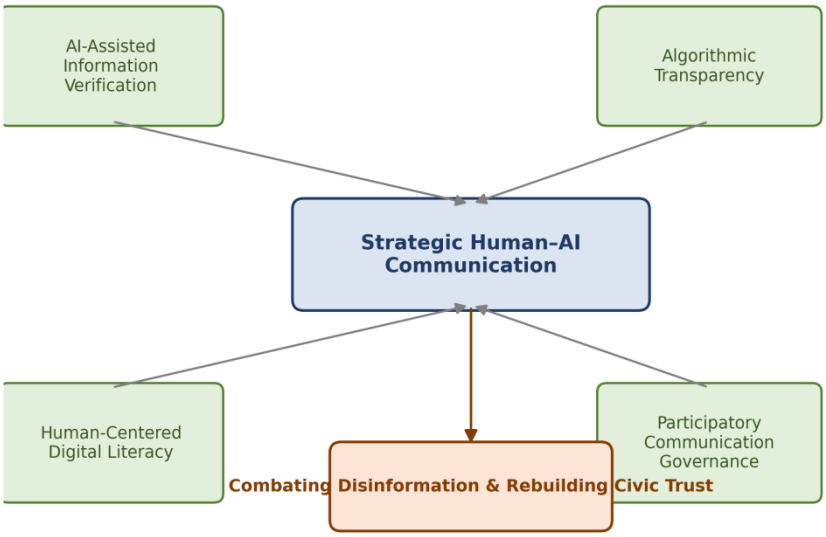


Figure 1. Conceptual Framework of Strategic Human–AI Communication

(Source: author's analysis)

AI-Assisted Information Verification. The first dimension concerns the use of AI as a tool that assists, rather than replaces, human fact-checking. Large-scale evaluations of machine-learning pipelines that combine natural language processing with human-curated training data show that hybrid systems trained on fact-checked datasets achieve markedly higher precision than models trained on unverified social media data alone, and that real-time performance still depends on continuous human oversight to contextualize ambiguous or sarcastic claims [10]. This pattern is reinforced by evidence that fact-checking itself is evolving from an ad hoc journalistic corrective into a form of epistemic infrastructure distributed across newsroom, platform, and civil-society actors, in which AI tools accelerate claim detection but institutional arrangements still determine whose verifications are trusted and how corrections reach affected audiences [11]. Analyses of automated information-disorder detection similarly emphasize that AI functions simultaneously as an authoritative verification agent and, when misused, as a spreader of misleading content, underscoring that the technology's net effect on information integrity depends on the governance arrangements surrounding its deployment rather than on its technical capability alone [3].

Nonetheless, the reviewed evidence also cautions against overstating AI's verification capacity. Detection systems evaluated under controlled benchmark conditions frequently underperform when applied to the rapidly evolving, context-dependent claims that circulate on live social media feeds, and adversarial actors can adapt content deliberately to evade classifiers trained on earlier disinformation patterns [10]. This asymmetry between benchmark and real-world performance is precisely why

Copyright © 2026, Suhendra, Tiara Nesva Laraskana, Widya Permata Sari, and Indira Sartika

DOI: 10.33830/osc.v4i1.9289 / ISSN: 3032-2227

the reviewed literature converges on hybrid designs: AI systems are best understood as a triage and prioritization layer that surfaces candidate claims for human fact-checkers, rather than as an autonomous adjudicator of truth, a division of labor that preserves interpretive nuance including sarcasm, cultural context, and contested framing that current models still struggle to resolve independently [10], [11].

Algorithmic Transparency. The second dimension addresses the disclosure of how algorithmic systems curate, rank, and moderate content. Foundational newsroom-based research distinguishes several layers at which disclosure can occur the underlying data, the model, the inference, and the user interface and finds that journalists and audiences alike see value in transparency but resist disclosure that overwhelms end users with technical detail [12]. Complementing this, critical accounts of algorithmic accountability caution that transparency should not be treated as an unqualified remedy: making a system's logic visible does not automatically make it comprehensible or contestable, particularly when disclosure is selective or when institutional incentives favor limited, strategic forms of openness [13]. Taken together, this literature suggests that algorithmic transparency contributes to civic trust only when it is calibrated to the audience's capacity to interpret it and is accompanied by accountability mechanisms that allow disclosed information to be acted upon, rather than treated as a one-way signal from platform to user.

This dimension also illuminates why transparency reforms sometimes fail to restore trust even when technically implemented. Where disclosure is designed primarily to satisfy regulatory compliance rather than genuine audience comprehension, citizens may perceive it as procedural rather than substantive, reproducing rather than resolving the opacity concerns transparency was meant to address [13]. Reconciling this tension requires treating algorithmic transparency as an ongoing communicative relationship between platforms and users one that is iteratively tested for comprehension and revised accordingly rather than as a static disclosure artifact appended to a platform's terms of service.

Human-Centered Digital Literacy. The third dimension concerns the capacity of citizens to critically evaluate AI-mediated information. Large-scale randomized evidence from real-world digital media literacy campaigns demonstrates that even brief, scalable interventions can measurably improve people's ability to discern between mainstream and false news, with effects observed across markedly different media and political contexts [14]. A recent meta-analysis spanning 49 experimental studies corroborates this finding at a broader scale, reporting a consistent positive effect of media literacy interventions on resilience to misinformation, an effect that strengthens with multiple-session designs and varies systematically with cultural and educational context [15]. These findings position digital literacy not as a substitute for platform-level intervention but as a necessary complement: even highly accurate AI verification systems cannot rebuild trust if citizens lack the interpretive skills to engage critically with the corrected information they receive.

At the same time, the literacy literature cautions against treating individual-level education as a sufficient remedy on its own. Digital literacy gains have been found to strengthen accuracy discernment more consistently than they reduce the likelihood that people share unverified content, suggesting that literacy programming addresses citizens'

Copyright © 2026, Suhendra, Tiara Nesva Laraskana, Widya Permata Sari, and Indira Sartika

DOI: 10.33830/osc.v4i1.9289 / ISSN: 3032-2227

cognitive evaluation of information more effectively than the social and emotional incentives that drive sharing behavior on social media [14]. This finding reinforces the rationale for treating human-centered digital literacy as one complementary pillar within the broader Strategic Human–AI Communication framework rather than as a stand-alone solution capable of resolving disinformation independently of platform design and governance interventions.

Participatory Communication Governance. The fourth dimension concerns the institutional arrangements through which citizens, platforms, and public authorities co-determine the rules governing digital communication. Case-study evidence from public-interest participatory platforms shows that deliberative principles inclusion, publicity, reason-giving, transparency, and traceability can be operationalized through concrete institutional design choices, and that such platforms function as public-interest infrastructures rather than purely technical products when governance is sustained through political commitment rather than software features alone [16]. Complementary work on direct and liquid forms of digital democracy similarly finds that participatory digital arrangements can mitigate the very disinformation and polarization risks associated with unmediated social media exposure, provided they are embedded in legal and constitutional safeguards [6]. This dimension reframes civic trust as something that is negotiated through ongoing participation rather than restored solely through top-down correction of false content.

The governance dimension also clarifies why some of the world's most extensive investments in fact-checking and platform disclosure have not fully arrested declining institutional trust: without a participatory mechanism through which citizens can contest, appeal, or co-design the rules governing their information environment, verification and transparency measures can be perceived as unilaterally imposed rather than legitimately negotiated [4], [16]. Fact-checking arrangements that function as recognized epistemic infrastructure, embedded within accountable platform-governance processes rather than operating as an external corrective, appear more capable of sustaining long-term public confidence than isolated verification services [11].

Table 2. Synthesis of the Four Dimensions of Strategic Human–AI Communication

Dimension	Locus of Intervention	Key Mechanism	Representative Sources
AI-assisted information verification	Content and claim analysis	Hybrid human–AI fact-checking pipelines	[3], [10], [11]
Algorithmic transparency	Platform disclosure design	Layered, audience-calibrated disclosure	[12], [13]
Human-centered digital literacy	Citizen capability	Repeated, context-adapted literacy interventions	[14], [15]
Participatory	Institutional design	Deliberative,	[6], [16]

Copyright © 2026, Suhendra, Tiara Nesva Laraskana, Widya Permata Sari, and Indira Sartika

DOI: 10.33830/osc.v4i1.9289 / ISSN: 3032-2227

communication governance	accountable platform governance

(Source: author's analysis)

Table 2 summarizes how each dimension addresses a distinct locus of intervention technical verification, platform disclosure, citizen capability, and institutional design while sharing the common outcome of supporting civic trust. This complementary structure explains why the reviewed studies converge on hybrid, human-in-the-loop solutions: a technically accurate verification system offers little protection if its outputs are not disclosed transparently, comprehensible only to already-literate citizens, or disconnected from participatory mechanisms through which citizens can contest or appeal automated decisions.

Integration and Theoretical Implications. Read together, the four dimensions support a central proposition of this study: Strategic Human–AI Communication functions most effectively when AI operates as a collaborative epistemic partner embedded within human-governed verification, disclosure, education, and participation structures, rather than as an autonomous replacement for any one of them. This reframes the widely discussed "human-in-the-loop" principle from a narrow technical safeguard into a strategic communication design requirement that spans newsroom workflows, platform interface design, public education curricula, and civic governance institutions simultaneously. Theoretically, the framework extends the Human–Machine Communication research agenda [1] by specifying the conditions under which communicative AI contributes to, rather than undermines, the broader information ecosystem's legitimacy. It also responds to calls in the disinformation literature for integrated sociotechnical responses that combine literacy, regulation, and strengthened journalism, rather than relying on any single computational or legal intervention in isolation [2], [3].

Practical Implications. For policymakers, the findings suggest that regulation should target the interaction between AI systems and the governance structures surrounding them for example, mandating layered transparency disclosures and funding independent verification infrastructure rather than regulating algorithms or content in isolation. For journalists and media organizations, the results reinforce the value of hybrid verification workflows in which AI tools triage and prioritize claims while trained journalists retain final editorial judgment, consistent with evidence that fully automated fact-checking has not displaced human epistemic authority even where it improves throughput [10], [11]. For digital platform developers, the transparency literature implies that interface-level disclosure should be designed iteratively with user comprehension testing rather than treated as a compliance checkbox [12], [13]. Finally, for educators and civil-society organizations, the digital literacy evidence supports investment in repeated, contextually adapted media literacy programming rather than one-off awareness campaigns [14], [15].

Copyright © 2026, Suhendra, Tiara Nesva Laraskana, Widya Permata Sari, and Indira Sartika

DOI: 10.33830/osc.v4i1.9289 / ISSN: 3032-2227

CONCLUSION

This study set out to develop a conceptual framework for Strategic Human–AI Communication capable of addressing the dual challenge of combating disinformation and rebuilding civic trust in AI-mediated social media environments. Through a PRISMA-guided systematic literature review and thematic synthesis of 42 peer-reviewed studies, the analysis identified four interrelated dimensions AI-assisted information verification, algorithmic transparency, human-centered digital literacy, and participatory communication governance that jointly determine whether AI functions as a trustworthy communicative partner or as a source of further epistemic harm. The review's central conclusion is that AI contributes most reliably to information integrity and civic trust when it augments human judgment within transparent, participatory, and educationally supported institutional arrangements, rather than when it operates as a stand-alone technical solution.

These findings carry implications for both theory and practice. Theoretically, the study extends strategic communication and Human–Machine Communication scholarship by offering an integrated framework specific to the disinformation and civic trust domain, addressing a gap left by research traditions that have tended to treat AI, social media, and misinformation as separate objects of inquiry. Practically, the framework offers policymakers, journalists, media organizations, and platform developers a structured basis for designing hybrid verification workflows, layered transparency mechanisms, sustained digital literacy programming, and participatory governance arrangements that reinforce rather than compete with one another.

Beyond its immediate empirical scope, the Strategic Human–AI Communication framework contributes a vocabulary that allows researchers and practitioners across disciplines—computer science, journalism studies, and political communication—to locate their specific interventions within a shared conceptual structure, potentially reducing the current fragmentation in which technical, editorial, and institutional responses to disinformation are designed and evaluated in isolation from one another.

This study is not without limitations. As an SLR restricted to English-language, peer-reviewed literature indexed in Scopus and Web of Science, the review may underrepresent non-English scholarship and practitioner or policy documents that also shape disinformation governance in practice. The proposed framework is furthermore conceptual and has not yet been empirically tested across specific platforms, political systems, or cultural contexts. Future research should therefore pursue comparative empirical validation of the Strategic Human–AI Communication framework—for example, through cross-national surveys of algorithmic transparency perceptions, field experiments on hybrid human–AI fact-checking workflows, and longitudinal studies of participatory governance platforms to assess how the four dimensions interact under varying regulatory, cultural, and technological conditions, and ultimately to strengthen the resilience of digital communication ecosystems worldwide.

Copyright © 2026, Suhendra, Tiara Nesva Laraskana, Widya Permata Sari, and Indira Sartika

DOI: 10.33830/osc.v4i1.9289 / ISSN: 3032-2227

REFERENCES

[1] A. L. Guzman and S. C. Lewis, "Artificial intelligence and communication: A Human–Machine Communication research agenda," New Media Soc., vol. 22, no. 1, pp. 70–86, 2020, doi: 10.1177/1461444819858691.
[2] C. von Sikorski and M. Hameleers, "Disinformation in the age of artificial intelligence (AI): Implications for journalism and mass communication," Journalism Mass Commun. Q., 2025, doi: 10.1177/10776990251375097.
[3] A. Tomassi, A. Falegnami, and E. Romano, "Mapping automatic social media information disorder: The role of bots and AI in spreading misleading information in society," PLoS ONE, vol. 19, no. 5, p. e0303183, 2024, doi: 10.1371/journal.pone.0303183.
[4] I. Castelló, E. Colleoni, A. G. Scherer, and H. Trittin-Ulbrich, "Social media is a threat for democracy! A political perspective for analysing and diminishing harm," J. Manage. Stud., 2026, doi: 10.1111/joms.70053.
[5] O. S. Kiratli, "Social media effects on public trust in the European Union," Public Opin. Q., 2023, doi: 10.1093/poq/nfad029.
[6] P. Rattanasevee, Y. Akarapattananukul, and Y. Chirawut, "Direct democracy in the digital age: Opportunities, challenges, and new approaches," Humanit. Soc. Sci. Commun., vol. 11, art. 1681, 2024, doi: 10.1057/s41599-024-04245-1.
[7] M. J. Page et al., "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," BMJ, vol. 372, p. n71, 2021, doi: 10.1136/bmj.n71.
[8] M. J. Page et al., "PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews," BMJ, vol. 372, p. n160, 2021, doi: 10.1136/bmj.n160.
[9] J. Thomas and A. Harden, "Methods for the thematic synthesis of qualitative research in systematic reviews," BMC Med. Res. Methodol., vol. 8, art. 45, 2008, doi: 10.1186/1471-2288-8-45.
[10] G. Gondwe, "Can AI outsmart fake news? Detecting misinformation with AI models in real-time," Emerging Media, vol. 3, no. 2, pp. 252–274, 2025, doi: 10.1177/27523543251325902.
[11] D. Shin, D. D. Perlmutter, J. S. Lim, and K. Jitkajornwanich, "Fact-checking as epistemic infrastructure: From journalistic gatekeeping to platform governance," Emerging Media, 2025, doi: 10.1177/27523543251344972.
[12] N. Diakopoulos and M. Koliska, "Algorithmic transparency in the news media," Digit. Journalism, vol. 5, no. 7, pp. 809–828, 2017, doi: 10.1080/21670811.2016.1208053.
[13] M. Ananny and K. Crawford, "Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability," New Media Soc., vol. 20, no. 3, pp. 973–989, 2018, doi: 10.1177/1461444816676645.
[14] A. M. Guess, M. Lerner, B. Lyons, J. M. Montgomery, B. Nyhan, J. Reifler, and N. Sircar, "A digital media literacy intervention increases discernment between mainstream and false news in the United States and India," Proc. Natl. Acad. Sci. U.S.A., vol. 117, no. 27, pp. 15536–15545, 2020, doi: 10.1073/pnas.1920498117.
[15] G. Huang, W. Jia, and W. Yu, "Media literacy interventions improve resilience to misinformation: A meta-analytic investigation of overall effect and moderating factors," Commun. Res., 2024, doi: 10.1177/00936502241288103.
[16] J. C. Correia, "Alternative digital platforms and the renewal of the public sphere: Decidim and the democratic governance of participatory infrastructures," Soc. Sci., vol. 15, no. 3, art. 166, 2026, doi: 10.3390/socsci15030166.

Copyright © 2026, Suhendra, Tiara Nesva Laraskana, Widya Permata Sari, and Indira Sartika