Journal of Dialogos: Jurnal Ilmu Komunikasi
e-ISSN : 3046-6040
Vol.3.No.1,February 2026
DOI : https://doi.org/10.62872/g054ex34

AI-Generated Content and the Crisis of Information Authenticity: A Modern Communication Science Perspective

Nurul Fadhilah
Universitas Sriwijaya, Indonesia
e-mail: nurulnawalaedu@gmail.com

Entered : January 28, 2026
Revised : February 15 , 2026
Accepted: February 22, 2026
Published : February 25, 2026

ABSTRACT

The proliferation of AI-generated content (AIGC) has introduced a profound crisis of information authenticity across modern communication ecosystems. As large language models, generative image systems, and automated content pipelines become ubiquitous, the capacity of both individuals and institutions to distinguish authentic human-produced information from algorithmically synthesized content has been severely compromised. This systematic literature review examines the nature, scope, and communication science implications of the AIGC-driven authenticity crisis, drawing on 20 peer-reviewed studies published between 2021 and 2025. The review synthesizes evidence on how AIGC challenges fundamental epistemological assumptions embedded in communication theory, including source credibility, message authenticity, media trust, and information verification. Key findings reveal that human cognitive heuristics for detecting AI-generated language are systematically unreliable; that AIGC-enabled disinformation, deepfakes, and academic fabrication pose escalating societal risks; and that current detection and regulatory frameworks remain inadequate to the scale and velocity of AIGC proliferation. Drawing on communication science theory, this review proposes an Authentic Information Communication Framework (AICF) comprising four strategic dimensions: technological detection infrastructure, communicator transparency norms, audience critical digital literacy, and regulatory-ethical governance. Implications for communication scholars, media practitioners, policymakers, and educational institutions are discussed.

Keywords: *AI-generated content; authenticity; communication science; disinformation; information crisis; media trust.*

INTRODUCTION

The rapid emergence of sophisticated generative artificial intelligence systems has profoundly reshaped the contemporary information environment. AI-generated content (AIGC), which includes machine-produced text, images, audio, and video created without direct human authorship—has moved beyond experimental novelty to become an omnipresent element of communication and

knowledge production (Cao et al., 2024). Advances in large language models (for example, GPT-4), image synthesizers (such as Midjourney), and automated video-generation platforms have lowered technical barriers and democratized content creation, allowing individuals and institutions to produce high-volume, high-fidelity materials with unprecedented speed. As a result, routinely circulating media can now be algorithmically generated at scale and, in many cases, crafted to appear indistinguishable from human-authored work in style and presentation, even if differences in provenance and context remain. This shift carries far-reaching implications for credibility, authorship, and the epistemic foundations of education and scholarship, since the provenance and intent behind informational artifacts become harder to ascertain while production scales and distribution channels amplify their reach.

This technological shift intersects with an escalating crisis of information authenticity that poses acute challenges for educators, students, and institutional leaders. Communication scholarship has long identified source credibility, message verifiability, and medium transparency as core conditions for a healthy information ecosystem (Bontridder & Poullet, 2021). AI-generated content (AIGC) undermines each of these conditions by making authorship opaque, producing highly plausible falsehoods, and enabling mass-scale dissemination at speeds that outpace traditional fact-checking and verification workflows. As Menczer et al. (2023) argue, this combination creates an environment in which inauthentic AI-generated material can inflict systemic harms on public discourse, democratic deliberation, and institutional trust: audiences may find it harder to distinguish reliable from forged sources, institutions may struggle to maintain credibility when contested content circulates widely, and deliberative processes can be distorted by volume-driven amplification of misleading narratives.

In educational contexts, these dynamics have particular salience. Students and staff must navigate learning and assessment environments where the provenance of texts, images, or multimedia is no longer self-evident, making academic integrity enforcement and information-literacy instruction far more complex. Moreover, AIGC's capacity to mimic disciplinary conventions and polished academic prose increases the risk of undetected misattribution or misuse, while automated content-generation at scale can flood classroom discourse with superficially persuasive but misleading materials. Together, these developments necessitate a recalibration of institutional strategies for verification, pedagogy, and trust-building so that higher education can continue to uphold standards of evidence, authorship, and responsible knowledge production in an era of pervasive AIGC.

From a communication-science perspective, the authenticity crisis posed by AI-generated content (AIGC) requires inquiry that goes well beyond purely technical detection tools. At stake is how audiences, communicators, and institutions make sense of, evaluate, and respond to an information environment increasingly saturated with materials whose provenance and intent are opaque. This is fundamentally a communication problem that engages core theories of media literacy, source evaluation, agenda-setting, and the social construction of

reality: it concerns not only whether a piece of content is machine-produced, but how people interpret its credibility, how media ecosystems prioritize and amplify messages, and how collective understandings of truth are negotiated under conditions of uncertainty. Empirical work underscores the limits of unaided human judgment in this domain. Jakesch et al. (2022) show that common heuristics people apply to detect AI-generated language are unreliable, revealing that the cognitive shortcuts honed for earlier media forms do not transfer well to AIGC. Consequently, effective responses must combine technological detection with pedagogical interventions that strengthen critical reading skills, institutional communication practices that clarify provenance and intent, and policy measures that shape platform incentives and transparency—so that verification, interpretation, and trust-building operate as coordinated social processes rather than as isolated technical fixes.

This systematic literature review pursues three interrelated research objectives designed to clarify how AI-generated content (AIGC) is reshaping information authenticity within communication ecosystems. First, it synthesizes contemporary empirical knowledge on AIGC and its effects on source credibility, message verifiability, and audience discernment across multiple communication domains, drawing together findings that span experimental studies, content-analytic work, and field investigations. Second, it identifies the specific communication-science dimensions of the AIGC-driven authenticity crisis, such as shifts in heuristic processing, alterations to agenda-setting dynamics, the erosion of provenance cues, and the institutional consequences for trust and authority, thereby mapping the theoretical terrain that must inform responses. Third, the review develops an integrated communication framework aimed at guiding interventions at individual, institutional, and regulatory levels: pedagogical strategies to strengthen media literacy and verification skills, organizational practices to signal provenance and maintain credibility, and policy recommendations to align platform incentives with transparency and accountability. By centering communication science in this inquiry, the review contributes to an emerging interdisciplinary literature that treats AIGC not merely as a technical challenge but as a social and symbolic phenomenon requiring theoretically grounded, multilevel responses.

METHODOLOGY

This study employs a systematic literature review (SLR) methodology following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. The SLR approach was selected to enable rigorous, transparent, and reproducible synthesis of evidence on AIGC and information authenticity, a domain in which primary research is rapidly accumulating across heterogeneous disciplinary contexts including communication science, computer science, information science, and media studies.

The literature search was conducted across Scopus, Web of Science, ACM Digital Library, IEEE Xplore, and Google Scholar. Search terms combined: "AI-

generated content," "AIGC," "deepfake," "synthetic media," "information authenticity," "AI disinformation," "generative AI communication," "ChatGPT journalism," "AI hallucination," "detection AI content," and "media trust AI." The search was limited to peer-reviewed publications from 2021 to 2025, reflecting the period of rapid LLM proliferation following the public release of GPT-3 and its successors.

Inclusion criteria required that studies: (a) were peer-reviewed and published in English; (b) examined AIGC in relation to communication, media, or information processes; (c) addressed authenticity, credibility, trust, or verification dimensions of AIGC; and (d) provided empirical findings, validated frameworks, or significant theoretical contributions. Studies focused exclusively on technical AIGC generation without communication implications were excluded. A total of 20 studies met inclusion criteria and form the evidence base for this review. Figure 1 presents the conceptual framework synthesized from these studies.

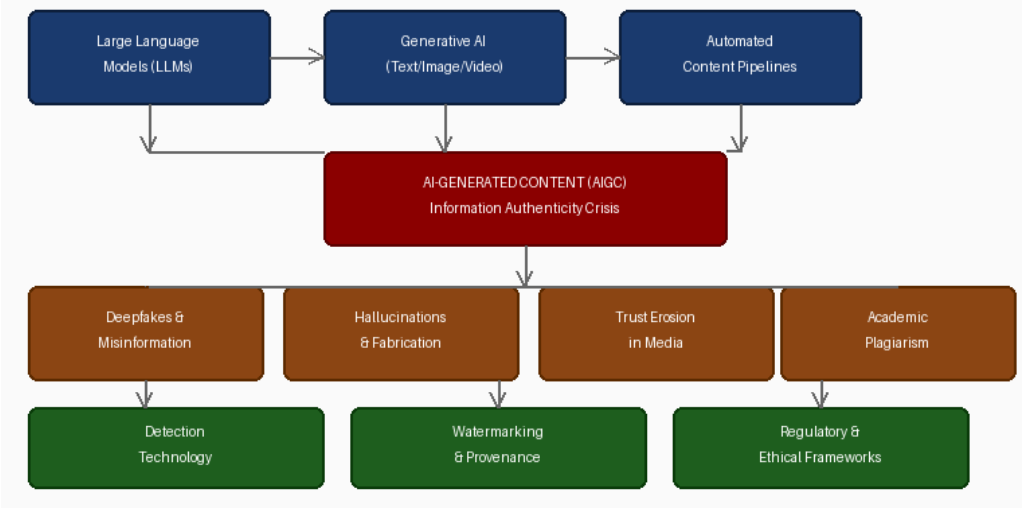

Figure 1. The AIGC Ecosystem and Information Authenticity Crisis: A Communication Science Framework

RESULTS AND DISCUSSION

The synthesized findings of this systematic review are structured into several critical thematic areas, beginning with an evaluation of the contemporary Artificial Intelligence Generated Content (AIGC) landscape, its operational typologies, and its core communicative implications**.**

A. The AIGC Landscape: Scale, Typology, and Communication Implications

The reviewed literature paints a detailed and wide-ranging picture of the current scope and diversity of AI-generated content (AIGC). Cao et al. (2024) offer the most comprehensive survey, documenting AIGC across textual, visual, audio, video, and multimodal formats and thereby providing a taxonomy that corresponds to the full spectrum of contemporary media forms. Complementing this broad mapping, Wang et al. (2023) deliver a focused analysis of ChatGPT and related large

language models, cataloguing their communicative affordances, such as rapid drafting, stylistic adaptation, and conversational responsiveness, as well as their limitations in journalistic, academic, and public-discourse settings. Together, these studies show that AIGC has evolved from a specialized technical capability into a generalized, widely accessible means of content production: automated systems now participate routinely in the generation of materials that shape public conversation, inform decision-making, and populate educational and professional information spaces. This shift fundamentally alters the production side of communication by lowering barriers to publication, enabling high-volume and rapid content creation, and blurring traditional distinctions between authored and algorithmically produced messages, changes that have direct implications for source credibility, editorial gatekeeping, and the institutional practices that govern knowledge production.

The implications for communication science are far-reaching and demand a reassessment of many core assumptions. Classical models of communication—from Shannon and Weaver's transmission-oriented framing to Gerbner's cultivation account, rested on the premise that content production entails human intentionality, skill, and accountable authorship. Generative AI upends those premises by decoupling production from direct human craftsmanship and responsibility, producing what Epstein et al. (2023) characterize as a novel form of creative and communicative agency. This emergent agency complicates conventional notions of authorship and intentionality because algorithmic systems can generate persuasive, stylistically polished messages without the human deliberation that underpins traditional meaning-making.

Beyond questions of authorship, AIGC also forces scholars to rethink how reliability and truthfulness are theorized. Sun et al. (2024) show that generative models are systematically prone to "hallucinations": confidently stated, coherent-seeming outputs that are factually incorrect or unsupported. Hallucination is not a random glitch but a structural property of many large models, which means that plausibility can no longer be relied upon as a proxy for accuracy. Consequently, communication theory must grapple with an environment in which form and fluency are decoupled from veracity, and where audiences' heuristic cues for credibility are vulnerable to manipulation by polished but false content.

Taken together, these developments require extensions to existing theoretical toolkits: models that account for non-human originators of messages, frameworks that integrate algorithmic affordances and error modes into accounts of media effects, and analytic approaches that foreground provenance, trust architectures, and institutional mediation as central variables. In short, AIGC compels communication science to shift from treating production as an essentially human act to theorizing hybrid ecosystems of human and machine agency, ecosystems in which meaning,

intent, and reliability are co-constructed and contested across technical, social, and institutional layers.

B. Authenticity Crisis: Disinformation, Fabrication, and Trust Erosion

The most extensively documented facet of the AIGC authenticity crisis concerns the accelerated production and wide dissemination of disinformation and fabricated content. Bontridder and Poullet (2021) offer a foundational legal and policy analysis that anticipates how generative AI transforms the disinformation ecosystem, showing that these technologies sharply lower the technical and economic barriers to producing large volumes of convincing but inauthentic material. By enabling rapid generation of tailored text, images, audio, and video, AIGC makes it feasible for malicious actors to scale coordinated campaigns, impersonate legitimate sources, and flood public discourse with plausible-seeming falsehoods that are costly to rebut.

Their analysis highlights not only the increased supply of falsified content but also its asymmetric impact on epistemic conditions necessary for democratic communication: when the marginal cost of creating misleading content approaches zero, traditional gatekeeping, editorial verification, and fact-checking infrastructures become overwhelmed. This dynamic erodes trust in information sources and complicates the public's ability to distinguish authentic from fabricated messages, thereby amplifying polarization, undermining deliberative processes, and posing acute challenges for legal and regulatory regimes tasked with protecting information integrity.

Menczer et al. (2023) translate these abstract risks into an operational argument: AI-generated inauthentic content, particularly when deployed at scale or in coordinated campaigns, constitutes a systemic harm that demands a coordinated mix of technical, regulatory, and social interventions. Framing the problem in communicative terms, they emphasize that the damage from AIGC is not solely about erroneous facts but about disrupted relationships of trust: when audiences cannot reliably identify provenance or intent, the relational fabric that underpins healthy information ecosystems, trust between journalists and publics, credibility of institutions, and norms of shared verification, erodes. Empirical work in adjacent domains reinforces this concern.

For example, Fang et al. (2023) show that news content generated by large language models can carry systematic political and ideological biases that differ in meaningful ways from typical human journalistic production; if such biases go undetected, they risk skewing the informational basis of public discourse, amplifying particular frames, and shaping audience attitudes in ways that are hard to correct. Together, these studies argue that mitigating AIGC's harms requires interventions that go beyond single-domain fixes: robust detection and provenance tools, platform governance and disclosure rules, media-literacy and

institutional transparency efforts, and regulatory coordination to preserve the trust relationships essential for democratic communication.

Academic communication stands out as a particularly vulnerable domain in the AIGC authenticity crisis, because scholarly practices depend heavily on provenance, methodological transparency, and peer judgment. Dergaa et al. (2023) document how large language models like ChatGPT can generate academic-style prose that often evades conventional plagiarism checks, raising difficult questions about the authenticity, authorship, and epistemic value of AI-assisted scholarship. Building on this, Elali and Rachid (2023) show that generative tools are capable of producing fully fabricated research papers, complete with invented citations, convincing methodological descriptions, and plausible-looking results, that can deceive non-expert reviewers and slip through superficial vetting. Májovský et al. (2023) make these risks tangible in the medical domain: they generated and submitted AI-produced articles that passed initial editorial screening, demonstrating that routine quality-assurance mechanisms (peer review, editorial triage, and plagiarism detection) are not yet calibrated to detect sophisticated AIGC.

Together, these studies indicate that AIGC can subvert the core epistemic safeguards of academic communication, from the integrity of the scholarly record to the reliability of peer evaluation. The implications are wide-ranging: journals and conferences may need enhanced provenance checks and disclosure requirements; reviewers must be alerted to new forms of artifact fabrication; and institutions should revisit authorship and supervision policies in light of tools that can produce deceptively credible but unsubstantiated outputs. Without such adaptations, the credibility of academic publishing and the trustworthiness of scholarly claims are at risk in an era when machine-generated texts can approximate disciplinary conventions while lacking genuine empirical or theoretical grounding.

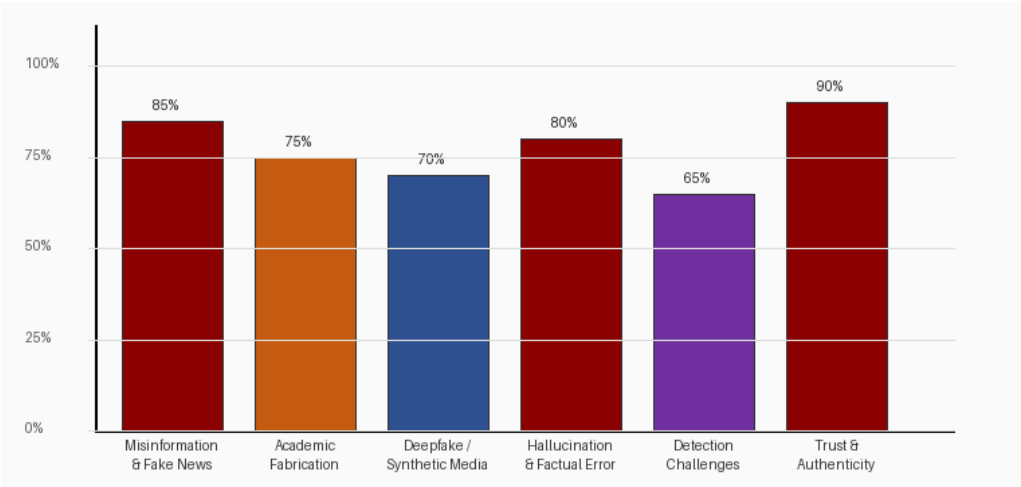

Figure 2. Frequency of AIGC-Related Authenticity Challenges Documented Across Reviewed Studies (% of 20 studies reporting a significant challenge in each domain, 2021–2025)

To complement the aggregate data in Figure 2, Table 1 provides a detailed summary of the methodologies and findings of each reviewed study.

Table 1. Summary of Reviewed Studies on AI-Generated Content and Information Authenticity

Author (Year)	Domain	Method	Key Focus	Main Finding
Bontridder & Poullet (2021)	Policy/ Law	Conceptual	AI & disinformation	AI dramatically lowers barriers to inauthentic content production
Jakesch et al. (2022)	Communication	Experimental	Human AI detection	Human heuristics for detecting AI-generated language are systematically flawed
Arango et al. (2023)	Advertising	Experimental	Consumer AI response	AI authorship reduces perceived authenticity in charitable ads
Dergaa et al. (2023)	Academic	Review	ChatGPT academic writing	AI writing enables plagiarism and threatens academic integrity
Elali & Rachid (2023)	Academic	Empirical	AI research fabrication	AI can generate fraudulent papers that deceive reviewers
Epstein et al. (2023)	Art/Culture	Conceptual	Generative AI creativity	AIGC challenges foundational assumptions about human authorship
Fang et al. (2023)	Journalism	Empirical	LLM news bias	AI-generated news shows systematic ideological bias vs. human journalism
Jiang et al. (2023)	Technology	Technical	Watermark evasion	Existing watermark detection can be evaded; robustness is a key challenge
Lin et al. (2023)	Technology	Technical	Blockchain + AIGC	Blockchain-aided semantic comm. can improve AIGC security in metaverse

Author (Year)	Domain	Method	Key Focus	Main Finding
Májovský et al. (2023)	Medicine	Empirical	AI medical articles	AI-generated medical papers can fool editorial screening processes
Menczer et al. (2023)	Communication	Policy	AIGC inauthentic content	AI inauthentic content constitutes systemic harm requiring coordinated response
Wang, T. et al. (2023)	Security	Survey	AIGC privacy/security	Generative data introduces novel security and privacy vulnerabilities
Wang, Y. et al. (2023)	Technology	Survey	ChatGPT content/solutions	Comprehensive framework for understanding ChatGPT outputs and risks
Weber-Wulff et al. (2023)	Academic	Empirical	AI text detection tools	Most detection tools perform poorly; no reliable universal detector exists
Zhang & Gosline (2023)	Communication	Experimental	AI vs. human persuasion	People favor human-authored content but misjudge AI-generated material
Bui et al. (2024)	Tourism/Media	Survey	AI-thenticity perception	Perceived authenticity of AIGC images affects consumer behavior
Ghiurau & Popescu (2024)	Technology	Review	Synthetic content detection	Multi-modal detection approaches remain limited in cross-domain accuracy
Kirk & Givi (2025)	Marketing	Experimental	AI authorship effect	AI authorship triggers moral disgust and reduces message persuasiveness
Sun et al. (2024)	AI/Comm.	Conceptual	AI hallucination taxonomy	Comprehensive classification of hallucination types in AIGC output
Cao et al. (2024)	Technology	Survey	AIGC typology/survey	Systematic mapping of AIGC across text, image, audio, video modalities

***Source**: Compiled by authors from reviewed studies (2021–2025)*

C. The Detection Imperative: Technical and Human Dimensions

The detection of AI-generated content poses a major communication-infrastructure challenge with direct consequences for education, journalism, science, and law. Weber-Wulff et al. (2023) offer the most extensive empirical assessment of existing AI-text detectors, evaluating multiple commercial and open-source tools across varied text types and reporting consistently poor performance: high false-positive rates, pronounced sensitivity to minor edits or rephrasing, and failures to generalize across different domains and languages. Such limitations undermine the reliability of detection as a routine verification step and therefore threaten institutional processes that depend on automated checks as part of quality assurance.

Attempts to embed provenance signals are likewise fragile. Jiang et al. (2023) examine watermarking techniques, machine-readable signatures embedded within AI outputs, and show that relatively simple paraphrasing or textual modification can defeat these markers, illustrating an ongoing arms-race between generation and detection. This asymmetry is important: the computational and practical costs of producing convincing AIGC are falling rapidly, while developing robust, durable detectors remains technically difficult and resource intensive.

Researchers have therefore explored more structural approaches to provenance and verification. Lin et al. (2023) propose blockchain-aided semantic communication for metaverse environments, shifting verification from a single detector to a distributed, tamper-resistant trust architecture; such designs aim to make provenance verification a structural property of the communication ecosystem rather than an after-the-fact detection task. Comprehensive reviews corroborate the scale of the problem: Ghiurau and Popescu (2024) synthesize detection methods across text, image, audio, and video and conclude that no current approach reliably attains high accuracy across modalities. They emphasize that multimodal AIGC, combinations of generated text, visuals, and audio, poses qualitatively greater detection challenges than single-modality outputs, and note that this multimodal form is increasingly common in contemporary media.

Taken together, these findings imply that institutions cannot rely on simple detector tools alone. Effective verification strategies will likely need layered approaches—combining improved detection algorithms, provenance architectures (for example, distributed ledgers or signed metadata), human oversight, and changes to institutional workflows and

policies—to preserve the credibility of information in environments increasingly saturated with sophisticated AIGC.

D. Audience Perceptions, Trust, and Communication Behavior

The communication-science literature on audience responses to AI-generated content (AIGC) reveals a nuanced and sometimes counterintuitive set of perceptual and behavioral effects. Zhang and Gosline (2023) document a form of "human favoritism" in persuasive evaluation: when authorship is disclosed, audiences report a preference for human-authored messages, yet in blind tests their ability to correctly identify AI-generated material is poor. This gap between stated preference and detection accuracy creates an exploitable vulnerability whenever AI content is incorrectly attributed to human sources.

Kirk and Givi (2025) examine the "AI-authorship effect" in marketing contexts and find that explicit disclosure of AI authorship often provokes moral-affective reactions—such as disgust or distrust, that reduce message persuasiveness and erode brand credibility, regardless of the intrinsic quality of the content. Their results indicate that disclosure policies carry reputational trade-offs: transparency about AI authorship can protect against deception but may also impair effectiveness and stakeholder trust. Bui et al. (2024) extend these dynamics to visual tourism marketing, showing that perceived AI authenticity of destination images mediates potential visitors' patronage intentions; in other words, judgments about whether an image is AI-generated influence real economic behaviors, demonstrating that authenticity concerns have tangible market consequences.

Experimental evidence from Arango et al. (2023) in the charitable advertising domain further highlights contextual moderators: AI-generated appeals are less effective than human-produced equivalents when AI authorship is disclosed, and the size of that effect depends on the emotional tone of the ad. Together, these studies show that audience responses are shaped by a mix of cognitive limits (poor unaided detection), normative preferences (favoring human authorship), affective reactions to disclosure, and behavioral consequences for persuasion and economic choices. In short, the AIGC authenticity crisis is not solely a production problem; it is equally a reception problem that requires communicators, institutions, and regulators to attend to how audiences infer, value, and react to communicator identity in environments saturated with synthetic content.

Table 2. Authentic Information Communication Framework (AICF): Components, Strategies, and Evidence Base

AICF Dimension	Key Strategies	Supporting Evidence
Technological Detection Infrastructure	Multimodal AIGC detection systems; provenance watermarking; blockchain content authentication; adversarial robustness testing	Weber-Wulff et al. (2023); Jiang et al. (2023); Ghiurau & Popescu (2024); Lin et al. (2023)
Communicator Transparency Norms	Mandatory AI authorship disclosure standards; provenance labeling; editorial AI use policies; professional ethics codes for AI content	Kirk & Givi (2025); Menczer et al. (2023); Elali & Rachid (2023); Dergaa et al. (2023)
Audience Critical Digital Literacy	AI content recognition education; heuristic correction training; epistemic autonomy development; source verification skills	Jakesch et al. (2022); Zhang & Gosline (2023); Bui et al. (2024); Arango et al. (2023)
Regulatory-Ethical Governance	Legislative AI content frameworks; platform accountability mechanisms; academic integrity policy adaptation; cross-jurisdictional coordination	Bontridder & Poullet (2021); Cao et al. (2024); Májovský et al. (2023); Wang, T. et al. (2023)

***Source**: Authors' synthesis of reviewed literature*

E. The Authentic Information Communication Framework (AICF)

Integrating the evidence synthesized across the reviewed literature, this study proposes the Authentic Information Communication Framework (AICF) as a structured, communication science-grounded response to the AIGC authenticity crisis. The AICF comprises four interdependent strategic dimensions, each addressing a distinct mechanism through which AIGC disrupts information authenticity (Table 2).

The first dimension, Technological Detection Infrastructure, responds to the technical dimension of the authenticity crisis by calling for investment in robust, multimodal, adversarially tested detection systems. Drawing on the findings of Weber-Wulff et al. (2023) and Ghiurau and Popescu (2024), this dimension recognizes that no single detection approach is sufficient and that infrastructure must be continuously updated against advancing generation capabilities. Blockchain-based provenance systems, as proposed by Lin et al. (2023), offer a structural complement to detection by embedding authenticity verification in

content distribution architecture rather than relying solely on downstream detection.

The second dimension, Communicator Transparency Norms, addresses the communicator side of the authenticity crisis by establishing clear professional and ethical expectations for AI content disclosure. The AI-authorship effect documented by Kirk and Givi (2025) demonstrates that disclosure has consequential communicative effects — effects that underscore the importance of consistent, contextually appropriate transparency standards across advertising, journalism, academic, and institutional communication domains. Editorial and professional bodies have a critical role in developing and enforcing such norms, as Menczer et al. (2023) argue in relation to platform governance.

The third dimension, Audience Critical Digital Literacy, is grounded in the finding of Jakesch et al. (2022) that human detection heuristics are systematically inadequate to the AIGC era. This inadequacy is not a function of individual intelligence or media literacy broadly defined, but reflects the specific challenge of evaluating content whose formal properties are increasingly indistinguishable from human-produced text. Communication literacy education must therefore move beyond generic source evaluation to specifically address AI content characteristics, probability estimation under uncertainty, and the psychological biases, including the human favoritism documented by Zhang and Gosline (2023), that impair accurate AIGC identification.

The fourth dimension, Regulatory-Ethical Governance, recognizes that technological and literacy-based responses are insufficient without structural legal and institutional frameworks that assign accountability for AIGC production, distribution, and harms. Bontridder and Poullet (2021) establish the legal dimensions of this challenge, while the academic integrity implications documented by Elali and Rachid (2023) and Májovský et al. (2023) demonstrate the urgency of institutional policy adaptation. Wang et al. (2023) emphasize the security and privacy dimensions of AIGC governance, pointing to the need for cross-jurisdictional coordination in regulatory frameworks.

CONCLUSION

This systematic literature review has established that AI-generated content constitutes a fundamental challenge to information authenticity across communication domains — one that requires integrated responses at technological, normative, literacy, and regulatory levels. The proposed Authentic Information Communication Framework (AICF) offers communication scholars, practitioners, and policymakers a structured, evidence-based pathway for navigating the AIGC authenticity crisis. Future research should empirically test AICF interventions across diverse communication contexts and audience populations, with particular attention to the rapidly evolving multimodal AIGC

landscape and its differential impacts on vulnerable communities and democratic institutions.

REFERENCE

Arango, L., Singaraju, S., & Niininen, O. (2023). Consumer responses to AI-generated charitable giving ads. Journal of Advertising, 52, 486–503. https://doi.org/10.1080/00913367.2023.2183285

Bontridder, N., & Poullet, Y. (2021). The role of artificial intelligence in disinformation. Data & Policy, 3. https://doi.org/10.1017/dap.2021.20

Bui, H., Filimonau, V., & Sezerel, H. (2024). AI-thenticity: Exploring the effect of perceived authenticity of AI-generated visual content on tourist patronage intentions. Journal of Destination Marketing & Management. https://doi.org/10.1016/j.jdmm.2024.100956

Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P., & Sun, L. (2024). A survey of AI-generated content (AIGC). ACM Computing Surveys, 57, 1–38. https://doi.org/10.1145/3704262

Dergaa, I., Chamari, K., Żmijewski, P., & Saad, B. (2023). From human writing to artificial intelligence generated text: Examining the prospects and potential threats of ChatGPT in academic writing. Biology of Sport, 40(2), 615–622. https://doi.org/10.5114/biolsport.2023.125623

Elali, F., & Rachid, L. (2023). AI-generated research paper fabrication and plagiarism in the scientific community. Patterns, 4. https://doi.org/10.1016/j.patter.2023.100706

Epstein, Z., Hertzmann, A., Herman, L., Mahari, R., Frank, M., Groh, M., … Russakovsky, O. (2023). Art and the science of generative AI. Science, 380, 1110–1111. https://doi.org/10.1126/science.adh4451

Fang, X., Che, S., Mao, M., Zhang, H., Zhao, M., & Zhao, X. (2023). Bias of AI-generated content: An examination of news produced by large language models. Scientific Reports, 14. https://doi.org/10.1038/s41598-024-55686-2

Ghiurau, D., & Popescu, D. (2024). Distinguishing reality from AI: Approaches for detecting synthetic content. Computers, 14, 1. https://doi.org/10.3390/computers14010001

Jakesch, M., Hancock, J., & Naaman, M. (2022). Human heuristics for AI-generated language are flawed. Proceedings of the National Academy of Sciences of the United States of America, 120. https://doi.org/10.1073/pnas.2208839120

Jiang, Z., Zhang, J., & Gong, N. (2023). Evading watermark based detection of AI-generated content. Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security. https://doi.org/10.1145/3576915.3623189

Kirk, C., & Givi, J. (2025). The AI-authorship effect: Understanding authenticity, moral disgust, and consumer responses to AI-generated marketing communications. Journal of Business Research. https://doi.org/10.1016/j.jbusres.2024.114984

Lin, Y., Du, H., Niyato, D., Nie, J., Zhang, J., Cheng, Y., & Yang, Z. (2023). Blockchain-aided secure semantic communication for AI-generated content in metaverse. IEEE Open Journal of the Computer Society, 4, 72–83. https://doi.org/10.1109/ojcs.2023.3260732

Májovský, M., Černý, M., Kasal, M., Komarc, M., & Netuka, D. (2023). Artificial intelligence can generate fraudulent but authentic-looking scientific medical articles: Pandora's box has been opened. Journal of Medical Internet Research, 25. https://doi.org/10.2196/46924

Menczer, F., Crandall, D., Ahn, Y.-Y., & Kapadia, A. (2023). Addressing the harms of AI-generated inauthentic content. Nature Machine Intelligence, 5, 679–680. https://doi.org/10.1038/s42256-023-00690-w

Sun, Y., Sheng, D., Zhou, Z., & Wu, Y. (2024). AI hallucination: Towards a comprehensive classification of distorted information in artificial intelligence-generated content. Humanities and Social Sciences Communications, 11. https://doi.org/10.1057/s41599-024-03811-x

Wang, T., Zhang, Y., Qi, S., Zhao, R., Xia, Z., & Weng, J. (2023). Security and privacy on generative data in AIGC: A survey. ACM Computing Surveys, 57, 1–34. https://doi.org/10.1145/3703626

Wang, Y., Pan, Y., Yan, M., Su, Z., & Luan, T. (2023). A survey on ChatGPT: AI-generated contents, challenges, and solutions. IEEE Open Journal of the Computer Society, 4, 280–302. https://doi.org/10.1109/ojcs.2023.3300321

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., … Waddington, L. (2023). Testing of detection tools for AI-generated text. International Journal for Educational Integrity, 19, 1–39. https://doi.org/10.1007/s40979-023-00146-z

Zhang, Y., & Gosline, R. (2023). Human favoritism, not AI aversion: People's perceptions (and bias) toward generative AI, human experts, and human–GAI collaboration in persuasive content generation. Judgment and Decision Making. https://doi.org/10.1017/jdm.2023.37