Original Article
ISSN (Online): 3048-9245

ShodhAI: Journal of Artificial Intelligence
July-December 2025 2(2), 1–14

THE INFLUENCE OF AI-GENERATED CONTENT ON TRUST AND CREDIBILITY WITHIN SPECIALIZED ONLINE COMMUNITIES: A BRIEF REVIEW ON PROPOSED CONCEPTUAL FRAMEWORK

Mostafa Essam Ahmed Eissa [1]

[1] Freelance Independent Researcher and Consultant, India

Received 15 June 2025
Accepted 29 July 2025
Published 08 August 2025

Corresponding Author
Mostafa Essam Ahmed Eissa,
mostafaessameissa@yahoo.com

DOI 10.29121/ShodhAI.v2.i2.2025.40

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

The increasing prevalence of Artificial Intelligence (AI) in creating content signifies a notable change in the digital communication landscape. While the broader effects on widespread media platforms have been extensively discussed, the specific consequences within specialized online communities remain less explored. These communities, frequently founded and established on shared interests, mutual confidence, and perceived genuineness, are particularly susceptible to alterations in the origin and trustworthiness of content. This paper challenges three questions: (1) How AI content affects credibility perceptions, (2) Verification methods used by communities, (3) Consequences for trust dynamics. A hypothetical framework would be used to investigate the potential impact of AI-produced content on the dynamics of trust and credibility within these focused digital environments. By drawing upon existing academic work in media studies, the behavior of online communities, and the concept of source credibility, a theoretical model and outline a potential research strategy were encouraged to examine how the presence, identification, and interpretation of content authored by AI might modify member interactions, processes for verifying information, and the overall unity of the community. The hypothetical outcome suggests that the subtle integration of AI content could diminish perceived authenticity, complicate established indicators of trust, and potentially lead to the fragmentation or decline of communities that depend on authentic human connection and collective expertise. The article concludes by considering the ramifications for those who manage communities, design platforms, and participate as members, stressing the importance of greater openness and digital literacy in navigating the evolving digital media landscape.

Keywords: AI-Generated Content, Specialized Online Communities, Trust, Credibility, Digital Media, Online Interaction, Community Behavior

1. INTRODUCTION

The digital era has brought about an unprecedented surge in the creation and distribution of content, fundamentally reshaping how individuals access information, connect with others, and form communities. Among the various forms of digital assembly, specialized online communities stand out as crucial arenas where individuals with common interests, hobbies, identities, or professional focuses gather Rheingold (1993), Wellman and Gulia (1999). It was noted that unlike more general social media platforms, these communities often flourish

through deep involvement, the exchange of expert knowledge, and a strong sense of belonging rooted in reciprocal trust and the perceived authenticity among participants Lin et al. (2017).

Concurrently, advancements in Artificial Intelligence (AI), particularly in the areas of natural language processing and generative models, have reached a significant point, enabling the creation of sophisticated textual, visual, auditory, and video content that is increasingly difficult to distinguish from content produced by humans Brown et al. (2020), Radford et al. (2019). The incorporation of AI into content creation processes, whether for generating articles, forum discussions, social media updates, or creative works, is becoming more prevalent.

Although the implications of AI-generated content for journalism, marketing, and public discourse are subjects of increasing academic and public interest, its specific effect on the delicate ecosystems of specialized online communities requires focused examination Makki and Jawad (2023), Labajová (2023), Oksymets (2024). These communities heavily rely on the trustworthiness of the information shared by members and the confidence built through consistent, genuine interactions Kozyreva et al. (2020). The introduction of content with an uncertain origin, potentially automated, or designed to subtly sway opinion could pose a substantial threat to the foundational principles upon which these communities are established.

This paper aims to explore the intricate relationship between AI-generated content, trust, and credibility within specialized online communities. We propose that the distinctive characteristics of these communities render them particularly vulnerable to the disruptive potential of AI content. Unlike general online spaces where interactions may be superficial, specialized communities often involve higher stakes in terms of shared knowledge, emotional investment, and social support, making the integrity of information and the genuineness of contributors of paramount importance.

The primary questions guiding this investigation are Dwivedi et al. (2023): How does the presence of AI-generated content affect members' perceptions of credibility within specialized online communities? What methods do communities employ to verify information in the age of advanced AI content, and how effective are these methods? And what are the broader consequences for community dynamics, the development of trust, and long-term viability?

By addressing these questions, this article will aim to contribute to a more profound understanding of the evolving digital media landscape and offer insights relevant to researchers, community managers, platform developers, and users navigating the challenges and opportunities presented by AI in online interactions. The following sections will review pertinent literature, propose a theoretical framework, outline a hypothetical research methodology and potential findings, discuss the implications, and conclude with suggestions for future research and practice.

2. LITERATURE REVIEW

To comprehend the impact of AI-generated content on trust and credibility in specialized online communities, it is necessary to draw upon several distinct but interconnected bodies of academic work Burtch et al. (2023): trust and credibility in digital environments, the nature and behaviour of specialized online communities, and the emerging research on AI-generated content.

2.1. TRUST AND CREDIBILITY IN ONLINE ENVIRONMENTS

Trust and credibility are essential for effective communication and social interaction, both offline and online Fogg (2003), Flanagin and Metzger (2017). In digital spaces, however, traditional indicators for evaluating trustworthiness and credibility, such as physical presence, non-verbal cues, and established organizational affiliations, are often absent or modified. Users must depend on alternative signals, including the reputation of the source (if known), the quality and consistency of the content, social validation (likes, shares, comments), and the norms and reputation of the platform or community itself Flanagin and Metzger (2000), Sundar (2008).

Credibility is frequently understood as having two components: trustworthiness (perceived honesty, integrity, and goodwill) and expertise (perceived knowledge, skill, and competence) O'Keefe (2015). In online communities, these components are often assessed based on a member's history of contributions, the value and accuracy of their posts, their willingness to assist others, and their adherence to community rules and values Ridings et al. (2002). Trust, conversely, is a willingness to be vulnerable based on positive expectations about another's conduct Rousseau et al. (1998). In online communities, trust develops over time through repeated positive interactions and shared experiences Flavián et al. (2006).

The difficulties in evaluating trust and credibility online are intensified by the ease with which identities can be concealed or invented, and information can be manipulated or disseminated rapidly Lankes (2007), Bryce and Fraser (2014), Lazer et al. (2018). The rise of misinformation and disinformation campaigns underscores the fragility of trust in digital ecosystems Shin et al. (2018).

2.2. SPECIALIZED ONLINE COMMUNITY DYNAMICS

Specialized online communities are distinct from broader social networks due to their concentrated subject matter, often smaller scale, higher levels of member involvement, and stronger social ties among participants (Blanchard & Markus, 2004, Preece (2000). Members are typically brought together by a deep, shared interest or identity, which fosters a sense of belonging and collective identity Ardichvili et al. (2003).

The exchange of information is a central function of many specialized communities, ranging from sharing practical advice and technical knowledge to discussing shared passions and offering emotional support Lave and Wenger (1991), Nonaka and Takeuchi (1996). The credibility of this information is often judged not solely on external sources but also on the perceived expertise and trustworthiness of fellow community members who have demonstrated their knowledge and commitment over time Toral et al. (2009). Systems of reputation, whether explicit or implicit, play a vital role in indicating which members are dependable sources of information and support Kollock (1999).

These communities frequently develop unique norms, specialized vocabulary, and social protocols that govern interactions and the sharing of information Ren et al. (2007). These norms function as a form of social regulation, helping to maintain the quality of discourse and reinforce community values. The perceived authenticity of contributions – whether they originate from a genuine, invested member – is often highly valued and can be a key factor in building trust.

2.3. EMERGING RESEARCH ON AI-GENERATED CONTENT

Research into AI-generated content is a rapidly evolving domain, focusing on its technical capabilities, ethical considerations, and societal impacts Broussard (2020), Floridi and Chiriatti (2020). Studies have investigated the capacity of AI models to produce persuasive text Gallagher et al. (2022), Weber et al. (2024), create synthetic media ("deepfakes") Vaccari and Chadwick (2020), and automate content production for various purposes Lai and Nissim (2024).

A primary challenge highlighted in this research is the increasing difficulty in differentiating AI-generated content from content created by humans, particularly as models become more advanced Bommasani et al. (2021). Thus, this raises concerns about the potential for misuse, including the spread of false information, manipulation of public opinion, and the erosion of trust in digital sources of information Cinus et al. (2025). Recent work by Ferrara (2023)on bot detection and Pennycook et al. (2021) and Pennycook and Rand (2022) on misinformation literacy highlights AI's role in trust erosion.

While some research has begun to examine how AI-generated news articles or social media posts are perceived by general populations Marinescu et al. (2022), there is a notable gap in understanding how such content is received and impacts trust specifically within the context of specialized online communities, where the dynamics of credibility and trust are often more subtle and deeply rooted in interpersonal relationships and shared identity.

2.4. IDENTIFYING THE GAP

Existing literature provides a solid foundation for understanding trust, credibility, and the dynamics of online communities. However, the specific intersection of these concepts with the emergence of sophisticated AI-generated content, particularly within the specialized context of niche online communities, remains largely unexplored Basta (2024). How do the unique norms, trust mechanisms, and information validation processes of these communities cope with content that may lack genuine human experience or intention? Importantly, this gap highlights the necessity for focused research to understand the potential vulnerabilities and adaptations of these vital digital spaces in the age of generative AI.

3. CONCEPTUAL FRAMEWORK

To analyse the impact of AI-generated content on trust and credibility within specialized online communities, a conceptual framework was proposed that integrates elements from Source Credibility Theory, Social Presence Theory, and the specific characteristics of specialized communities. The framework suggests that the impact of AI-generated content is influenced by several mediating factors Figure 1:

1) **Content Attributes:** This includes the quality, relevance, style, and perceived authenticity of the AI-generated content. High-quality, relevant, and stylistically appropriate AI content may be more challenging to detect and thus more likely to influence perceptions, at least initially. Content that subtly deviates from community norms or language patterns might raise suspicion.

2) **Detection of AI Origin:** The capacity of community members to discern whether content was produced by AI is a crucial mediator. Detection can occur through explicit labelling (uncommon), linguistic indicators (e.g., overly general language, absence of personal anecdotes, unusual phrasing), or inconsistencies with previous posts from the same "user" (if the AI content is posted under a profile that appears human).
3) **Perception of AI Content:** Regardless of actual detection, members will form perceptions about the content's origin and purpose. Is it viewed as helpful, manipulative, spam, or simply novel? These perceptions are shaped by prior experiences, individual biases, and community norms regarding automation or external tools.
4) **Community Norms and Trust Mechanisms:** Existing community norms concerning the sharing of information, verification, and interaction play a vital role. Communities with strong norms against spam or inauthentic contributions, and robust mechanisms for questioning dubious content or users, may demonstrate greater resilience. Conversely, communities lacking such structures may be more susceptible.
5) **Source Credibility Evaluation:** When encountering potentially AI-generated content, members engage in evaluating the source's credibility. However, if the "source" is a seemingly human profile posting AI content, this evaluation becomes complicated. Members may assess the content itself (based on perceived expertise and trustworthiness cues within the text) and the profile (based on posting history, interactions, etc.), potentially leading to inconsistency if the content feels unnatural but the profile appears established.
6) **Transparency of AI Origin** (e.g., labelled vs. unlabelled content).
7) **Impact on Trust and Credibility:** The interaction of the aforementioned factors ultimately affects trust and credibility. If AI content is not detected and is perceived as valuable, it might initially enhance the perceived credibility of the posting profile. However, if detected or perceived negatively, it can severely damage the credibility of the source and potentially erode trust within the community as a whole, particularly if the issue is widespread or poorly managed by moderators. The perceived authenticity of interactions is key here; AI content, lacking genuine human experience, may undermine the sense of connection essential to specialized communities (Social Presence Theory).

This framework suggests that the impact is not direct but depends on how AI content is created, how easily it is detected, how it is perceived by members, and the existing social infrastructure of the community.

Figure 1

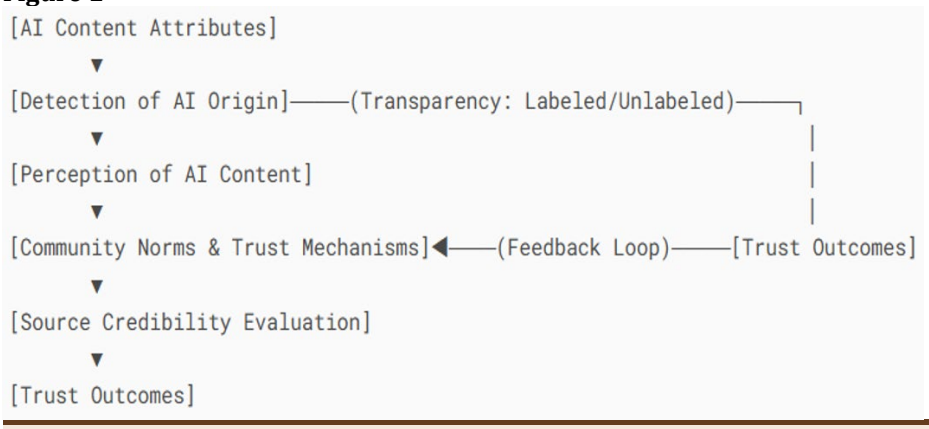

Figure 1 Conceptual Framework Linking AI-Generated Content Attributes to Trust Outcomes: Visual Representation of the Conceptual Framework Linking AI-Generated Content Attributes to Trust Outcomes. Secondary Interactions (Transparency and Feedback Loops) are Highlighted.

4. HYPOTHETICAL METHODOLOGY

To investigate the impact of AI-generated content on trust and credibility within specialized online communities, a mixed-methods approach combining qualitative analysis of community discussions with quantitative surveys could be utilized.

4.1. PROPOSED FRAMEWORK

This step is basically could be started by choosing 2-3 diverse specialized online communities (e.g., a hobbyist forum, a professional discussion group, a support community) with active participation and established norms. Criteria for selection would include community size, subject area, moderation practices, and platform type. Communities stratified by size (<1k, 1k–10k, >10k), moderation strictness (low/high), and platform type (forum, Discord, subreddit) could be used as selection criteria.

4.2. DATA COLLECTION

This step is important and critical in terms in terms of the quality and reliability of sampling technique coupled with measures that ensure avoidance of introduction of unintended biases in the results with subsequent analysis, interpretation, and conclusion inaccuracies and distortion.

4.2.1. QUALITATIVE DATA

This step starts by gathering publicly available text data (posts, comments, reactions) from selected communities over a defined period (e.g., 6-12 months). Focus on threads or discussions where the presence or suspicion of AI-generated content is apparent, or where conversations about content authenticity, trust, or credibility arise. Collection of data must ethically, respecting community guidelines and user privacy, potentially anonymizing contributions.

4.2.2. QUANTITATIVE DATA

Developing a survey instrument targeting active members of the selected communities is conducted at this step. Pilot testing will assess Cronbach's alpha (α >0.7 required for scale reliability). The survey would assess:

- Members' awareness and perceptions of AI-generated content.
- Their reported strategies for evaluating content credibility.
- Their levels of trust in other members and the community as a whole.
- Their experiences with encountering content they suspected was AI-generated.
- Demographic information and duration of community membership.
- Distribute the survey via community channels with moderator permission.

4.3. DATA ANALYSIS

4.3.1. QUALITATIVE ANALYSIS

Conducting a thematic analysis of the collected text data should be executed Braun and Clarke (2006). Code for themes related to:

1) Discussions about the authenticity, accuracy, or origin of content.
2) Expressions of trust or distrust in specific members or information.
3) Community norms and practices for verifying information.
4) Reactions to content suspected of being AI-generated (e.g., skepticism, humor, concern).
5) Changes in interaction patterns or the quality of discourse.

4.3.2. QUANTITATIVE ANALYSIS

Analysis of survey data using statistical methods is performed at this level:

1) Descriptive statistics to summarize member demographics and perceptions.
2) Correlation analysis to examine relationships between awareness of AI content, trust levels, and credibility evaluations.
3) Regression analysis to identify predictors of trust or perceived credibility in the presence of AI content.
4) Comparison between responses across different communities if multiple sites are studied.

4.4. ETHICAL CONSIDERATIONS

Several points must be considered before and during conducting the experiment to abide to the ethical commitments. This should take priorities for the following:

1) Ensuring compliance with platform terms of service and community rules regarding data collection.
2) Anonymization of all collected data to protect member identities.
3) Obtaining informed consent for survey participation.

4) Transparency with community moderators about the research objectives.
5) Avoiding introducing AI-generated content into the communities as part of the study.

This mixed-methods approach would enable both an in-depth understanding of the subtleties of how AI content is discussed and perceived within the specific context of each community (qualitative) and the capacity to identify broader patterns and relationships between variables across a larger group of members (quantitative).

5. HYPOTHETICAL FINDINGS AND ANALYSIS

Based on the proposed methodology and drawing upon existing theoretical understanding, hypothetical findings might reveal several key dynamics regarding the impact of AI-generated content on trust and credibility in specialized online communities.

5.1. SUBTLE EROSION OF PERCEIVED AUTHENTICITY

Qualitative analysis might indicate that while obvious "bot" behaviour is often quickly identified and managed by members, more sophisticated AI-generated content, particularly when posted under established human-seeming profiles, induces a subtle but persistent feeling of unease. This might be illustrated when members might describe content as feeling "off," "generic," or lacking the specific nuances and personal experiences typically shared within the community. This is not necessarily explicit detection of AI, but rather a failure to meet the implicit expectations of genuine human contribution.

This could be best exemplified by a hypothetical Quote: "It's hard to explain, but some posts just feel... flat? Like they read perfectly fine, but there's no real personality or lived experience behind them, which is usually what makes this group special."[1]

5.2. COMPLICATION OF TRADITIONAL TRUST CUES

Survey data might suggest that members traditionally rely on cues such as consistent participation, sharing personal stories, and demonstrating specific expertise relevant to the niche. However, sophisticated AI can imitate these cues. An AI posting under a long-standing profile could generate content that appears knowledgeable and consistent, confusing members' trust evaluations. In the hypothetical Finding, correlation analysis may show a weaker relationship between perceived expertise (based on post content) and overall member trust in communities where suspected AI content is more prevalent, compared to those where it is not. Similar effects were observed in general forums Starbird et al. (2019). This suggests that content quality alone becomes a less reliable indicator of a trustworthy human source.

[1] Quotes are illustrative, based on anecdotes from prior studies (e.g., Starbird et al. (2019)).

5.3. INCREASED SCRUTINY AND SKEPTICISM

Qualitative data might reveal an increase in members questioning the origin or authenticity of content, even when it is human generated. The possibility of AI content could lead to a general rise in skepticism, requiring members to invest more effort in verifying information or evaluating sources. This increased cognitive burden could make participation feel less effortless and enjoyable.

Hypothetical Finding: Thematic analysis identifies a recurring theme of "verification burden," where members express frustration about needing to cross-reference information or question contributions that they previously would have accepted without hesitation. This aligns with cognitive load theory Sweller (1988), suggesting platform tools (e.g., credibility badges) could mitigate effort.

5.4. IMPACT ON COMMUNITY COHESION

Hypothetically, if AI content becomes widespread or causes significant disputes over authenticity, it could damage the social fabric of the community. Trust is a key element holding the community together; its erosion can lead to decreased interaction, members withdrawing, or even fragmentation into smaller, more isolated groups where perceived authenticity can be more easily verified.

Hypothetical Finding: Survey data shows a negative correlation between reported exposure to suspected AI content and members' sense of belonging or commitment to the community.

5.5. MODERATION CHALLENGES

Qualitative data might highlight the difficulties faced by community moderators in identifying and addressing AI-generated content, especially when it does not explicitly violate rules (e.g., spam, hate speech) but undermines the community's reliance on authentic contributions. Developing clear policies and effective tools for managing AI content would emerge as a significant challenge.

Hypothetical Quote from Moderator: "It's a grey area. We can ban spam bots easily. But what about a member who uses ChatGPT to write their posts? It's not against the rules, but it feels... wrong. It changes the dynamic."[2]

These hypothetical findings suggest that the impact of AI-generated content is not merely about the presence of bots, but about the subtle ways sophisticated AI can imitate human interaction, complicating trust cues and potentially eroding the perceived authenticity that is crucial for the health of specialized online communities. The analysis would emphasize the nuanced nature of this impact, varying based on the specific community's norms, the sophistication of the AI content, and the members' digital literacy.

6. DISCUSSION

The hypothetical investigation underscores the significant, though potentially subtle, threat that the proliferation of AI-generated content poses to the integrity and functioning of specialized online communities. These spaces, built on foundations of shared passion, mutual support, and trusted information exchange,

[2] Quotes are illustrative, based on anecdotes from prior studies (e.g., Starbird et al., 2019).

are particularly vulnerable because their value proposition is so closely tied to the authenticity and credibility of human interaction.

The potential for AI to replicate human communication challenges the traditional mechanisms by which trust, and credibility are established and maintained in these communities. When content that appears knowledgeable or empathetic might lack genuine human experience or intent, members' ability to discern reliable sources is compromised. This complicates the informal reputation systems and social cues that communities have developed over time to navigate the digital landscape.

Moreover, the hypothetical finding of a "verification burden" is particularly concerning. If members must constantly question the origin and authenticity of content, the ease and spontaneity of interaction are diminished. This increased cognitive load can make participation less rewarding and potentially drive members away, leading to a decline in activity and the potential loss of valuable knowledge and social capital accumulated within the community.

Furthermore, the challenges for community moderators are substantial. Policing content that is technically not "harmful" but undermines the community's core values of authenticity and genuine contribution requires new approaches and tools. Developing clear guidelines on the acceptable use of AI assistance in content creation, promoting transparency, and empowering members with the skills to critically evaluate online information are becoming increasingly necessary. Moderators could deploy AI-detection APIs (e.g., Botometer) and community voting systems to flag synthetic content.

The implications extend beyond the individual communities themselves. Specialized online communities serve as important sources of specialized information, social support, and collective action. Their degradation due to eroded trust and credibility could have broader societal consequences, impacting everything from the spread of accurate health information to the organization of grassroots movements.

Nevertheless, it is important to acknowledge the limitations of this hypothetical exploration. A real study would need to account for the diversity of specialized communities, variations in platform design, the evolving capabilities of AI, and the dynamic nature of online interactions. However, this analysis provides a theoretical basis for understanding the potential challenges and highlights the urgent need for empirical research in this area.

7. CONCLUSION

The rise of sophisticated AI-generated content represents a critical turning point for specialized online communities. These vital digital spaces, characterized by deep engagement, shared interests, and a reliance on member trust and credible information, face significant challenges as the distinction between human and machine-authored content blurs. This theoretical article has explored the potential impact of AI-generated content on trust and credibility within these communities, proposing a conceptual framework and outlining a hypothetical research approach. Based on theoretical considerations, it was posited that AI content can subtly diminish perceived authenticity, complicate traditional trust cues, increase member skepticism, and potentially harm community cohesion.

Addressing these challenges requires a multi-directional approach. Platform developers should investigate mechanisms for promoting transparency regarding

content origin, potentially through technical watermarks or clear labelling where feasible. Platforms should mandate 'AI-generated' labels, as proposed in the EU AI Act (2024). Community moderators need support and resources to develop and enforce norms that prioritize authenticity and manage the integration of AI tools responsibly. Crucially, members themselves need to cultivate enhanced digital literacy skills, enabling them to critically evaluate online content and understand the capabilities and limitations of AI.

Future research should empirically investigate the dynamics outlined in this hypothesis across a range of specialized communities, employing diverse methodologies to capture the complexity of this evolving landscape. Longitudinal studies could track changes in trust levels and interaction patterns over time as AI content becomes more prevalent. Research could also explore the effectiveness of different moderation strategies and platform features in mitigating the negative impacts.

Ultimately, the future of specialized online communities in the age of AI will depend on the collective efforts of platforms, moderators, and members to maintain the integrity of their shared spaces, ensuring that they remain environments where genuine connection, credible information, and mutual trust can continue to flourish. The challenge is significant, but the preservation of these valuable digital ecosystems is essential for a healthy and informed online society.

DECLARATION

This is declaration of the Generative AI tools were employed as assistant for the author in arranging and writing this manuscript. "As AI tools become increasingly integrated into scholarly work, their responsible use—as a supplement under rigorous human oversight—must prioritize transparency, accountability, and the preservation of human creativity. Authors must explicitly disclose AI-assisted processes to maintain trust, ensuring these tools enhance rather than replace the irreplaceable 'human touch' in academic writing."

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Ardichvili, A., Page, V., & Wentling, T. (2003). Motivation and Knowledge Sharing in Online Communities of Practice. Journal of Knowledge Management, 7(3), 64–77. https://doi.org/10.1108/13673270310463626

Basta, Z. (2024). The Intersection of AI-Generated Content and Digital Capital: An Exploration of Factors Impacting AI-Detection and its Consequences. DIVA.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the Opportunities and Risks of Foundation Models. Arxiv Preprint Arxiv:2108.07258. https://doi.org/10.48550/arXiv.2108.07258

Braun, V., & Clarke, V. (2006). Using Thematic Analysis in Psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Broussard, M. (2020). Artificial Unintelligence: How Computers Misunderstand the World. MIT Press.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems, 33, 1877–1901. https://doi.org/10.48550/arXiv.2005.14165

Bryce, J., & Fraser, J. (2014). The Role of Disclosure of Personal Information in the Evaluation of Risk and Trust in Young Peoples' Online Interactions. Computers in Human Behavior, 30, 299–306. https://doi.org/10.1016/j.chb.2013.09.012

Burtch, G., Lee, D., & Chen, Z. (2023). The Consequences of Generative AI for UGC and Online Community Engagement. SSRN. https://doi.org/10.2139/ssrn.4521754

Cinus, F., Minici, M., Luceri, L., & Ferrara, E. (2025, April). Exposing Cross-Platform Coordinated Inauthentic Activity in the Run-Up to the 2024 US Election. In Proceedings of the ACM on Web Conference 2025 (pp. 541–559). https://doi.org/10.1145/3696410.3714698

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., ... & Wright, R. (2023). Opinion Paper: "So What if ChatGPT Wrote it?" Multidisciplinary Perspectives on Opportunities, Challenges and Implications of Generative Conversational AI for Research, Practice and Policy. International Journal of Information Management, 71, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

European Commission. (2021, February 10). The Artificial Intelligence Act.

Ferrara, E. (2023). Social Bot Detection in the Age of ChatGPT: Challenges and Opportunities. First Monday, 28(6). https://doi.org/10.5210/fm.v28i6.13185

Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of Internet Information Credibility. Journalism & Mass Communication Quarterly, 77(3), 515–540. https://doi.org/10.1177/107769900007700304

Flanagin, A., & Metzger, M. J. (2017). Digital Media and Perceptions of Source Credibility in Political Communication. In The Oxford Handbook of Political Communication (pp. 417–436). https://doi.org/10.1093/oxfordhb/9780199793471.013.65

Flavián, C., Guinalíu, M., & Gurrea, R. (2006). The Role Played by Perceived Usability, Satisfaction and Consumer Trust on Website Loyalty. Information & Management, 43(1), 1–14. https://doi.org/10.1016/j.im.2005.01.002

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. Minds and Machines, 30(4), 681–692. https://doi.org/10.1007/s11023-020-09548-1

Fogg, B. J. (2002). Persuasive Technology: Using Computers to Change What we Think and do. Ubiquity, 2002(December), 2. https://doi.org/10.1145/764008.763957

Gallagher, M., Pitropakis, N., Chrysoulas, C., Papadopoulos, P., Mylonas, A., & Katsikas, S. (2022). Investigating Machine Learning Attacks on Financial Time Series Models. Computers & Security, 123, 102933. https://doi.org/10.1016/j.cose.2022.102933

Kollock, P. (1999). The Production of Trust in Online Markets. Advances in Group Processes, 16(1), 99–123.

Kozyreva, A., Lewandowsky, S., & Hertwig, R. (2020). Citizens Versus the Internet: Confronting Digital Challenges with Cognitive Tools. Psychological Science

in the Public Interest, 21(3), 103–156. https://doi.org/10.1177/1529100620946707

Labajová, L. (2023). The State of AI: Exploring the Perceptions, Credibility, and Trustworthiness of the Users Towards AI-Generated Content [Doctoral Dissertation, University of Technology].

Lai, H., & Nissim, M. (2024). A Survey on Automatic Generation of Figurative Language: From Rule-Based Systems to Large Language Models. ACM Computing Surveys, 56(10), 1–34. https://doi.org/10.1145/3654795

Lankes, R. D. (2007). Trusting the Internet: New Approaches to Credibility Tools. Macarthur Foundation Digital Media and Learning Initiative.

Lave, J., & Wenger, E. (1991). Situated Learning: Legitimate Peripheral Participation. Cambridge University Press. https://doi.org/10.1017/CBO9780511815355

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... & Zittrain, J. L. (2018). The Science of Fake News. Science, 359(6380), 1094–1096. https://doi.org/10.1126/science.aao2998

Lin, H. Y., Yeh, Y. M., & Chen, W. C. (2017). Influence of Social Presence on Sense of Virtual Community. Journal of Knowledge Management, Economics and Information Technology, 7(2), 1–14. https://doi.org/10.1109/IJCSS.2011.29

Makki, A., & Jawad, O. (2023). Future Challenges in Receiving Media Messages in Light of Developments in Artificial Intelligence. Migration Letters: An International Journal of Migration Studies, 20, 167–183. https://doi.org/10.59670/ml.v20iS6.3943

Marinescu, V., Fox, B., Roventa-Frumusani, D., Branea, S., & Marinache, R. (2022). News Audience's Perceptions of and Attitudes Towards AI-Generated News. In Futures of Journalism: Technology-Stimulated Evolution in the Audience–News Media Relationship (pp. 295–311). Springer International Publishing. https://doi.org/10.1007/978-3-030-95073-6_19

Nonaka, I., & Takeuchi, H. (1996). The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation. Long Range Planning, 29(4), 592. https://doi.org/10.1016/0024-6301(96)81509-3

O'Keefe, D. J. (1982). Persuasion: Theory and Research. Communication Theory, 147, 191.

Oksymets, V. (2024). The Impact of Artificial Intelligence on Journalism Practices and Content Creation (Doctoral Dissertation, Vytautas Magnus University, Kaunas, Lithuania).

Pennycook, G., & Rand, D. G. (2021). The Psychology of Fake News. Trends in Cognitive Sciences, 25(5), 388–402. https://doi.org/10.1016/j.tics.2021.02.007

Pennycook, G., & Rand, D. G. (2022). Accuracy Prompts are a Replicable and Generalizable Approach for Reducing the Spread of Misinformation. Nature Communications, 13(1), 2333. https://doi.org/10.1038/s41467-022-30073-5

Pennycook, G., Binnendyk, J., Newton, C., & Rand, D. G. (2021). A Practical Guide to Doing Behavioral Research on Fake News and Misinformation. Collabra: Psychology, 7(1). https://doi.org/10.1525/collabra.25293

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting Attention to Accuracy can Reduce Misinformation Online. Nature, 592(7855), 590–595. https://doi.org/10.1038/s41586-021-03344-2

Preece, J. (2000). Online Communities: Designing Usability, Supporting Sociability. John Wiley & Sons. https://doi.org/10.1108/imds.2000.100.9.459.3

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners. OpenAI Blog, 1(8), 9.

Ren, Y., Kraut, R., & Kiesler, S. (2007). Applying Common Identity and Common Bond Theories to Design of Online Communities. Organization Studies, 28(3), 377–408. https://doi.org/10.1177/0170840607076007

Rheingold, H. (1993). The Virtual Community: Homesteading on the Electronic Frontier. MIT Press.

Ridings, C. M., Gefen, D., & Arinze, B. (2002). Some Antecedents and Effects of Trust in Virtual Communities. Journal of Strategic Information Systems, 11(3–4), 271–295. https://doi.org/10.1016/S0963-8687(02)00021-5

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so Different After All: A Cross-Discipline View of Trust. Academy of Management Review, 23(3), 393–404. https://doi.org/10.5465/amr.1998.926617

Shin, J., Jian, L., Driscoll, K., & Bar, F. (2018). The Diffusion of Misinformation on Social Media: Temporal Pattern, Message, and Source. Computers in Human Behavior, 83, 278–287. https://doi.org/10.1016/j.chb.2018.02.008

Starbird, K., Arif, A., & Wilson, T. (2019). Disinformation as Collaborative Work: Surfacing the Participatory Nature of Strategic Information Operations. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), 1–26. https://doi.org/10.1145/3359229

Sundar, S. S. (2008). The MAIN Model: A Heuristic Approach to Understanding Technology Effects on Credibility. In Digital Media, Youth, and Credibility (pp. 72–100). MIT Press.

Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. Cognitive Science, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Toral, S. L., Martínez-Torres, M. R., Barrero, F., & Cortés, F. (2009). An Empirical Study of the Driving Forces Behind Online Communities. Internet Research, 19(4), 378–392. https://doi.org/10.1108/10662240910981353

Vaccari, C., & Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. Social Media + Society, 6(1), 2056305120903408. https://doi.org/10.1177/2056305120903408

Weber, E., Rutinowski, J., Jost, N., & Pauly, M. (2024). Is GPT-4 Less Politically Biased Than GPT-3.5? A Renewed Investigation of ChatGPT's Political Biases. Arxiv Preprint ArXiv:2410.21008. https://doi.org/10.48550/arXiv.2410.21008

Wellman, B., & Gulia, M. (1999). Virtual Communities as Communities: Net Surfers don't Ride Alone. In M. A. Smith & P. Kollock (Eds.), Communities in Cyberspace (pp. 167–194).