INTERNATIONAL REVIEW OF ARTIFICIAL INTELLIGENCE IN MARKETING

NAWALA

Volume 4, Number 2, 2025

Artificially Generated Reviews in Digital Marketing: Impacts on Trust Perception and Decision-Making

Farhan Alfredo[1*]

[1] Sekolah Tinggi Ilmu Ekonomi Widya Wiwaha, Yogyakarta, Indonesia

Abstract

Article history:
Received: 29 September 2025
Revised: 11 October 2025
Accepted: 04 November 2025
Published: 30 December 2025

Keywords:
AI Reviews, Consumer Decisions, Digital Trust, Online Reviews, Trust Perception.

Identifier:
Nawala
Page: 108-121
https://nawala.io/index.php/iraim

This article examines how artificially generated reviews influence trust perception and consumer decision-making in contemporary digital marketing environments. It situates AI generated and AI manipulated reviews within the broader challenges of fake online reviews and asks under what conditions AI involvement in review creation undermines or reshapes their informational value. Using a systematic literature review of peer reviewed studies published between 2020 and 2024, the study synthesizes experimental, survey, and analytical evidence from e commerce and platform-based settings. The findings show that disclosure or suspicion of AI involvement typically reduces perceived authenticity, usefulness, and trust, with downstream effects on purchase intention, willingness to follow recommendations, and confidence in platforms and brands. The review also highlights that most detection-oriented research remains model centric and only weakly connected to user facing governance and design. By integrating these strands, the article identifies key conceptual and methodological gaps and proposes a research agenda for managing artificially generated reviews in a way that safeguards trust and decision quality.

*Corresponding author:
(Farhan Alfredo)

1. Introduction

Online reviews have become a central information source in digital marketing, shaping how consumers evaluate product quality, service reliability, and brand credibility. Trust in these reviews is therefore critical for their persuasive impact and for the effectiveness of firms' review-based strategies across search, platforms, and social commerce. Integrative work shows that trust in online reviews depends on a complex interplay of reviewer identity, platform design, and message features, and that even subtle credibility cues can alter perceived usefulness and diagnosticity (Borchers, 2023). At the same time, concerns about authenticity have intensified as deceptive or manipulated content has proliferated, threatening both consumer welfare and the legitimacy of review-driven marketing practices (Wu et al., 2020). This tension is amplified by the growing role of AI systems in both generating and moderating review content.

The emergence of generative artificial intelligence transforms the fake review problem from sporadic manipulation into the possibility of automated, large-scale production of highly fluent, contextually tailored text. Earlier research on deceptive reviews mapped their prevalence, typologies, and market impact, and catalogued algorithmic detection techniques that often frame the issue as a machine-learning classification task (Wu et al., 2020; Salminen et al., 2022). Recent studies show that large language models can create or paraphrase product reviews so effectively that even experienced consumers find it difficult to distinguish genuine from AI-generated content, with measurable consequences for perceived authenticity and trust in review platforms (Xylogiannopoulos et al., 2024). In parallel, advertising and

communication scholars argue that synthetic AI content such as deepfakes and AI-generated ads requires updated theoretical frameworks to understand how manipulated media shape persuasion, perceived falsity, and the need for consumer protection (Campbell et al., 2021). Together, this work suggests that artificially generated reviews are not simply another form of opinion spam but part of a broader shift toward AI-mediated influence in digital marketplaces.

Within this evolving landscape, studies have begun to explore consumer responses when they suspect or are explicitly told that content is AI generated. Experiments on AI-generated charitable giving ads show that disclosure of AI involvement can dampen prosocial responses by reducing perceived authenticity and emotional engagement, even when message quality is held constant (Arango et al., 2023). In the context of travel decision-making, ethical and quality concerns surrounding ChatGPT recommendations undermine acceptance and satisfaction primarily via reduced perceived trustworthiness and heightened risk perceptions (Kim et al., 2023). In online review settings specifically, consumers rate reviews they believe were written with ChatGPT as less useful, trustworthy, and authentic than identical reviews attributed to human authors, and this distrust can spill over to the platform and brand (Amos & Zhang, 2024). These findings indicate that artificially generated reviews occupy a paradoxical position: linguistically polished and scalable, yet prone to triggering authenticity concerns that reshape trust perception and decision-making.

Despite rapid developments, the literature on artificially generated reviews remains fragmented across review analytics, AI ethics, human-AI interaction, and

advertising research. There is no consolidated overview that focuses explicitly on AI-generated or AI-manipulated reviews and their implications for trust, credibility judgments, and consumer choice in digital marketing environments. This article addresses that gap by conducting a systematic literature review of peer-reviewed studies published between 2020 and 2024 that examine AI-generated, AI-assisted, or AI-paraphrased online reviews. The review maps how artificially generated reviews are conceptualized and operationalized, synthesizes empirical evidence on their effects on trust perception, diagnostic use of reviews, and downstream decisions such as purchase intention and platform choice, and identifies methodological and conceptual limitations that constrain cumulative knowledge. By doing so, the study aims to clarify the conditions under which artificially generated reviews undermine or potentially enhance promotional effectiveness, and to outline a research agenda and practical implications for managing AI-mediated reviews in contemporary digital marketing.

2. Literature Review

Research on online reviews in digital marketing has long emphasized their central role in shaping perceived credibility, diagnosticity, and purchase decisions. Foundational work on fake reviews shows that deceptive content can be highly persuasive and difficult to detect, even for trained observers, and that human and algorithmic detectors each have systematic blind spots (Wu et al., 2020; Plotkina et al., 2020). More recent integrative reviews argue that trust in online reviews depends on reviewer cues, platform design, and message characteristics, and call for more

nuanced models of how authenticity perceptions emerge in real consumption settings (Borchers, 2023). However, this earlier stream typically assumes that fake reviews are written manually, whereas current concerns increasingly center on scalable, AI mediated manipulation.

A second, rapidly growing stream focuses on detecting fake or low credibility reviews using machine learning and deep learning. Studies have examined linguistic, behavioral, and network-based features, showing that hybrid approaches combining text and reviewer behavior often outperform purely content-based models (Salminen et al., 2022; He et al., 2022). More recent work leverages transformer-based language models and advanced feature fusion to improve classification accuracy and robustness, demonstrating the value of contextual embeddings and deep architectures for review classification (Gupta et al., 2021; Mohawesh et al., 2024; Sun et al., 2024). A comprehensive review synthesizes these advances and concludes that the field has become heavily model centric, with limited attention to how detection tools are integrated into platform governance and how they influence user trust and market dynamics (Gupta et al., 2024). Complementary work on reputation systems highlights that rating fraud and fake reviews should be viewed as a strategic challenge for reputation mechanisms and platform efficiency, rather than a purely technical classification problem (Ananthakrishnan et al., 2020; Wang & Chen, 2020).

Only a smaller subset of studies explicitly addresses artificially generated reviews and AI mediated persuasive content. Experimental work on generative models shows that synthetic or paraphrased reviews can be linguistically fluent yet

distort consumer judgments, and that ChatGPT based paraphrases can undermine trust in otherwise genuine review ecosystems (Xylogiannopoulos et al., 2024). In parallel, research on AI generated ads and AI assisted recommendations finds that disclosure of AI involvement often reduces perceived authenticity, emotional engagement, and willingness to act on the message, even when message quality is held constant (Campbell et al., 2021; Arango et al., 2023; Kim et al., 2023). In direct review contexts, consumers evaluate reviews they believe were written with ChatGPT as less useful, trustworthy, and authentic than identical content attributed to human authors, with distrust spilling over to platforms and brands (Amos & Zhang, 2024). Taken together, these strands suggest that artificially generated reviews sit at the intersection of two partly disconnected literatures: one on fake review detection and one on AI mediated persuasion and trust, which this article aims to integrate.

3. Methods

The study employs a systematic literature review design to synthesize current evidence on artificially generated reviews and their effects on trust perception and decision-making in digital marketing contexts. A structured search was conducted in major academic databases such as Scopus, Web of Science, ScienceDirect, and Google Scholar for peer-reviewed articles published between 2020 and 2024. Search strings combined terms related to online reviews and deception (for example, "online reviews", "fake reviews", "review manipulation", "opinion spam") with terms related to artificial intelligence and generation (for example, "AI generated",

"ChatGPT", "large language model", "generative AI"). The search was restricted to English-language articles. After removing duplicates, titles and abstracts were screened to identify studies that explicitly addressed AI-generated, AI-assisted, or AI-paraphrased reviews in a digital marketing or consumer decision context.

Full texts of the remaining articles were then assessed using predefined inclusion and exclusion criteria. Studies were included if they examined artificially generated or AI-manipulated reviews, reported empirical findings (experimental, survey, field, or simulation) or analytical evaluations related to trust, credibility, usefulness, or decision outcomes, and focused on consumer-facing digital environments such as e-commerce platforms, review sites, apps, or social media. Conceptual papers without empirical or analytical results, technical detection papers that did not connect AI generation to consumer perception, and studies on non-review content (for example, purely advertising or chatbot interactions) were excluded unless they directly informed mechanisms relevant to review-based decision-making. For each included study, a standardized extraction template captured information on research design, platform or context, type of AI involvement, operationalization of artificially generated reviews, measures of trust and decision-making, key findings, and noted limitations. This structured approach enabled cross-study comparison and synthesis around the core themes of trust perception and decision outcomes.

4. Results and Discussion

Across the studies included in this review, artificially generated reviews are conceptualized in three main ways: fully AI written reviews, human written reviews edited or paraphrased with generative tools, and synthetic reviews used as stimuli to model future AI driven manipulation. Most empirical work uses experimental or scenario-based designs in e-commerce or platform style contexts, often with product review screenshots or mock interfaces that manipulate whether content is described as human written or AI generated (Arango et al., 2023; Kim et al., 2023; Amos & Zhang, 2024; Xylogiannopoulos et al., 2024). A smaller number of studies draw on real platform data or combine behavioral traces with survey measures, but controlled experiments remain the dominant method. Compared with the broader fake review literature, which emphasizes prevalence and detection (Plotkina et al., 2020; Wu et al., 2020; Salminen et al., 2022), this subset places trust perception and decision outcomes at the center of analysis.

A consistent finding is that perceived AI involvement tends to depress trust and perceived authenticity, even when the textual content is held constant. In review contexts, consumers who are told or led to believe that a review was written with ChatGPT rate it as less trustworthy, less useful, and less authentic than an identical review attributed to a human reviewer, and this effect can spill over to trust in the platform and brand (Amos & Zhang, 2024; Xylogiannopoulos et al., 2024). Similar patterns appear in adjacent domains. AI generated charitable giving ads are evaluated as less authentic and emotionally engaging, with disclosure of AI involvement further dampening prosocial responses (Arango et al., 2023). In travel decision

scenarios, ethical and quality concerns about ChatGPT recommendations translate into lower perceived trustworthiness and higher risk, which in turn reduce acceptance and satisfaction with AI suggestions (Kim et al., 2023). These results align with broader review work showing that trust in online reviews is highly sensitive to authenticity cues and perceived manipulation (Borchers, 2023).

The consequences of these trust shifts for decision-making are substantial. Studies find that lower trust in AI generated or AI assisted reviews is associated with reduced purchase intention, lower likelihood of following a recommendation, and weaker positive attitudes toward the reviewed product or service (Arango et al., 2023; Kim et al., 2023; Amos & Zhang, 2024). In some experiments, AI paraphrasing of genuine reviews not only reduces trust in the specific review but also undermines confidence in the surrounding review set, making consumers more uncertain and less willing to rely on reviews as a decision aid (Xylogiannopoulos et al., 2024). This suggests a possible negative externality: even if only a subset of reviews is artificially generated, the mere perception that AI could be involved may erode the informational value of the entire review corpus. The evidence also indicates that platform level and brand level trust are affected, raising strategic concerns for firms that depend heavily on review driven discovery and conversion.

At the same time, the review highlights that artificially generated reviews operate within a larger ecosystem of fraud and detection practices. Work on reputation fraud and fake review campaigns shows that fraudulent activity can strategically exploit platform mechanisms, and that decisions about whether to display suspicious or flagged reviews involve trade-offs between completeness,

transparency, and perceived credibility (Ananthakrishnan et al., 2020; Wang & Chen, 2020). Detection oriented research documents significant progress in machine learning and deep learning methods, including hybrid models that combine textual, behavioral, and network features, as well as transformer-based architectures that capture rich contextual signals (Gupta et al., 2021, 2024; He et al., 2022; Salminen et al., 2022; Mohawesh et al., 2024; Sun et al., 2024). However, most of these studies focus on classification accuracy and benchmark performance, with relatively little attention to how detection systems are integrated into platform governance or how users interpret and respond to detected or flagged content. From the perspective of artificially generated reviews, this means that even sophisticated detection may not automatically repair trust if users remain uncertain about the prevalence and handling of AI manipulation.

Overall, the findings indicate that artificially generated reviews sit at the intersection of two partially disconnected literatures: fake review detection and AI mediated persuasion. The empirical evidence suggests that, under current conditions, AI involvement in reviews is more likely to undermine than enhance trust and decision quality, especially when disclosed explicitly or suspected by users. Yet several limitations qualify these conclusions. Most studies rely on short term, self-reported outcomes in controlled scenarios with limited ecological complexity. Operationalizations of artificially generated reviews vary widely, making it difficult to compare effect sizes across contexts. There is little work on long term behavioral outcomes, cross platform dynamics, or culturally diverse consumer groups, and only a few studies explore potential design solutions such as nuanced disclosure formats,

interactional cues, or hybrid human-AI review systems. Addressing these gaps will be essential for understanding whether and how artificially generated reviews can be managed in ways that preserve informational value while mitigating risks to trust and decision-making.

5. Conclusion

This review shows that artificially generated reviews have a clear and mostly negative impact on trust perception and decision-making in digital marketing contexts. Across experimental and scenario-based studies, disclosure or suspicion of AI involvement tends to lower perceived authenticity, usefulness, and trust, with effects that extend beyond individual reviews to platforms and brands. At the same time, the broader ecosystem of fake review detection and reputation management reveals that technical advances in identifying deceptive content are not automatically translated into restored consumer confidence, particularly when platform governance and user facing communication remain opaque.

However, these conclusions are constrained by several shortcomings in the underlying studies. Most evidence is drawn from short term, self-reported outcomes in controlled settings, often using student or convenience samples, with limited attention to real behavioral data, cross platform dynamics, or cultural variation. Operationalizations of "artificially generated review" vary widely, and many studies focus on single products or narrow categories, which may limit generalizability. These methodological and contextual limits reduce the strength of causal claims and

leave open questions about how robust the observed effects would be in more complex, long run market environments.

Future research should therefore move beyond isolated experiments and integrate multiple levels of analysis. Longitudinal and field-based designs could examine how exposure to AI mediated reviews shapes repeat behavior, loyalty, and platform choice over time. Comparative studies are needed to test different disclosure formats, interface cues, and hybrid human–AI review arrangements, assessing not only immediate trust but also perceived fairness and control. Finally, stronger links between detection research and consumer facing design would help clarify how technical tools, governance policies, and user understanding jointly determine whether artificially generated reviews irreparably erode or can be responsibly integrated into digital marketing practice.

References

Amos, C., & Zhang, L. (2024). Consumer reactions to perceived undisclosed ChatGPT usage in an online review context. *Telematics and Informatics, 93*, 102163.

Ananthakrishnan, U. M., Li, B., & Smith, M. D. (2020). A tangled web: Should online review portals display fraudulent reviews?. *Information Systems Research, 31*(3), 950-971.

Arango, L., Singaraju, S. P., & Niininen, O. (2023). Consumer responses to AI generated charitable giving ads. *Journal of Advertising, 52*(4), 486–503.

Borchers, N. S. (2023). Why do we trust in online reviews? Integrative literature review and future research directions. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 17*(2), Article 7.

Campbell, C. L., Plangger, K., Sands, S., & Kietzmann, J. (2021). Preparing for an era of deepfakes and AI-generated ads: A framework for understanding responses to manipulated advertising. *Journal of Advertising, 51*(1), 22-38.

Gupta, P., Gandhi, S., & Chakravarthi, B. R. (2021). Leveraging transfer learning techniques BERT, RoBERTa, ALBERT and DistilBERT for fake review detection. In *Proceedings of the 13th Annual Meeting of the Forum for Information Retrieval Evaluation*.

Gupta, R., Jindal, V., & Kashyap, I. (2024). Recent state of the art of fake review detection: A comprehensive review. *The Knowledge Engineering Review, 39*, e8.

He, S., Hollenbeck, B., Overgoor, G., Proserpio, D., & Tosyali, A. (2022). Detecting fake review buyers using network structure: Direct evidence from Amazon. *Proceedings of the National Academy of Sciences, 119*(47), e2211932119.

Kim, J. H., Kim, J., Kim, C., & Kim, S. (2023). Do you trust ChatGPTs? Effects of the ethical and quality issues of generative AI on travel decisions. *Journal of Travel & Tourism Marketing, 40*(9), 779–801.

Mohawesh, R., Bany Salameh, H., Jararweh, Y., Alkhalaileh, M., & Maqsood, S. (2024). Fake review detection using transformer based enhanced LSTM and RoBERTa. *International Journal of Cognitive Computing in Engineering, 5*, 250-258.

Plotkina, D., Munzel, A., & Pallud, J. (2020). Illusions of truth: Experimental insights into human and algorithmic detections of fake online reviews. *Journal of Business Research, 109*, 511-523.

Salminen, J., Kandpal, C., Kamel, A. M., Jung, S., & Jansen, B. J. (2022). Creating and detecting fake reviews of online products. *Journal of Retailing and Consumer Services, 64*, 102771.

Sun, P., Bi, W., Zhang, Y., Wang, Q., Kou, F., Lu, T., & Chen, J. (2024). Fake review detection model based on comment content and review behavior. *Electronics, 13*(21), 4322.

Wang, Z., & Chen, Q. (2020). Monitoring online reviews for reputation fraud campaigns. *Knowledge-Based Systems, 195*, 105685.

Wu, Y., Ngai, E. W. T., Wu, P., & Wu, C. (2020). Fake online reviews: Literature review, synthesis, and directions for future research. *Decision Support Systems, 132*, 113280.

Xylogiannopoulos, K. F., Xanthopoulos, P., Karampelas, P., & Bakamitsos, G. A. (2024). ChatGPT paraphrased product reviews can confuse consumers and undermine their trust in genuine reviews. *Information Processing & Management, 61*(6), 103842.