Journal of Management Studies and Development
Vol.5, No. 03, pp. 511-527
journal.iistr.org/index.php/JMSD
DOI: 10.56741/IISTR.jmsd.002651

The AI Transparency Paradox in Digital Marketing: A Systematic Integrative Review with Comparative Evidence from the Philippines and Thailand

[1]Maria Joy C. Iglesia*, [1]Emjay Cervas, [2]Basel Abdulwahab, [3]Suhairee Berngacha
Corresponding Author: * joyidian@cspc.edu.ph
[1] Camarines Sur Polytechnic Colleges, Nabua, Philippines
[2] Azizstan Commerce Technological College, Pattani, Thailand
[3] Yala Rajabhat University, Yala, Thailand

ARTICLE INFO

Article history
Received 11 September 2026
Revised 30 October 2026
Accepted 08 November 2026

ABSTRACT

Artificial intelligence (AI) is increasingly embedded in advertising, personalization, influencer marketing, recommender systems, and customer interaction. Yet greater transparency about AI involvement does not produce uniformly positive consumer responses. This systematic integrative review synthesizes evidence on the emerging AI transparency paradox and compares how the underlying mechanisms are represented in the Philippines and Thailand. A structured search conducted on September 8, 2026 covered accessible publisher and journal platforms, including Innovative Marketing/Business Perspectives, ScienceDirect, Wiley, Springer, SAGE, Emerald, EconJournals, ThaiJO/TCI, and related scholarly repositories. The primary evidence window was 2020–2026; foundational theoretical sources were retained for conceptual framing. Twenty-eight recent academic studies were included in the substantive synthesis, with country-specific evidence examined separately for the Philippines and Thailand. Four recurring mechanisms emerged. First, AI disclosure often activates an authorship or credibility penalty, particularly in emotional, service, and authenticity-sensitive contexts. Second, clearer role disclosure and verification cues can restore process credibility and trust. Third, human presence and perceived authenticity remain important trust stabilizers. Fourth, personalization creates value while simultaneously intensifying privacy and surveillance concerns. Philippine evidence is comparatively trust- and

influencer-centered, whereas Thai evidence more strongly emphasizes personalization, privacy, and governance. The review proposes an AI Transparency Paradox Framework in which transparency signaling and persuasion-knowledge activation operate simultaneously, with authenticity, human oversight, privacy assurance, AI attitudes, and market context as boundary conditions.

Keywords
AI Disclosure
Consumer Trust
Digital Marketing
Personalization
Transparency

This is an open-access article under the CC–BY-SA license.

Introduction

Artificial intelligence has moved from a back-end analytical capability to a visible and increasingly autonomous participant in marketing communication. Marketers now use generative AI to produce advertising copy and images, recommendation algorithms to personalize product exposure, conversational AI to manage customer interactions, and virtual or AI-generated influencers to represent brands. This transformation changes not only marketing efficiency but also the social meaning of persuasion. Consumers are increasingly asked to evaluate messages whose authorship, targeting logic, and degree of human control may be opaque. The resulting managerial problem is no longer simply whether AI improves marketing performance, but whether consumers trust marketing when they know—or suspect—that AI has shaped the interaction [1], [2].

Transparency appears to offer an intuitive solution. Organizational-transparency research links disclosure, clarity, and accuracy with stakeholder judgments of trustworthiness [3], [4]. In an AI setting, disclosure can signal that a firm is willing to reveal otherwise hidden information about the production of persuasive content. However, the rapidly expanding advertising literature shows that the effect of such disclosure is unstable. Wortel et al. [5] found that AI labels reduced advertising attitudes, while Kirk and Givi [6] showed that AI authorship can reduce authenticity, word of mouth, and loyalty for emotional communications. Grigsby et al. [7] similarly reported lower trust and ad attitudes when service advertisements disclosed AI generation. Bui [8] found lower advertising value and purchase intention following AI-involvement disclosure. These findings suggest that transparency can reveal a source cue that consumers interpret negatively rather than rewarding disclosure as honesty.

At the same time, newer studies demonstrate that disclosure is not a unitary treatment. Le et al. [9] show that greater clarity about the functional role of AI increases perceived process credibility and advertising evaluation. Sarran and Datt [10] find that verification signals,

particularly third-party certification/assurance cues, can increase trust in AI-generated advertising. Nguyen et al. [11], in an Innovative Marketing study of Vietnamese Gen Z consumers, report that AI disclosure increased trust and indirectly increased purchase intention, although the effect was moderated by collectivistic orientation. Schreiner et al. [12], however, show that explanatory disclosures do not reliably improve evaluations in realistic Instagram-style settings and that AI aversion remains a strong predictor of negative responses. The emerging literature therefore points to a paradox: transparency can simultaneously operate as a positive signal of openness and a negative cue that activates skepticism, authenticity concerns, or persuasion knowledge.

This contradiction is especially important in Southeast Asia, where digital commerce, social media, influencer marketing, and AI-enabled personalization are expanding rapidly. Innovative Marketing has already published work on AI disclosure in Vietnam, social-media advertising in Thailand, and consumer engagement in the Philippines, indicating increasing scholarly attention to digitally mediated persuasion in the region [11], [13], [14]. Yet the Philippines and Thailand have not been systematically compared through the specific lens of AI-mediated trust. Recent Philippine research highlights AI-driven customer interaction, virtual influencers, brand trust, and loyalty [15], [16], [17]. By comparison, Thai research gives greater attention to AI personalization, recommender systems, privacy, virtual influencers, and responsible marketing communication [18], [19], [20], [21].

This study therefore synthesizes the competing mechanisms through which AI-mediated marketing shapes consumer trust and compares how those mechanisms are represented in Philippine and Thai evidence. The review asks three questions. First, under what conditions does AI transparency strengthen consumer trust and when does it undermine it? Second, which psychological mechanisms explain the competing effects of AI disclosure, personalization, and algorithmic brand representation? Third, what does the available evidence suggest about the comparative digital-marketing contexts of the Philippines and Thailand? The review develops an AI Transparency Paradox Framework that integrates signaling theory, the Persuasion Knowledge Model, algorithm aversion/appreciation, authenticity, and privacy-related mechanisms.

Literature Review and Conceptual Foundation

A. AI-mediated marketing and the changing identity of the persuader

Traditional digital marketing already blurred the boundaries between mass and interpersonal communication, but generative AI intensifies this shift because the persuasive agent may no longer be fully human. AI can write messages, optimize offers, emulate social presence, select content, recommend products, and represent a brand through a synthetic persona. Puntoni et al. [2] argue that consumer experiences with AI include data capture,

classification, delegation, and social interaction, each of which can create value while also generating psychological costs. Davenport et al. [1] similarly anticipated that AI would transform targeting, personalization, service, and marketing decision-making.

The literature does not support a simple assumption that consumers systematically reject algorithms. Algorithm-aversion research shows that people may abandon algorithmic advice after observing errors [22], whereas algorithm-appreciation research finds that people can prefer algorithmic advice to human judgment in some decision contexts [23]. This tension suggests that AI response depends on the social meaning of the task. Marketing is particularly sensitive because persuasion requires consumers to infer motives, credibility, authenticity, and responsibility.

B. Transparency signaling versus persuasion-knowledge activation

Signaling theory provides the positive side of the transparency mechanism. When information asymmetry exists, observable signals can communicate otherwise uncertain qualities [24]. Applied to AI advertising, disclosure can signal openness, procedural accountability, and willingness to reveal production practices. Organizational-transparency research further clarifies that transparency is not merely the quantity of disclosed information. It involves perceived disclosure, clarity, and accuracy [3], [4]. Thus, a minimal label stating "AI-generated" and a detailed explanation of AI's specific role are theoretically different signals.

The negative pathway is explained by the Persuasion Knowledge Model. Friestad and Wright [25] argue that consumers develop knowledge about persuasion agents, their motives, and their tactics, and use this knowledge to cope with influence attempts. Disclosure may therefore reveal more than production information; it can change the meaning of the advertisement by drawing attention to the persuasive machinery behind it. Campbell [26] showed that consumers infer manipulative intent when they perceive advertising tactics as disproportionately serving the advertiser. AI disclosure can activate similar concerns about automation, cost saving, scalable persuasion, synthetic authenticity, and psychological targeting.

The core proposition of this review is that both mechanisms can operate at the same time. AI transparency may improve trust by reducing informational opacity while simultaneously increasing scrutiny of the marketer's technological persuasion strategy. The observed consumer response is therefore the net result of competing positive and negative inferences rather than a single disclosure effect.

Methodology

A. Review design

The study used a systematic integrative review design because the relevant evidence spans experiments, cross-sectional surveys, structural-equation models, conceptual reviews, and

country-specific marketing studies. The approach combines a structured search and screening procedure with theory-driven thematic synthesis. The primary evidence window was January 2020 to September 8, 2026, reflecting the period in which AI became increasingly visible in consumer-facing marketing. Foundational studies published before 2020 were retained only for theoretical framing and were not counted as part of the recent evidence corpus.

B. Search strategy

Searches were executed on September 8, 2026 across accessible publisher and journal platforms and scholarly repositories, including Business Perspectives/Innovative Marketing, ScienceDirect, Wiley Online Library, Springer Nature, SAGE Journals, Emerald, Taylor & Francis, EconJournals, ThaiJO/TCI, institutional repositories, and reference-linked scholarly pages. Search strings combined AI terms (artificial intelligence, generative AI, AI-generated, algorithmic marketing, AI influencer, AI personalization, recommender system, chatbot) with marketing mechanisms (advertising, social media marketing, personalization, influencer marketing) and outcomes (trust, transparency, disclosure, authenticity, manipulation, privacy, credibility, engagement, purchase intention). Country-specific searches appended Philippines/Filipino and Thailand/Thai terms. Backward and forward citation chaining were used to identify additional directly relevant studies.

Table 1. Search Components and Operational Keywords Used in the Literature Review

Search component	Operational terms
AI technology	artificial intelligence; generative AI; AI-generated; algorithmic; chatbot; recommender system; AI/virtual influencer
Marketing context	advertising; social media marketing; personalization; influencer marketing; digital branding; e-commerce
Consumer mechanism	trust; transparency; disclosure; credibility; authenticity; manipulation; privacy; AI aversion
Behavioral outcome	purchase intention; engagement; loyalty; ad attitude; brand attitude
Country layer	Philippines; Filipino; Thailand; Thai

C. Inclusion and exclusion criteria

Studies were included when they (a) examined AI-mediated marketing or advertising, (b) reported consumer trust, credibility, authenticity, privacy, or behavioral outcomes, (c) were peer-reviewed journal articles or substantive scholarly reviews, and (d) provided sufficient methodological or conceptual detail for synthesis. Country-specific evidence was included when the study was conducted in the Philippines or Thailand and directly addressed AI-enabled marketing, virtual influencers, personalization, recommender systems, privacy, or related digital consumer behavior. Purely technical AI papers, non-marketing AI applications, duplicate versions, and commentary without an identifiable scholarly basis were excluded. Official policy documents from the Philippine National Privacy Commission and Thailand's Electronic Transactions Development Agency were used only for contextual interpretation and were not counted as academic evidence.

D. Evidence corpus and analytic procedure

The final recent academic evidence set comprised 28 studies used in the substantive synthesis. These included global mechanism studies on AI disclosure, authorship, verification, personalization, authenticity, and trust; six studies providing Philippine digital-marketing evidence; six providing Thai evidence; and additional Innovative Marketing studies used to position the review within the journal's current research conversation. Because the web-accessible search interfaces used here do not expose stable database-level result counts comparable to a Scopus or Web of Science export, the review does not fabricate a PRISMA record-flow count. Instead, it reports an auditable included-study corpus and explicitly treats exhaustive database coverage as a limitation. Studies were coded for AI mechanism, context, method, positive trust mechanism, negative trust mechanism, behavioral outcome, and boundary condition. Thematic comparison generated four higher-order patterns.

Results

A. Characteristics of the emerging evidence base

The literature is recent and methodologically heterogeneous. Experimental studies dominate the direct AI-disclosure stream, while survey-based SEM studies are more common in country-specific Southeast Asian work. The global disclosure literature concentrates on source labels, AI authorship, process credibility, ad attitudes, authenticity, and purchase intention. Philippine studies focus heavily on brand trust, AI-generated or virtual influencers, adoption intention, and digital-brand interaction. Thai studies emphasize personalization, recommender systems, privacy, virtual influencers, and the governance of AI-enabled marketing communication.

Table 2. Summary of Empirical Studies on AI Mechanisms and Consumer Responses in Digital Marketing

Study/context	Design	Primary AI mechanism	Main implication
Wortel et al. [5]	3-condition experiment	AI disclosure in Instagram ads	Disclosure reduced ad attitude; effects depended partly on AI aversion and manipulation inference.
Kirk & Givi [6]	7 preregistered experiments	AI authorship	AI authorship reduced authenticity and downstream loyalty/WOM for emotional communication.
Grigsby et al. [7]	3 experiments	AI disclosure in service ads	Disclosure lowered trust and ad attitudes; selective human presence mitigated the penalty.
Bui [8]	Experiment, N=358	AI involvement disclosure	Disclosure reduced ad value and purchase intention via lower credibility.
Le et al. [9]	4 studies	Role disclosure transparency	Clearer explanation of AI's functional role improved process credibility and ad evaluation.

Study/context	Design	Primary AI mechanism	Main implication
Schreiner et al. [12]	3 studies	Explanatory/beneficiary framing	Detailed explanations often failed to overcome negative priors in realistic ad settings.
Nguyen et al. [11], Vietnam	Experiment + PLS-SEM, N=400	AI disclosure	Disclosure increased trust; trust mediated purchase intention; collectivism weakened effect.
Sarran & Datt [10]	Randomized within-subject experiment	Verification signals	Third-party assurance and QR cues increased trust; behavioral translation varied.
Becaro-Lapiz et al. [16], Philippines	PLS-SEM, N=473	AI-generated influencers	Authenticity and parasocial interaction influenced purchase intention; trust was a key mediator.
Pithuk et al. [18], Thailand	CB-SEM, N=450	AI personalization	AI personalization positively influenced trust, engagement, and value; effects varied by generation.
Jahed et al. [19], Thailand	Survey, N=403	AI recommender systems	Privacy, transparency, and fairness predicted trust; personalization and autonomy predicted satisfaction.
Songja & Cheyjunya [20], Thailand	Survey, N=420	Marketing technology and data disclosure	Privacy orientation shaped willingness to disclose data for improved customer experience.

B. Theme 1: AI disclosure frequently produces an authorship and credibility penalty

The first theme is a recurring penalty associated with revealed AI authorship. Wortel et al. [5] reported lower advertising attitudes after AI disclosure in Instagram advertising. Kirk and Givi [6] found a broader AI-authorship effect across seven experiments: when consumers believed emotional marketing messages were written by AI rather than a human, perceived authenticity declined, moral disgust increased, and positive word of mouth and loyalty weakened. The effect was attenuated when communications were factual or when AI merely edited rather than authored the content. This distinction is important because it suggests that the consumer response depends on the perceived replacement of human expressive effort, not simply on the presence of AI.

Service advertising shows a similar mechanism. Grigsby et al. [7] found that AI disclosure reduced trust in service providers and produced less favorable ad attitudes. The penalty was especially pronounced when AI generated intangible or human-centered service cues. Trust could be partially restored when the human service provider remained real while AI generated only tangible background elements. Bui [8] likewise found that disclosure reduced advertising value and purchase intention, with advertising credibility operating as a negative mediator. Kučinskas and Survilaite [27] reported that disclosure of AI involvement reduced perceived authenticity and, through that mechanism, lowered brand trust and purchase intention.

These studies indicate that disclosure can function as a negative source cue when the marketing task depends on sincerity, emotion, human effort, or relational authenticity. In such

contexts, the consumer does not simply learn that AI was used; the disclosure changes the inferred identity of the persuader.

C. Theme 2: Transparency quality and verification can convert disclosure into a trust signal

The second theme qualifies the disclosure penalty. Le et al. [9] demonstrate that consumers respond more favorably when the disclosure communicates the role AI played in the ad-creation process rather than merely labeling the final content as AI-generated. Greater role-disclosure transparency enhanced ad-creation process credibility, which in turn improved advertising attitudes. Regulatory-compliance signals and disclosure-motive cues further strengthened the effect. This evidence supports the view that clarity and informativeness matter more than disclosure presence alone.

Verification cues provide another route to trust. Sarran and Datt [10] distinguish institutional credibility signals, such as third-party assurance seals, from transparency signals, such as QR codes linking to provenance information. Both can improve trust, but their effects on purchase intention differ. This distinction is consistent with signaling theory: a signal's credibility depends on what it represents and how costly or verifiable it is.

The most important counterexample comes from Nguyen et al. [11], whose Innovative Marketing experiment in Vietnam found that AI disclosure increased trust and indirectly increased purchase intention. The direct disclosure-to-purchase path was not significant, implying that disclosure worked primarily through a trust mechanism rather than immediate behavioral persuasion. Collectivistic orientation weakened the disclosure-trust relationship, illustrating that transparency effects are contingent rather than universal.

However, explanatory disclosure is not a guaranteed remedy. Schreiner et al. [12] found that specific explanations performed better than minimal labels in a controlled text setting, but the benefits weakened or disappeared in Instagram-style advertising. AI aversion predicted negative responses across studies, and disclosure attention was itself a critical condition. This suggests that marketers cannot assume consumers will notice, process, or accept the intended meaning of a transparency statement.

D. Theme 3: Human presence, authenticity, and relational credibility remain trust stabilizers

The third theme concerns the continued importance of perceived humanness. Kirk and Givi [6] show that the negative authorship effect is strongest when communication is emotional and therefore normatively associated with human sincerity. Grigsby et al. [7] show that retaining a real human service provider can buffer negative responses to AI-generated service environments. Qadri and Moustafa [28] similarly report that authenticity mediates responses

to AI-driven personalized influencer content and that the effects of personalization depend on AI transparency and perceived influencer well-being.

Evidence on AI influencers reinforces this conclusion. Zhang and He [29] found that AI rather than human influencer endorsements lowered brand trustworthiness and purchase intention through reduced mind perception and trustworthiness. Yet country-specific studies indicate that consumers can accept synthetic influencers when authenticity and relational cues are sufficiently strong. In the Southern Philippines, Becaro-Lapiz et al. [16] found that perceived authenticity, parasocial interaction, and anthropomorphism all contributed to purchase intention, with trust serving as the more important mediator than attitude. Biason et al. [30] likewise found that virtual influencer marketing affected purchase intention among Filipino consumers through product involvement and brand familiarity.

The implication is not that human communication is always superior. Rather, human involvement operates as an accountability and authenticity cue when consumers expect emotion, responsibility, or relational intention. This helps explain why disclosures that specify human review or human responsibility may be more effective than labels that simply announce AI production.

E. Theme 4: Personalization generates value and a parallel privacy-surveillance cost

The fourth theme broadens the transparency paradox beyond advertising authorship. AI personalization can improve relevance, convenience, engagement, and satisfaction, but it also makes the underlying data infrastructure more salient. Srivastava and Gurme [31] report that AI-enabled social-media experiences involve simultaneous relationships among personalization, trust, loyalty, and privacy concern. This pattern is consistent with the classic personalization-privacy paradox: consumers value tailored experiences while remaining sensitive to profiling, data capture, and loss of control.

Thai evidence strongly reflects this duality. Jahed et al. [19], using 403 Thai online consumers, found that privacy, transparency, and fairness were positively associated with trust in AI-driven recommender systems, while personalization, autonomy, and diversity were associated with satisfaction. Pithuk et al. [18] found that AI-powered personalization positively influenced brand trust, engagement, and perceived value across Thai generations, with the strength of relationships varying by cohort. Songja and Cheyjunya [20] show that willingness to disclose personal data for improved customer experience differs across privacy-personality types, demonstrating that the value of personalization cannot be separated from consumer perceptions of data risk and control.

F. Comparative evidence: Philippines versus Thailand

The comparative pattern is not a direct country effect because the existing studies use different samples, constructs, and designs. Nevertheless, the evidence reveals distinct

emphases in the two national literatures. Philippine studies are comparatively concentrated on trust as a relational mechanism connecting AI-driven interaction, virtual influencers, brand reputation, adoption, and purchase or loyalty outcomes. Thai studies more frequently integrate personalization, recommender systems, data privacy, fairness, virtual-influencer exposure, and marketing-technology governance.

Table 3. Comparative Analysis of AI-Driven Marketing, Consumer Responses, and Governance in the Philippines and Thailand

Dimension	Philippines	Thailand	Comparative interpretation
Dominant AI applications	AI-driven brand interaction; AI/virtual influencers; personalization	AI personalization; recommender systems; virtual influencers; marketing technology	Both markets are highly platform-mediated, but Thai evidence is more system/data oriented.
Central psychological construct	Brand trust; authenticity; parasocial interaction; adoption intention	Trust; satisfaction; privacy; fairness; autonomy	Trust is common to both, but antecedent structures differ.
Typical behavioral outcome	Purchase intention; brand loyalty; engagement	Purchase intention; consumer behavior; willingness to disclose data	Philippine work emphasizes relationship conversion; Thai work adds privacy trade-offs.
Key tension	Authenticity of AI-mediated representation	Personalization value versus privacy/governance	The transparency paradox may manifest through different proximal concerns.
Governance context	Data Privacy Act and NPC AI guidance	PDPA plus ETDA AI-governance guidelines	Both contexts increasingly frame AI trust through transparency and data responsibility.

Philippine evidence includes Jesus et al. [15], who found that AI-driven customer interaction and related digital-branding strategies predicted trust and purchase intention among Metro Cebu consumers; Secretaria [17], who found that trust in AI-driven marketing predicted adoption intention and brand loyalty among Gen Z consumers; and Sario et al. [32], who examined AI-enhanced personalization in relation to self-congruity and engagement. Ramos [14], although not an AI study, demonstrates the central role of consumer engagement in converting social-media attachment into purchase intention among Filipino Facebook users, helping explain why relational trust mechanisms are particularly relevant to AI-mediated brand interaction.

Thai evidence includes Prueksuralai [21], who found generally positive consumer opinions about AI-influencer personality, credibility, and presentation, and Thein [13], whose Innovative Marketing study shows that entertainment and convenience shape attitudes toward social-media advertising and that those attitudes strongly predict purchase intention. Tancharoen [33] synthesizes Thai AI-marketing communication as a combination of improved

targeting, personalization, and predictive capability alongside risks of privacy invasion, misinformation, weakened human connection, and erosion of long-term trust. These findings support a more privacy- and governance-sensitive interpretation of the Thai context.

Discussion

A. The AI Transparency Paradox Framework

The synthesis supports a dual-path model. The first path is a transparency-signaling mechanism: meaningful disclosure reduces opacity, clarifies AI's functional role, signals organizational openness, and can increase process credibility and trust. The second is a persuasion-knowledge mechanism: disclosure increases awareness of automated persuasion, activates motive attribution, and can lower authenticity or increase skepticism and manipulative-intent inference. Consumer trust is the point at which these competing evaluations are integrated.

The framework therefore rejects the binary proposition that AI disclosure is either beneficial or harmful. The more theoretically useful question is which mechanism dominates under a particular configuration of message, product, technology, and consumer. The evidence reviewed here identifies at least five boundary conditions: (1) emotional versus factual communication; (2) full AI authorship versus AI assistance or editing; (3) minimal labeling versus role-specific explanation; (4) human presence, review, or accountability; and (5) consumer predispositions such as AI aversion, privacy orientation, and familiarity.

A sixth boundary condition is the market context. The Philippines and Thailand should not be reduced to static cultural stereotypes. The available evidence instead suggests different proximal trust problems. In Philippine studies, consumers often evaluate AI through relational constructs—authenticity, parasocial interaction, brand trust, and loyalty. In Thai studies, trust is more frequently connected to personalization quality, privacy, fairness, data disclosure, and governance. These differences may reflect the topics selected by researchers as much as actual population differences; nevertheless, they provide a useful agenda for comparative empirical testing.

Table 4. Conceptual Framework of Positive and Negative Consumer Responses to AI-Mediated Marketing Cues

AI-mediated marketing cue	Positive pathway	Negative pathway	Boundary conditions	Outcome
Disclosure/ personalization/ AI authorship/ AI influencer	Transparency → process credibility → trust	Persuasion knowledge → manipulation/authenticity/privacy concern → lower trust	Human oversight; AI aversion; privacy assurance; message type; product/service type; country context	Ad attitude; engagement; purchase intention; loyalty

B. Theoretical contributions

First, the review connects two literatures that are often treated separately. Transparency research explains why disclosure can build trust, while persuasion-knowledge research explains why the same disclosure can undermine persuasion. Integrating them resolves part of the apparent contradiction in recent AI-advertising findings.

Second, the review reframes AI disclosure from a categorical label into a communication-design variable. Le et al. [9], Sarran and Datt [10], and Schreiner et al. [12] collectively show that the content, specificity, credibility, and attention received by disclosure are as important as disclosure itself. Future research should therefore distinguish source disclosure, role disclosure, motive disclosure, verification disclosure, and human-oversight disclosure.

Third, the review places authenticity and human oversight inside the trust process rather than treating them as peripheral controls. When AI performs tasks associated with emotion, sincerity, creativity, or social presence, consumers may interpret automation as a reduction in human effort or accountability. Human involvement can therefore function as a substantive signal, not merely an aesthetic preference.

Fourth, the Philippines–Thailand comparison expands the regional conversation beyond a single Southeast Asian market. Innovative Marketing's Vietnam study established that AI disclosure can build trust in a collectivistic emerging market [11]. The present synthesis shows that regional generalization should remain cautious: country-specific evidence points to different combinations of authenticity, trust, personalization, privacy, and governance.

C. Managerial implications

The findings imply that marketers should avoid a one-size-fits-all AI label. A minimal "AI-generated" disclosure may satisfy a transparency requirement while simultaneously creating an authorship penalty. When disclosure is necessary, firms should explain the functional role of AI in concise, concrete terms and identify where human judgment or review remains responsible.

Second, verification may be more persuasive than explanation alone. Third-party assurance, provenance information, and clear data-use controls can provide stronger trust signals than lengthy justifications. This is particularly relevant in personalization contexts, where transparency about data sources, user control, and privacy safeguards may matter more than transparency about content generation.

Third, firms should align AI use with the communicative task. AI is less likely to damage trust when used for technical, factual, background, or efficiency-oriented elements than when it is perceived to replace human emotion, expertise, or moral responsibility. Hybrid human-AI production should therefore be communicated as augmentation rather than wholesale substitution when authenticity is strategically important.

Fourth, regional strategies should reflect the trust problem most salient to consumers. Philippine campaigns using AI influencers or automated brand interaction should prioritize authenticity, relational credibility, and visible human accountability. Thai personalization and recommendation strategies should place stronger emphasis on privacy, fairness, autonomy, and transparent data governance.

Conclusion

AI is changing marketing not only by automating tasks but by changing what consumers believe about the identity, motives, and accountability of the persuader. The literature reviewed here shows that transparency is neither uniformly beneficial nor uniformly harmful. AI disclosure can build trust when it clarifies process, reduces uncertainty, or provides credible verification; the same disclosure can reduce trust when it signals inauthentic authorship, reduced human effort, manipulative intent, or intrusive data practices. This duality constitutes the AI transparency paradox. The comparative evidence further suggests that Philippine scholarship currently emphasizes relational trust and authenticity, whereas Thai scholarship places greater weight on personalization, privacy, fairness, and governance. The proposed framework positions transparency signaling and persuasion-knowledge activation as competing pathways whose balance is shaped by human oversight, authenticity, privacy assurance, AI attitudes, message type, and market context. For marketing theory, this reframes AI disclosure as a contingent signaling problem. For practice, it suggests that responsible AI marketing should make AI use understandable without removing human accountability.

Limitations And Future Research

This review has four limitations. First, the evidence base is developing extremely rapidly, particularly in 2025–2026, so conclusions should be treated as provisional. Second, the search was systematic and auditable across accessible publisher and journal platforms but did not use a direct Scopus or Web of Science export. Stable database-level hit counts were therefore unavailable, and no fabricated PRISMA record-flow numbers are reported. A future submission could strengthen reproducibility by repeating the same search strings in Scopus and Web of Science and archiving the exported metadata.

Third, the Philippines–Thailand comparison is based on separate national studies rather than matched cross-country experiments. Differences in constructs may reflect research emphasis rather than true population differences. A next-step empirical study should expose matched Filipino and Thai samples to identical AI-disclosure stimuli and test measurement invariance before interpreting structural differences.

Fourth, most existing studies measure stated trust, attitudes, or purchase intentions rather than actual behavior. Future work should use click-through, choice, purchase, opt-out,

data-disclosure, and longitudinal loyalty measures. Particularly valuable designs would compare no disclosure, minimal AI disclosure, role-specific disclosure, and AI-plus-human-oversight disclosure while manipulating message type and privacy assurance.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] T. Davenport, A. Guha, D. Grewal, and T. Bressgott, "How artificial intelligence will change the future of marketing," Journal of the Academy of Marketing Science, vol. 48, pp. 24–42, 2020. doi: 10.1007/s11747-019-00696-0.

[2] S. Puntoni, R. W. Reczek, M. Giesler, and S. Botti, "Consumers and artificial intelligence: An experiential perspective," Journal of Marketing, vol. 85, no. 1, pp. 131–151, 2021. doi: 10.1177/0022242920953847.

[3] A. K. Schnackenberg, and E. C. Tomlinson, "Organizational transparency: A new perspective on managing trust in organization–stakeholder relationships," Journal of Management, vol. 42, no. 7, pp. 1784–1810, 2016. doi: 10.1177/0149206314525202.

[4] A. K. Schnackenberg, E. C. Tomlinson, and C. Coen, "The dimensional structure of transparency: A construct validation of transparency as disclosure, clarity, and accuracy in organizations," Human Relations, vol. 74, no. 10, pp. 1628–1660, 2021. doi: 10.1177/0018726720933317.

[5] C. Wortel, I. Vanwesenbeeck, and F. Tomas, "Made with artificial intelligence: The effect of artificial intelligence disclosures in Instagram advertisements on consumer attitudes," Emerging Media, vol. 2, no. 3, pp. 547–570, 2024. doi: 10.1177/27523543241292096.

[6] C. P. Kirk and J. Givi, "The AI-authorship effect: Understanding authenticity, moral disgust, and consumer responses to AI-generated marketing communications," Journal of Business Research, vol. 186, Art. no. 114984, 2025. doi: 10.1016/j.jbusres.2024.114984.

[7] J. L. Grigsby, M. Michelsen, and C. Zamudio, "Service ads in the era of generative AI: Disclosures, trust, and intangibility," Journal of Retailing and Consumer Services, vol. 84, Art. no. 104231, 2025. doi: 10.1016/j.jretconser.2025.104231.

[8] H. T. Bui, "Examining the effect of AI advertising involvement disclosure on advertising value and purchase intentions," Journal of Research in Interactive Marketing, vol. 20, no. 5, pp. 613–632, 2026. doi: 10.1108/JRIM-02-2025-0066.

[9] K. B. Q. Le, H. Khan, F. Li, and W. H. Kunz, "More than saying 'It's AI ': How role disclosure transparency in AI-generated ads influences persuasion," Psychology & Marketing, 2026. doi: 10.1002/mar.70175.

[10] D. Sarran and V. Datt, "Consumer trust and purchase intentions in AI-generated advertising: The role of verification signals," Computers in Human Behavior Reports, vol. 22, Art. no. 101112, 2026. doi: 10.1016/j.chbr.2026.101112.

[11] H. P. Nguyen, P. U. Le, and X. T. Le, "From AI disclosure to purchase intention: The mediating role of trust and the moderating effect of collectivistic orientation in Vietnam," Innovative Marketing, vol. 22, no. 1, pp. 274–284, 2026. doi: 10.21511/im.22(1).2026.20.

[12] T. Schreiner, Y. Bart, K. Pauwels, C. Schütze, and F. Buder, "Made with AI" but why? How consumers interpret beneficiary-framed AI disclosures in advertising," Journal of Marketing Analytics, 2026. doi: 10.1057/s41270-026-00534-7.

[13] W. M. Thein, "Understanding users' gratification factors influence on fashion purchase behavior: Mediating effect of attitude towards social media advertising," Innovative Marketing, vol. 21, no. 2, pp. 170–182, 2025. doi: 10.21511/im.21(2).2025.14.

[14] W. Ramos, "Consumer engagement on Facebook: Modelling its antecedents and consequences," Innovative Marketing, vol. 21, no. 4, pp. 104–118, 2025. doi: 10.21511/im.21(4).2025.08.

[15] J. B. Jesus, M. R. A. Besario, N. L. E. Bueno, C. L. Bonghanoy, R. I. Navaja, G. L. Basilisco, J. M. Sumilhig, and C. Ople-Alviola, "Brand reputation in the age of AI: Impact of AI-driven customer interaction, social media engagement, and sustainability messaging on consumer trust and purchase intention," International Review of Management and Marketing, vol. 16, no. 1, pp. 161–169, 2026. doi: 10.32479/irmm.21424.

[16] J. M. Becaro-Lapiz, M. N. R. Ronquillo, R. Dingding, A. Beliganio, G. Baynosa, and M. Doblas, "AI-generated influencers and purchase intention: The mediating role of trust and attitude in HEIs in

Southern Philippines," International Review of Management and Marketing, vol. 16, no. 4, pp. 787–796, 2026. doi: 10.32479/irmm.23102.

[17] N. M. Secretaria, "Trust in AI-driven marketing and its impact on brand loyalty: The mediating role of adoption intention among Generation Z consumers," International Review of Management and Marketing, vol. 16, no. 3, pp. 566–574, 2026. doi: 10.32479/irmm.22426.

[18] L. Pithuk, J. Chanwichian, K. Seskhumbong, and S. Kanchanopast, "Beyond engagement: How sustainable marketing, AI personalization, and influencer marketing build brand trust among Thai digital consumers—A multi-generational SEM analysis," Rajapark Journal, 20(67), 231–252, 2026.

[19] B. Jahed, T. Shinasharkey, and P. Suesaowaluk, "The impact of AI-driven recommender systems on consumer behavior in Thailand: The roles of trust and satisfaction," International Research E-Journal on Business and Economics, vol. 9, no. 1, pp. 68–80, 2026. doi: 10.14456/irejbe.2026.5.

[20] R. Songja and P. Cheyjunya, "Willingness to disclose personal data for enhanced customer experience through marketing technology and online data privacy personality," Communication and Media in Asia Pacific, vol. 8, no. 2, pp. 97–114, 2025. doi: 10.14456/cmap.2025.8.

[21] N. Prueksuralai, "Exposure and opinion of consumers in Thailand towards the presentation of brands via AI influencers," Silpakorn University e-Journal (Social Sciences, Humanities, and Arts), vol. 44, no. 6, pp. 48–61, 2024. doi: 10.14456/sujthai.2024.46.

[22] B. J. Dietvorst, J. P. Simmons, and C. Massey, "Algorithm aversion: People erroneously avoid algorithms after seeing them err," Journal of Experimental Psychology: General, vol. 144, no. 1, pp. 114–126, 2015. doi: 10.1037/xge0000033.

[23] J. M. Logg, J. A. Minson, and D. A. Moore, "Algorithm appreciation: People prefer algorithmic to human judgment," Organizational Behavior and Human Decision Processes, vol. 151, pp. 90–103, 2019. doi: 10.1016/j.obhdp.2018.12.005.

[24] M. Spence, "Job market signaling," The Quarterly Journal of Economics, vol. 87, no. 3, pp. 355–374, 1973. doi: 10.2307/1882010.

[25] M. Friestad and P. Wright, "The Persuasion Knowledge Model: How people cope with persuasion attempts," Journal of Consumer Research, vol. 21, no. 1, pp. 1–31, 1994. doi: 10.1086/209380.

[26] M. C. Campbell, "When attention-getting advertising tactics elicit consumer inferences of manipulative intent," Journal of Consumer Psychology, vol. 4, no. 3, pp. 225–254, 1995. doi: 10.1207/s15327663jcp0403_02.

[27] G. Kučinskas and E. Survilaite, "Revealing AI involvement in ad creation: Effects on authenticity, brand perceptions and consumer intentions," Journal of Information Systems Engineering and Management, 10(16s), 2025. doi: 10.52783/jisem.v10i16s.2659.

[28] U. A. Qadri and A. M. A. Moustafa, "They made it just for me! How AI transparency and influencer well-being shape consumer responses to AI-driven content," Psychology & Marketing, vol. 43, no. 3, pp. 591–608, 2026. doi: 10.1002/mar.70075.

[29] Y. Zhang and A. He, "The impact of AI influencer endorsements on consumers' purchase intentions: The serial mediating roles of mind perception and brand trustworthiness," Journal of Product & Brand Management, vol. 34, no. 5, pp. 754–765, 2025. doi: 10.1108/JPBM-02-2024-4974.

[30] R. Biason, A. K. Elnagar, C. Tolete, H. A. S. Elsaadany, S. Hasan, and L. Santos, "Virtual influencer marketing: Mediating roles of product involvement and brand familiarity," International Journal of Electronic Marketing and Retailing, vol. 17, no. 6, pp. 1–23, 2026. doi: 10.1504/IJEMR.2026.153272.

[31] R. K. Srivastava and V. Gurme, "Artificial intelligence and consumer behaviour on social media: A study of personalization, trust, and privacy," Measurement: Digitalization, vol. 5, Art. no. 100021, 2026. doi: 10.1016/j.meadig.2025.100021.

[32] Z. M. Z. Sario, J. J. Dulay, K. N. Aliarte, and N. J. Novilla, "Improving marketing personalization with AI: How self-congruency influences consumer engagement," International Journal of Research and Innovation in Social Science, vol. 8, no. 12, pp. 537–549, 2024. doi: 10.47772/IJRISS.2024.8120042.

[33] B. Tancharoen, "The impact of artificial intelligence adoption on marketing communication processes in Thailand," Journal of Integrated Sciences, 23(1), 36–53, 2026.

Authors



Dr. Maria Joy C. Iglesia has a doctoral degree in Business Administration and is currently designated as the Dean of the College of Tourism, Hospitality, and Business Management in Camarines Sur Polytechnic Colleges, Philippines. She has several published research studies and is working on research related to market assessment, gender and development awareness in partnered entrepreneurs, the readiness of tertiary students to business concepts, theories, and applications, and the business management of slaughterhouses in the Philippines. She teaches business courses with expertise in microeconomics, macroeconomics, and other related subjects. (email: joyidian@cspc.edu.ph).



Emjay R. Cervas, MBA, is a core faculty member in the Bachelor of Science in Business Administration program, major in Financial Management, under the College of Tourism, Hospitality, and Business Management at Camarines Sur Polytechnic Colleges. He received his master's in business administration from the University of Saint Anthony in 2024. His professional background spans financial data analysis and quality assurance, having served as a Content Analyst at the London Stock Exchange Group and as an Internal Quality Auditor at the University of Saint Anthony. (email: emjaycervas@cspc.edu.ph).



Dr. Basel Abdulwahab is the Director of Azizstan Commerce Technological College in Pattani Province, Thailand. With extensive experience in education and institutional leadership, he plays a significant role in advancing commerce and technological studies while fostering cross-cultural academic collaboration. He is actively involved in higher education management and community development. (email: bs.azizstan@gmail.com).



Dr. Suhairee Berngacha is a lecturer in the Department of Educational Administration at the Faculty of Education, Yala Rajabhat University, Yala, Thailand. He earned his doctoral degree from Thaksin University, Thailand. His academic interests include educational leadership, school management, and educational policy, with a focus on improving institutional effectiveness and governance. He is actively involved in teaching, research, and community engagement. (email: suhairee.b@yru.ac.th).