REVIEW ARTICLE

Consumer Trust in AI-Generated Marketing Content: A Systematic Literature Review and Research Agenda

Kirill Baryshkov[1], Yana Kuzina[2], Iryna Smuk[3], Mila Tkachuk[4]

[1] Global Media & Measurement Lead, L'Oréal HQ, Clichy, France

[2] Marketing Manager, LeadingAge PA, USA

[3] Head of Marketing, BRKTHROUGH, USA

[4] Digital Marketing Manager, Burke Interiors, Los Angeles, USA

* Corresponding author

Abstract

Purpose. As generative artificial intelligence transforms marketing content production, questions about consumer trust have become urgent for both scholars and practitioners. This study systematically reviews the empirical and conceptual literature on consumer trust-related responses to AI-generated marketing content (AIGC), mapping antecedents, mediating mechanisms, moderating conditions, and downstream outcomes.

Design/methodology/approach. Following PRISMA 2020 reporting principles, a systematic search of Scopus and Google Scholar identified 59 records. After deduplication, screening, and eligibility assessment, 35 studies published between 2020 and 2026 were retained. Each study was coded along eleven dimensions including methodology, theoretical framework, independent and dependent variables, mediators, and moderators. Thematic synthesis organized findings into four themes: the disclosure dilemma, authenticity as central mediator, moderating conditions, and downstream outcomes.

Findings. AI disclosure activates persuasion knowledge and erodes trust-related outcomes across diverse marketing contexts, yet these effects are neither universal nor uniform. Perceived authenticity emerges as the primary mediating mechanism, with moral disgust operating as a parallel affective pathway. Content type (emotional versus rational), consumer AI literacy, cultural context, anthropomorphism cues, and disclosure framing moderate the strength and direction of effects. An integrative input-process-output framework synthesizes these findings.

Originality/value. To the authors' knowledge, this review provides one of the first focused syntheses of empirical and conceptual work on the trust construct in AI-generated marketing content. It proposes an integrative framework organizing antecedents, mediators, moderators, and outcomes, and advances five research propositions addressing habituation dynamics, cross-cultural variation, content-modality contingencies, disclosure framing, and behavioral measurement gaps.

Keywords: AI-generated content, consumer trust, brand authenticity, AI disclosure, persuasion knowledge, systematic literature review, Marketing

1. Introduction

Generative artificial intelligence has reshaped how brands produce marketing content at scale. From product descriptions and social media posts to sponsored vlogs and advertising visuals, AI-generated marketing content (AIGC) now permeates virtually every consumer touchpoint. Yet as organizations embrace these tools for efficiency and cost reduction, accumulating empirical evidence suggests that consumers may penalize brands when AI authorship becomes apparent (Bruns & Meissner, 2024; Kirk & Givi, 2025). This tension between operational advantage and consumer backlash places trust at the center of an emerging scholarly debate.

The academic literature on consumer responses to AIGC has expanded at a remarkable pace. Over ninety percent of the empirical studies identified in this review were published between 2024 and 2026, reflecting a field still in its formative stage. Early conceptual work modeled automated brand-generated content within computational advertising frameworks (Van Noort et al., 2020), and bibliometric overviews have mapped the broader AI-advertising research domain across four thematic clusters (Ford et al., 2023). However, no systematic review to date has synthesized empirical findings specifically around the trust construct and the psychological mechanisms that govern consumer reactions to AI-generated marketing content.

Apparent contradictions within the existing evidence base heighten the need for such a synthesis. Several studies report that disclosing AI involvement in content creation activates persuasion knowledge and erodes credibility, trust, and purchase intentions (Bui, 2025; Qiu et al., 2025). Other findings challenge this pattern. Kirkby et al. (2023) observed no significant effect of AI disclosure on brand attitudes or perceived authenticity, suggesting that the relationship between transparency and trust may be more contingent than initially assumed. Reconciling these divergent results requires a structured integrative effort.

A deeper structural problem compounds the challenge of reconciliation. Research on consumer responses to AIGC is dispersed across several disciplinary streams that have developed largely in isolation: disclosure research rooted in the Persuasion Knowledge Model, authenticity theory originating in branding scholarship, anthropomorphism and uncanny valley studies drawn from human-computer interaction, virtual influencer research situated in social media studies, and source credibility work from communication science. Each stream carries its own constructs, measurement traditions, and theoretical assumptions. The central problem confronting the field is therefore not a shortage of individual studies but a lack of integration across these parallel lines of inquiry. Without a unifying synthesis, findings from one stream remain difficult to reconcile with those from another, and the apparent contradictions in the evidence base may reflect not genuine empirical disagreement but the consequence of fragmented theoretical framing.

The present review addresses this gap by posing three research questions. First, what antecedents of consumer trust responses to AI-generated marketing content appear most frequently in the empirical literature (RQ1)? Second, through what mediating mechanisms does AIGC influence trust-related outcomes (RQ2)? Third, what moderating conditions alter the strength or direction of these effects (RQ3)? In pursuing these questions, the review makes three contributions. It provides one of the first focused syntheses of empirical and conceptual work that maps antecedents, mediators, moderators, and outcomes across thirty-five studies. It proposes an integrative input-process-output framework that organizes disparate findings into a coherent model. It advances a research agenda comprising

five propositions that address unresolved tensions, including longitudinal habituation dynamics, cross-cultural variation, disclosure framing effects, and behavioral measurement gaps.

The remainder of this article proceeds as follows. Section 2 establishes the theoretical background by reviewing AI in marketing communications, brand trust and authenticity theory, and the persuasion knowledge model as it applies to AI disclosure. Section 3 describes the systematic review methodology. Section 4 presents the thematic results. Section 5 discusses theoretical and managerial implications, and Section 6 outlines the proposed research agenda. Section 7 acknowledges limitations and concludes.

2. Theoretical Background

2.1 AI in Marketing Communications

AI-generated marketing content refers to text, images, video, and interactive brand communications created partially or fully by generative AI systems. A meaningful distinction exists between AI-assisted content, where human creators use AI as a collaborative tool, and fully AI-generated content, where the system autonomously produces the output (Liu, Lian, & Osman, 2025a). The trajectory of AI in marketing has progressed from programmatic ad targeting and recommendation algorithms to genuine content creation, a shift that fundamentally alters the nature of brand communication (Davenport et al., 2020). Early conceptual work proposed models of automated brand-generated content situated within computational advertising, highlighting how machine-produced messages might reshape consumer-brand interactions (Van Noort et al., 2020). More recent frameworks have extended this thinking into the era of agentic AI, where autonomous systems generate, distribute, and optimize marketing content with minimal human oversight (Kim, 2025). A bibliometric overview of AI advertising research identified trust as one of four central thematic clusters, confirming its prominence in the field (Ford et al., 2023). Virtual influencers represent a further extension of AIGC, as entirely AI-driven personas engage audiences on social media platforms, raising distinct questions about parasocial trust and perceived humanness (Su, 2025). Importantly, the distinction between AI-assisted and fully AI-generated content may function not as a categorical binary but as a continuum along which consumer tolerance shifts, a possibility that the moderating conditions reviewed in Section 4.4 begin to address.

2.2 Brand Trust and Authenticity Theory

Brand trust denotes a consumer's willingness to rely on a brand's capacity to perform its stated function under conditions of uncertainty (Chaudhuri & Holbrook, 2001; Delgado-Ballester, 2004). Closely related is the construct of perceived authenticity, which encompasses judgments of genuineness, originality, and sincerity directed toward a brand and its communications (Morhart et al., 2015). Both constructs depend on an implicit assumption that brand messages originate from human intentionality, creative effort, and emotional investment. AI authorship may violate this assumption.

Experimental evidence supports the existence of an "authenticity gap" triggered by AI-generated content. Bruns and Meissner (2024) demonstrated across three experiments that generative AI content

creation diminishes perceived brand authenticity, which in turn mediates negative follower reactions. Kirk and Givi (2025) extended these findings through seven experiments, revealing a dual-pathway model in which AI authorship of emotional marketing content reduces consumer attitudes, word-of-mouth intentions, and brand loyalty through both reduced perceived authenticity and heightened moral disgust. The affective pathway through moral disgust represents a distinctive contribution, connecting AI content evaluation to moral psychology rather than purely cognitive appraisal. In domain-specific contexts, AI disclosure has been shown to undermine brand authenticity in restaurant marketing, subsequently damaging brand image (Ali et al., 2025). However, anthropomorphic design cues appear to partially restore authenticity perceptions for AI-generated fashion content, suggesting that the authenticity gap is not fixed (Lee & Kim, 2024). Zhang & Hur (2025) similarly found that AI-generated advertising images reduce brand trust through diminished authenticity perceptions, reinforcing the centrality of this construct. The relationship between anthropomorphism and consumer acceptance appears to follow a nonlinear pattern: moderate humanlikeness may enhance social presence and trust, whereas excessive humanlikeness risks triggering uncanny valley discomfort and perceptions of manipulative intent, a pattern with implications for both virtual influencer design and AI content strategy.

2.3 AI Disclosure and the Persuasion Knowledge Model

The Persuasion Knowledge Model (Friestad & Wright, 1994) provides a well-established theoretical lens for understanding consumer responses to AI disclosure. The model posits that when consumers recognize a persuasion attempt, they activate coping mechanisms such as skepticism, counterarguing, and resistance. Labeling content as "AI-generated" may function as a disclosure cue that triggers persuasion knowledge activation, thereby reducing source credibility and downstream trust.

Empirical evidence within the corpus largely confirms this mechanism. Bui (2025) found that AI disclosure in prosocial advertising negatively affects credibility and purchase intentions through persuasion knowledge activation. Wortel et al. (2024) documented similar persuasion knowledge and attribution processes triggered by AI disclosures on Instagram advertisements. Qiu et al. (2025) demonstrated that in cause-related marketing, AI disclosure inhibits purchase intentions via the same persuasion knowledge pathway. These findings align with a broader pattern: across the corpus, persuasion knowledge serves as the most frequently invoked theoretical framework, appearing in seven of the thirty-five coded studies.

Yet the relationship between disclosure and trust is not uniformly negative. Kirkby et al. (2023) found that AI disclosure did not significantly affect brand attitudes or perceived authenticity, a result that may reflect boundary conditions related to brand strength or consumer familiarity with AI tools. This "disclosure paradox" is further complicated by regulatory imperatives. The EU AI Act and evolving FTC guidance increasingly mandate transparency about AI-generated content, creating a tension between legal compliance and potential trust erosion (Grigsby et al., 2025). Survey data indicate that seventy-five percent of consumers favor AI disclosure in advertisements, yet the same disclosures reduce trust (Grigsby et al., 2025). Resolving this paradox requires attention to disclosure framing, content domain, and individual differences in AI literacy, themes that the present review examines in its results and discussion.

3. Methodology

3.1 Review Approach and PRISMA Protocol

A systematic literature review (SLR) was adopted as the research methodology, following established guidelines for evidence-informed management knowledge development (Tranfield et al., 2003; Snyder, 2019). The SLR approach was selected over narrative or bibliometric reviews because the research questions demanded a structured, transparent, and replicable synthesis of empirical findings on a clearly defined phenomenon: consumer trust responses to AI-generated marketing content. This review follows PRISMA 2020 reporting principles (Page et al., 2021), which provided the procedural framework for identification, screening, eligibility assessment, and inclusion of studies. The review protocol was developed prior to the search process and covered database selection, search string construction, inclusion and exclusion criteria, and the coding scheme applied to each retained article.

3.2 Search Strategy

A systematic search was conducted across two major academic databases: Scopus and Google Scholar. Scopus was selected for its comprehensive coverage of peer-reviewed journals in business, management, and marketing, while Google Scholar was employed to capture additional relevant publications that might not appear in Scopus-indexed outlets. The search strings combined three conceptual blocks using Boolean operators: (a) the technology domain ("AI-generated" OR "generative AI" OR "artificial intelligence"), (b) the application context ("marketing" OR "advertising" OR "brand content"), and (c) the focal construct ("trust" OR "authenticity" OR "credibility" OR "consumer response"). These terms were applied to titles, abstracts, and keywords. The search covered publications from January 2018 through March 2026, a date range chosen to capture the emergence of generative AI applications in marketing while extending to the most recent scholarship. Only English-language publications were included, as English remains the dominant language of publication in the marketing and advertising disciplines represented in the target journals.

3.3 Inclusion and Exclusion Criteria

Several criteria governed the selection of studies. To be included, a publication had to meet all of the following conditions: (a) it was a peer-reviewed journal article, (b) it employed an empirical, conceptual, or review methodology, (c) it examined consumer-level responses rather than firm-level or technical performance metrics, (d) it addressed trust, perceived authenticity, credibility, or a closely related construct as a dependent variable or key theoretical construct, and (e) it situated its analysis within a marketing, advertising, or brand communication context involving AI-generated content. Conference papers and book chapters were excluded to ensure a consistent standard of peer-review quality across all retained studies. Purely technical papers that evaluated AI system performance without examining consumer outcomes were also excluded. Studies that focused exclusively on AI-powered chatbots or conversational service agents without a content generation dimension were deemed outside the scope of this review, as the phenomenon of interest was the

creation of marketing content rather than interactive service delivery.

3.4 Coding Scheme

A structured coding scheme was developed to extract and categorize data from each included study. Every article was coded along the following dimensions: (a) publication year, (b) journal, (c) research method, classified as experimental, survey-based, qualitative, mixed-methods, or conceptual/review, (d) geographic context of the sample or analysis, (e) type of AI-generated content examined, categorized as text, image, video, or mixed, (f) independent variables and antecedents, (g) mediating mechanisms, (h) moderating conditions, (i) dependent variables and outcomes, (j) key findings, and (k) the primary theoretical lens employed. The most frequently applied theoretical frameworks across the corpus were the Persuasion Knowledge Model (seven studies), attribution theory (five studies), and signaling theory (five studies), with additional studies drawing on uncanny valley theory, trust transfer theory, and the stimulus-organism-response framework. Coding was performed by the first author and verified by the second author, with disagreements resolved through discussion until consensus was reached.

3.5 PRISMA Flow

The systematic search yielded a total of 59 records across the two databases, with 38 records identified through Scopus and 21 through Google Scholar. After removing nine duplicate records that appeared in both databases, 50 unique records remained for screening. Title and abstract screening was applied to all 50 records, resulting in the exclusion of 13 articles that were determined to be off-topic upon closer inspection, as they did not substantively address consumer trust-related responses to AI-generated marketing content. The remaining 37 articles proceeded to full-text assessment for eligibility. At this stage, two additional articles were excluded because they focused exclusively on AI-powered chatbot interactions without a content generation component, falling outside the defined scope of this review. The final sample comprised 35 studies that met all inclusion criteria and formed the analytical corpus for this systematic review. The distribution of these studies reflected the recency of the research domain: one study was published in 2020, two in 2023, nine in 2024, twenty in 2025, and three in 2026, confirming that the overwhelming majority of scholarship on consumer trust in AI-generated marketing content has emerged within the past three years. The complete PRISMA flow is depicted in Figure 1.

4. Results

4.1 Descriptive Overview

The systematic search and screening process yielded 35 studies for inclusion. Table 1 presents the full distribution of included studies by year, journal, methodology, and geographic context. The temporal distribution reveals a field in rapid expansion: only one study was published in 2020, two appeared in 2023, and nine in 2024, whereas 20 studies were published in 2025 and three in 2026.
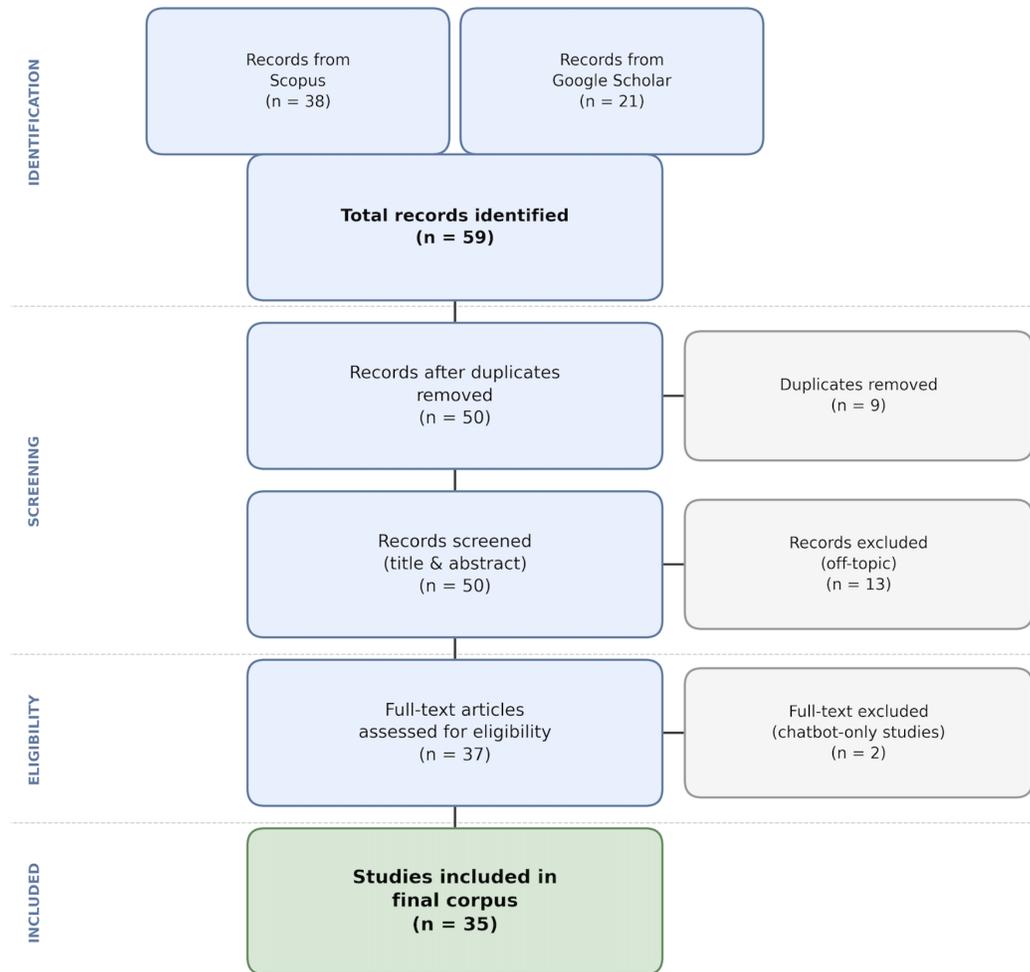
IDENTIFICATION

Records from
Scopus
(n = 38)

Records from
Google Scholar
(n = 21)

**Total records identified
(n = 59)**

SCREENING

Records after duplicates
removed
(n = 50)

Duplicates removed
(n = 9)

Records screened
(title & abstract)
(n = 50)

Records excluded
(off-topic)
(n = 13)

ELIGIBILITY

Full-text articles
assessed for eligibility
(n = 37)

Full-text excluded
(chatbot-only studies)
(n = 2)

INCLUDED

**Studies included in
final corpus
(n = 35)**

Figure 1. PRISMA 2020 Flow Diagram

This concentration means that over 91% of the empirical base has emerged within the most recent three years, reflecting both the novelty and the urgency of the topic. *Summary of Included Studies (N = 35)*

#	Authors (Year)	Journal	Method	Key Focus
1	Van Noort, Himelboim, Martin & Collinger (2020)	Journal of Advertising	Conceptual	Model of automated brand content and trust
2	Kirkby, Baumgarth & Henseler (2023)	Journal of Product & Brand Management	Experiment	AI disclosure impact on brand attitudes
3	Sands et al. (2023)	Journal of Advertising	Experiment	AI content barriers to charitable ad trust
4	Amos & Zhang (2024)	Telematics and Informatics	Experiment	Undisclosed ChatGPT reduces review trustworthiness
5	Brüns & Meißner (2024)	Journal of Retailing and Consumer Services	Experiment	GenAI content diminishes brand authenticity
6	Bui & Kozinets (2024)	International Journal of Advertising	Experiment	AI disclosure reduces credibility and purchase intent
7	Chen, Wang, Shi & Li (2024)	Journal of Business Research	Experiment	Ad appeal types and AI facial role
8	Ford, Jain, Wadhwani & Gupta (2024)	Journal of Business Research	Systematic review	Bibliographic review of AI advertising themes

Continued on next page

#	Authors (Year)	Journal	Method	Key Focus
9	Lee & Kim (2024)	Journal of Retailing and Consumer Services	Experiment	Anthropomorphism enhances AI fashion authenticity
10	Li, Zhang, Chang & Zhong (2024)	Journal of Retailing and Consumer Services	Experiment	AI celebrity disclosure and purchase decision
11	Mesut Cicek et al. (2024)	Journal of Hospitality Marketing & Management	Experiment	AI labeling lowers trust and purchase intent
12	Wortel, Vanwesenbeeck & Tomas (2024)	Emerging Media (SAGE)	Experiment	AI disclosure on Instagram ad attitudes
13	Wu, Dastio & Wen (2024)	Journal of Advertising	Experiment	AI disclosure effects are task-dependent
14	Ali, Ali, Abdalla & Abdalla (2025)	International Journal of Hospitality Management	Experiment	GenAI undermines restaurant brand authenticity
15	Baek et al. (2025)	Administrative Sciences	Experiment	AI images reduce authenticity and brand trust
16	Bui (2025)	Journal of Research in Interactive Marketing	Experiment	AI disclosure diminishes ad value, purchase intent
17	"Decoding the Trust Matrix" (2025)	Journal of Interactive Advertising	Survey	Predictors of trust in personalized AI ads
18	Du Plessis (2025)	Frontiers in Communication	Mixed methods	Ethical requirements for GenAI brand content
19	Grigsby, Mochesne & Zamudio (2025)	Journal of Retailing and Consumer Services	Experiment	AI disclosures reduce trust in service ads
20	Hardcastle, Verdier & Brower (2025)	Journal of Advertising	Mixed methods	AI personalization distrust in customer journeys
21	Israfilzade (2025)	Equilibrium. Quarterly Journal of Economics and Economic Policy	Experiment	AI vs human: ads, trust and purchase
22	Jung, Noghan, Lee & Kwon (2025)	Journal of Retailing and Consumer Services	Experiment	Trust and humanness in luxury AI ads
23	Kim (2025)	Journal of Interactive Advertising	Conceptual	Research agenda: AI, authenticity, and trust
24	Kirk & Givi (2025)	Journal of Business Research	Experiment	AI authorship, moral disgust, consumer responses
25	Worting & Vaisviari (2025)	Journal of Interactive Advertising	Experiment	AI disclosures affect ad and org trust
26	Liu, Osman, Lian & Ab Hamid (2025)	Journal of Theoretical and Applied Electronic Commerce Research	Survey	AI literacy moderates sponsored vlog purchases
27	Li et al. (2025)	Journal of Theoretical and Applied Electronic Commerce Research	Experiment	AIGC triggers authenticity concerns and disgust
28	Liu, Lian & Osman (2025)	Journal of Promotion Management	Experiment	AI disclosure reshapes quality perception in vlogs
29	Pierre (2025)	Journal of Current Issues & Research in Advertising	Survey	AI attitudes shape brand perception outcomes
30	Qiu et al. (2025)	Journal of Theoretical and Applied Electronic Commerce Research	Experiment	AI disclosure inhibits cause-related marketing intent
31	Su (2025)	AI Magazine	Survey	Virtual influencers build Gen Z trust
32	Wang, Sadia & Shurong (2025)	Journal of Consumer Marketing	Experiment	Anthropomorphism and transparency in GenAI marketing
33	Ni et al. (2025)	Journal of Theoretical and Applied Electronic Commerce Research	Survey	Trust-risk pathways for AI e-commerce context
34	Liu, Lian & Osman (2026)	Journal of Retailing and Consumer Services	Experiment	AI involvement reduces curiosity and engagement
35	Nguyen et al. (2026)	Journal of Global Scholars of Marketing Science	Survey	Cross-national AI ad effects on brands

In terms of publication outlets, the included studies span 19 distinct journals. The Journal of Retailing and Consumer Services contributes the largest share (n = 6), followed by the Journal of Advertising (n = 4), the Journal of Theoretical and Applied Electronic Commerce Research (n = 4), the Journal of Interactive Advertising (n = 3), and the Journal of Business Research (n = 3). The remaining 15 studies appear across outlets in advertising, hospitality, consumer marketing, and information systems, reflecting the interdisciplinary character of the research domain.

Experiments constitute the dominant methodology (n = 23, approximately 70%), followed by

survey designs (n = 5), mixed-methods studies (n = 2), conceptual and review papers (n = 2), and one prior systematic literature review. The heavy reliance on experimental designs reflects the field's orientation toward causal inference regarding disclosure and authorship manipulations. Geographically, most studies draw on Western samples (primarily the United States and Europe), with limited representation from East Asia (China, South Korea) and only one explicitly cross-national investigation. The Global South remains almost entirely absent from the evidence base.

4.2 Theme 1: The Disclosure Dilemma

AI disclosure – the explicit labeling of content as machine-generated – is the single most frequently examined antecedent in the corpus, investigated in nine of the included studies. The prevailing pattern is clear: disclosure activates consumer skepticism and erodes trust-related outcomes. Bui (2025) demonstrate that revealing AI involvement in prosocial advertising reduces ad credibility and purchase intentions through persuasion knowledge activation. Similarly, Koning and Voorveld (2025) find that disclosure diminishes trust in both advertisements and the organizations behind them, suggesting that the reputational cost extends beyond the individual message.

This negative disclosure effect replicates across varied marketing contexts. Qiu et al. (2025) report that AI disclosure inhibits purchase intentions specifically in cause-related marketing campaigns, where consumers appear to hold heightened expectations of human sincerity. In service advertising, the intangibility of the offering intensifies the problem: Grigsby et al. (2025) find that although 75% of consumers report favoring disclosure as a matter of principle, the same disclosures reduce their trust in the advertised service. Bui (2025) extends this pattern by showing that disclosure diminishes perceived advertising value, which in turn suppresses purchase intentions. In the context of AI-generated sponsored vlogs, Liu, Lian, and Osman (2025a) similarly demonstrate that AI disclosure affects perceived content quality and online shopping intentions. Wortel et al. (2024) confirm these findings in social media contexts, documenting how AI disclosures on Instagram advertisements trigger persuasion knowledge and attribution processes that shift consumer attitudes downward.

In contrast, the evidence is not uniformly negative. Kirkby et al. (2023) report that AI disclosure did not significantly affect brand attitudes or perceived authenticity, a finding that stands as a notable boundary condition. One possible explanation is that brand strength moderates the relationship: when established brands disclose AI involvement, consumers may perceive the disclosure as a signal of competence rather than deception. Wu et al. (2024) add further nuance, demonstrating that disclosure effects on word-of-mouth intentions vary depending on whether the AI was deployed for creative versus analytical tasks. Consumers penalize AI authorship more severely when the task is perceived as requiring human creativity and emotional investment.

Building on these findings, a synthesis across the corpus suggests that the disclosure-trust relationship follows a conditional pattern rather than a simple negative main effect. The direction and magnitude of the effect depend on content domain, product category, and disclosure framing, a set of moderating conditions explored further in Section 4.4. Notably, the evidence indicates that the format of disclosure itself matters. A bare label stating "Generated by AI" may activate persuasion knowledge more readily than a brand-voice disclosure framing such as "Our team used AI tools to enhance this content," which signals organizational honesty while preserving attributions of human oversight.

The null finding reported by Kirkby et al. (2023) is consistent with this interpretation, as their experimental stimuli employed a brand-voice transparency approach rather than a stark algorithmic label. This distinction between label-only and brand-voice disclosure formats represents a potentially important boundary condition that future research should examine directly.

4.3 Theme 2: Authenticity as Central Mediator

Perceived authenticity emerges as the most consistently identified mechanism through which AI-generated content influences downstream consumer responses. Seven studies in the corpus position authenticity as a formal mediator, and several additional studies measure it as a dependent variable. The pattern across these investigations points toward a dominant causal chain: awareness of AI authorship reduces perceived authenticity, which in turn diminishes trust, brand attitudes, and behavioral intentions.

Bruns and Meissner (2024) provide foundational evidence through three experiments showing that generative AI content creation diminishes perceived brand authenticity, which mediates negative follower reactions on social media. Kirk and Givi (2025) extend this work across seven experiments, identifying a dual-process pathway in which AI authorship reduces both perceived authenticity (a cognitive evaluation) and triggers moral disgust (an affective response), with each pathway independently suppressing attitudes, word-of-mouth intentions, and brand loyalty. The identification of moral disgust as a parallel mediator is distinctive within the corpus and connects the AI-content literature to moral psychology.

This pattern extends to specific industry contexts. In the restaurant industry, AI disclosure undermines perceived brand authenticity, which mediates subsequent declines in brand image and consumer behavior (Ali et al., 2025). Zhang & Hur (2025) replicate the authenticity-mediation pathway in visual advertising, showing that AI-generated images with disclosure erode authenticity perceptions and, through them, brand trust. Consumer reactions to perceived ChatGPT usage in online reviews follow the same logic: reviews suspected of AI generation are rated as less authentic, less trustworthy, and less useful (Amos & Zhang, 2024). Even when the level of AI involvement varies, the authenticity mechanism persists; one study finds that higher degrees of AI involvement progressively diminish perceived authenticity, creativity, and emotional depth, reducing content engagement (Liu, Lian, & Osman, 2025a).

Beyond authenticity, the corpus identifies several related mediating constructs that operate through parallel psychological pathways. Perceived humanness serves as a mediator in luxury advertising, where artificial creativity influences consumer trust and perceived humanness simultaneously (Jung et al., 2025). Social presence mediates the effects of anthropomorphic design cues on purchase intention and trust in generative AI marketing communication (Wang et al., 2025). Emotional trust functions as a mediator in hospitality contexts, where the mere use of the term "artificial intelligence" in product descriptions lowers emotional trust and suppresses purchase intentions (Cicek et al., 2024). Perceived risk operates as a complementary pathway on e-commerce platforms, where trust and risk jointly determine purchase behavior in response to AI-generated content (Yu et al., 2025).

Taken together, the mediator evidence answers RQ2 with considerable consistency. The dominant pathway runs from AI authorship awareness through reduced perceived authenticity to lower trust

and diminished behavioral intentions. Secondary pathways through perceived humanness, social presence, moral disgust, and emotional trust enrich the picture but have each been documented in only one or two studies, indicating avenues for replication.

4.4 Theme 3: Moderating Conditions

Findings across the corpus are heterogeneous, signaling that the effects of AI-generated content on consumer trust are not uniform. Thirteen studies identify explicit moderating variables, which can be organized into three categories: content-level, consumer-level, and contextual moderators.

At the content level, the type of appeal and the nature of the task emerge as important boundary conditions. Chen et al. (2024) demonstrate that emotional versus rational advertising appeals yield different consumer responses to AI-generated ads, with emotional appeals intensifying negative reactions because they invoke expectations of human warmth and sincerity. Wu et al. (2024) further show that consumers evaluate AI involvement differently when the task is creative (e.g., copywriting, design) versus analytical (e.g., data-driven product recommendations). AI disclosure reduces word-of-mouth intentions for creative tasks but not for analytical ones, suggesting that perceived appropriateness of AI use shapes consumer tolerance. The degree of AI involvement also matters: fully AI-generated content elicits stronger negative reactions than AI-assisted content, where human oversight is perceived to remain (Liu, Lian, & Osman, 2025a). Product category introduces additional variation, with services (particularly intangible offerings in hospitality) generating heightened trust concerns relative to tangible goods (Grigsby et al., 2025; Cicek et al., 2024).

Beyond these binary comparisons, consumer reactions appear to differ not merely between the poles of "AI-generated" and "human-created" but along a graduated spectrum of perceived AI involvement. This spectrum runs from minor AI assistance, through substantive human-AI collaboration, to fully autonomous AI generation. Liu, Lian, and Osman (2025a) provide initial evidence that increasing AI involvement progressively diminishes authenticity perceptions. Taken alongside the task-type findings of Wu et al. (2024), the pattern is consistent with a threshold model: below a certain level of perceived AI contribution, the disclosure penalty appears negligible, but once involvement crosses a critical boundary, consumer reactions shift sharply rather than scaling in proportion. Identifying where this threshold lies, and how it varies across content types and consumer segments, remains an open empirical question with substantial practical significance.

At the consumer level, individual differences moderate the AI content-trust relationship in several ways. AI literacy shapes how consumers process information from AI-generated sponsored vlogs, moderating the link between information adoption and purchase intention (Liu, Osman, et al., 2025b). Consumers with higher AI literacy may evaluate AI-generated content more critically yet also more accurately, producing a more balanced rather than uniformly negative response. Pre-existing attitudes toward AI in general influence brand evaluations when brands disclose their use of the technology (Pierre, 2025). Self-efficacy operates as an additional consumer-level moderator, interacting with appeal type to alter trust and perceived humanness judgments (Chen et al., 2024). Generational differences further segment the audience: Millennials and Gen Z demonstrate greater receptivity to AI-driven virtual influencers and the content they produce (Su, 2025).

At the contextual level, culture, brand transparency strategy, and anthropomorphic design cues shape

the strength and direction of AI content effects. The sole cross-national study in the corpus reveals significant differences in how consumers across countries respond to AI-generated advertising, pointing toward cultural dimensions (such as uncertainty avoidance and technology acceptance norms) as underexplored explanatory factors (Nguyen et al., 2026). Brand transparency framing moderates the relationship between AI use and consumer perceptions, with transparent communication about AI involvement partially offsetting negative reactions (Pierre, 2025). Anthropomorphism emerges as a particularly promising buffering mechanism: two studies show that endowing AI-generated content or its source with human-like characteristics enhances social presence, perceived authenticity, and trust (Lee & Kim, 2024; Wang et al., 2025). However, a qualitative comparative analysis cautions that such anthropomorphic strategies must be balanced against ethical transparency requirements to avoid perceptions of manipulation (Du Plessis, 2025).

4.5 Theme 4: Downstream Outcomes

Across the 35 studies, a broad range of behavioral and attitudinal outcomes affected by AI-generated marketing content, with purchase intention and brand trust measured most frequently. Purchase intention appears as a dependent variable in 21 studies, making it the single most examined outcome. Brand trust follows closely (n = 18), and advertising attitude is measured in 12 studies. Israfilzade (2025) provides direct experimental evidence that AI-generated advertisements reduce both consumer trust and purchase intent relative to human-created counterparts, reinforcing the centrality of these two outcome variables. This concentration reflects the field's applied orientation toward predicting consumer behavior in marketplace contexts.

Beyond these primary outcomes, several studies trace the effects of AI-generated content to word-of-mouth and sharing intentions. Kirk and Givi (2025) show that AI authorship reduces willingness to share and recommend marketing content, an effect that flows through both the authenticity and moral disgust pathways. Wu et al. (2024) replicate the word-of-mouth finding but demonstrate that the effect is conditional on task type. Brand attitude and brand image are measured as outcomes in studies spanning product branding (Kirkby et al., 2023), cross-national advertising (Nguyen et al., 2026), and restaurant marketing (Ali et al., 2025), with the general pattern showing negative or null effects of AI disclosure on these constructs.

Several studies extend the outcome space beyond traditional marketing metrics. Consumer engagement behaviors – including clicks, likes, and follows – decline when AI involvement is disclosed (Bruns & Meissner, 2024; Liu, Lian, & Osman, 2025a). Ethical judgments represent a distinct outcome category: Li et al. (2024) examine how AI identity disclosure influences consumer unethical behavior through social judgment processes, while Kirk and Givi (2025) document moral disgust as both a mediator and an outcome in its own right. Customer experience quality is diminished when AI-driven personalization triggers distrust (Hardcastle et al., 2025), and donation intentions decline in AI-generated charitable advertising (Arango et al., 2023). The pattern that emerges across these diverse outcomes supports a trust-centered causal chain: AI awareness feeds into authenticity and credibility judgments, which shape trust, which in turn determines behavioral intentions across purchasing, sharing, engaging, and donating contexts.

4.6 Tensions and Contradictions

Three unresolved tensions run through the corpus and resist straightforward synthesis, though a closer examination of boundary conditions offers partial resolution for each.

The first is the disclosure paradox. Consumers consistently report preferring transparency about AI involvement – Grigsby et al. (2025) find that 75% of respondents favor disclosure – yet the same disclosures reduce their trust and purchase intentions. This paradox places brands in a double bind that intensifies as regulatory mandates such as the EU AI Act make disclosure increasingly compulsory. However, the paradox becomes easier to address once disclosure is disaggregated by format. The studies reporting strong negative effects predominantly employ label-only disclosure conditions (a simple statement that content was "AI-generated" or "created by AI") which functions as a persuasion knowledge trigger. In contrast, the null finding of Kirkby et al. (2023) involved a brand-voice transparency framing in which the brand proactively communicated its use of AI tools. This distinction suggests that the disclosure paradox may dissolve when the analysis moves from a binary disclosure-versus-nondisclosure framework to a more granular examination of how disclosure is framed. Label-only disclosure activates skepticism; brand-voice disclosure may instead signal organizational honesty.

A second tension concerns the inconsistency between null and negative disclosure effects. Kirkby et al. (2023) find no significant impact of AI disclosure on brand attitudes or perceived authenticity, a result that contrasts with the majority of studies reporting negative effects (Bui, 2025; Qiu et al., 2025; Zhang & Hur, 2025). The divergence appears dependent on at least three factors operating simultaneously. Brand strength may serve as a buffer: established brands with pre-existing authenticity capital can absorb disclosure risk that would damage less familiar brands. Content type further conditions the effect, with negative results concentrated in emotional, prosocial, and cause-related contexts where expectations of human sincerity are highest. Disclosure format, as discussed above, adds a third layer. These factors likely interact, such that a strong brand disclosing AI assistance in informational content through a brand-voice frame may face negligible backlash, while an unfamiliar brand disclosing full AI authorship of emotional content faces compounding penalties. The corpus does not yet contain factorial designs that test these interactions jointly, but the pattern of available results is consistent with this multi-layered conditional account.

A third tension involves authenticity restoration through anthropomorphism. Two studies demonstrate that human-like cues can partially restore perceived authenticity and trust when content is AI-generated (Lee & Kim, 2024; Wang et al., 2025). Yet Du Plessis (2025) warns that such strategies risk crossing into manipulation, potentially producing a delayed backlash when consumers recognize the anthropomorphic framing as itself inauthentic. The tension may be explained by a nonlinear relationship between humanlikeness and consumer acceptance. At moderate levels, anthropomorphic cues enhance social presence, warmth, and perceived sincerity without triggering skepticism. At excessive levels, the gap between the apparent humanness of the communication and the known machine origin may produce uncanny valley discomfort or, worse, perceptions of deliberate deception. The field has not yet established where the tipping point lies, nor how it shifts across product categories and consumer segments, but the available evidence indicates that anthropomorphism is a balancing problem rather than a simple buffering strategy.

4.7 Conceptual Framework

Figure 2 presents an integrative framework that synthesizes the findings across all 35 included studies into an input-process-output model. On the left side, antecedents are organized into three tiers: content-level factors (AI authorship, disclosure presence, content type, AI involvement level, and appeal type), consumer-level factors (AI literacy, general AI attitudes, self-efficacy, and generational cohort), and contextual factors (product category, culture, and regulatory environment).
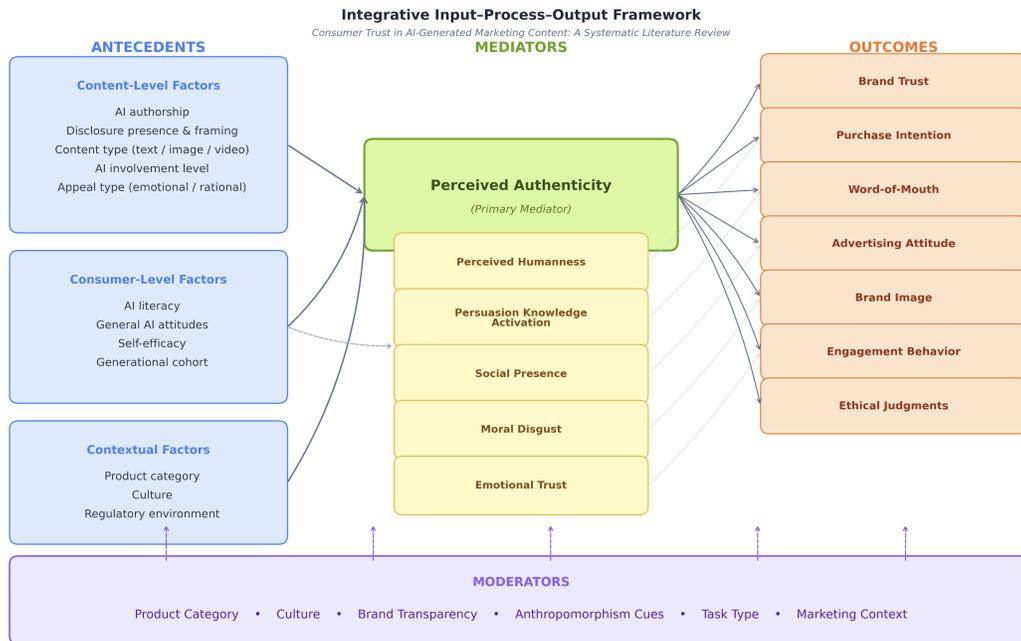

Integrative Input–Process–Output Framework
Consumer Trust in AI-Generated Marketing Content: A Systematic Literature Review

Figure 2. Integrative Input-Process-Output Framework of Consumer Trust Responses to AI-Generated Marketing Content

These antecedents feed into a set of mediating mechanisms positioned at the center of the framework. Perceived authenticity occupies the primary mediating role, supported by empirical evidence across seven studies. The mediating mechanisms operate through three distinguishable classes: cognitive pathways, including persuasion knowledge activation and source credibility assessment; affective pathways, including moral disgust and emotional trust; and perceptual pathways, including perceived authenticity, perceived humanness, and social presence. These classes are not mutually exclusive – in many consumer encounters with AI-generated content, cognitive, affective, and perceptual processes operate simultaneously, with boundary conditions determining which pathway dominates. On the right side, outcomes span trust in advertisements and brands, purchase intention, word-of-mouth, advertising attitude, brand image, engagement behavior, and ethical judgments.

Moderating conditions, including product category, culture, brand transparency framing, anthropomorphism cues, task type, and marketing context, are positioned to influence both the antecedent-to-mediator and mediator-to-outcome linkages. The framework highlights that the pathway with the strongest cumulative support runs from AI disclosure through reduced perceived authenticity to diminished brand trust and suppressed purchase intentions. Pathways through moral disgust and perceived humanness carry preliminary support and represent high-priority targets for future investigation, as discussed in Section 6.

5. Discussion

5.1 Theoretical Implications

Several implications for theory development at the intersection of artificial intelligence and consumer behavior emerge from this systematic review. First, the corpus confirms and extends the Persuasion Knowledge Model (Friestad & Wright, 1994) by demonstrating that AI disclosure functions as a novel persuasion-knowledge trigger in marketing communications. Multiple studies point to the finding that labeling content as AI-generated activates consumer coping mechanisms, which in turn reduce credibility and purchase intentions (Bui, 2025; Wortel et al., 2024; Qiu et al., 2025). Yet the Persuasion Knowledge Model alone cannot account for the heterogeneity observed across the evidence base. Kirkby et al. (2023) found no significant effect of AI disclosure on brand attitudes or perceived authenticity, suggesting that persuasion knowledge activation does not invariably translate into negative evaluations. This boundary condition points to the need to augment the Persuasion Knowledge Model with authenticity theory. Specifically, the negative consequences of persuasion knowledge activation appear to depend on whether the content domain requires perceived human intentionality – emotional and creative content – or whether it serves primarily informational and functional purposes, where machine authorship may be tolerated or even welcomed.

Second, perceived authenticity emerges from the present synthesis as the central construct in the AI-content-trust relationship. Across the reviewed studies, authenticity is the most frequently identified mediator linking AI authorship awareness to downstream consumer outcomes such as brand trust, purchase intention, and word-of-mouth (Bruns & Meissner, 2024; Kirk & Givi, 2025; Zhang & Hur, 2025). This pattern suggests that trust itself may be more accurately characterized as a downstream consequence of authenticity judgments rather than the proximal cognitive casualty of AI-generated content. The implication for brand management theory is substantial: as content creation is increasingly delegated to generative AI systems, the authenticity construct requires reconceptualization. Traditional conceptualizations of brand authenticity – grounded in genuineness, originality, and sincerity (Morhart et al., 2015) – assumed human creative agency. In an era of machine-generated communications, scholars must develop revised authenticity frameworks that accommodate hybrid human-AI content production without rendering the construct theoretically empty. One productive direction is the recognition that authenticity in marketing communications may involve two separable dimensions: content authenticity (is the output genuine and consistent with brand identity?) and process authenticity (was it created with genuine human intent?). This distinction, not yet formalized in the literature, could explain why high-quality AI outputs still incur authenticity penalties.

Third, the moral disgust pathway identified by Kirk and Givi (2025) introduces an affective dimension that enriches the predominantly cognitive accounts dominating the literature. Across seven experiments, these authors demonstrated that AI-authored emotional marketing reduces consumer attitudes, word-of-mouth intentions, and brand loyalty through a dual pathway: diminished perceived authenticity (cognitive) and heightened moral disgust (affective). This dual-process account resonates with moral foundations theory and suggests that consumers may perceive AI-generated emotional appeals as a form of moral transgression – a manipulation that violates expectations of sincerity and human care. Jiang et al. (2025) similarly find that AI-generated advertising content

triggers moral disgust unless actively mitigated. Future theorizing should integrate disgust sensitivity and moral psychology to specify the conditions under which affective reactions dominate cognitive appraisals.

Fourth, the substantial heterogeneity of findings across the 35 reviewed studies points toward a contingency theory of AI content effects. The negative impact of AI-generated content on trust is not universal or uniform; it depends on content type (emotional versus rational appeals), consumer characteristics (AI literacy, generational cohort, pre-existing AI attitudes), and contextual factors (culture, product category, anthropomorphism cues). This conditional perspective aligns with recent calls for nuanced, context-sensitive research on AI in marketing communications (Kim, 2025). The integrative framework proposed in Section 4.7 systematizes these conditional relationships and may serve as a foundation for developing a middle-range theory that specifies not merely whether AI-generated content harms trust but when, for whom, and through which mechanisms.

5.2 Managerial Implications

The synthesis yields several actionable recommendations for marketing practitioners navigating the tension between AI-enabled efficiency and consumer trust preservation. With respect to disclosure strategy, the evidence reveals a paradox: consumers express a preference for transparency – 75% favor disclosure of AI involvement in advertising (Grigsby et al., 2025) – yet disclosure consistently reduces trust and credibility. Given that regulatory mandates under the EU AI Act and FTC guidance are making disclosure increasingly unavoidable, managers cannot resolve this tension through non-disclosure. Instead, the evidence suggests that how AI involvement is disclosed matters as much as whether it is disclosed. Framing AI use as "AI-assisted, human-reviewed" rather than "AI-generated" may reduce negative reactions by preserving attributions of human oversight and intentionality. The finding that AI involvement level moderates consumer responses (Liu, Lian, & Osman, 2025a) lends preliminary support to this framing approach.

Regarding content-type segmentation, the reviewed evidence suggests that firms should differentiate their AI deployment strategy by content function. AI-generated content may be deployed with relatively lower risk for informational and functional tasks (product specifications, frequently asked questions, data-driven reports) where consumer authenticity expectations are modest. By contrast, emotional, creative, and purpose-driven content, such as brand storytelling, cause-related campaigns, and charitable appeals (Arango et al., 2023; Bui, 2025; Qiu et al., 2025), should retain substantial human involvement. The differential sensitivity of prosocial and cause-related contexts to AI disclosure effects reinforces this recommendation.

Anthropomorphism offers a partial buffer against trust erosion. Studies demonstrate that imbuing AI-generated content with human-like characteristics (personalized language, relatable tone, anthropomorphic visual cues) can enhance perceived authenticity and social presence, partially restoring consumer trust (Lee & Kim, 2024; Wang et al., 2025). However, practitioners must balance this strategy against ethical transparency requirements (Du Plessis, 2025), as excessive anthropomorphism of AI-generated content without disclosure could constitute a form of deception. The collaboration spectrum identified in Section 4.4 offers an additional mitigation pathway: positioning AI as a tool under human editorial direction, rather than as the autonomous author, may preserve both transparency and authenticity perceptions simultaneously.

Finally, AI literacy represents an emerging segmentation variable. As consumer familiarity with generative AI grows, reactions to AI-generated content may shift (Liu, Osman, et al., 2025b). Brands operating in tech-savvy markets or targeting digitally native cohorts (Su, 2025) may find that AI disclosure carries less stigma, whereas audiences with lower AI literacy may require more careful framing. Service brands and hospitality firms face heightened trust concerns due to the intangibility of their offerings (Grigsby et al., 2025; Cicek et al., 2024) and should exercise particular caution in adopting AI-generated customer-facing content.

6. Research Agenda

The gaps, contradictions, and underexplored areas identified in this review give rise to five research propositions intended to guide future inquiry.

First, the question of temporal dynamics deserves priority attention. All 35 studies in the present corpus employ cross-sectional designs. No study tracks how consumer reactions evolve as AI-generated content becomes normalized in everyday media consumption. The null finding reported by Kirkby et al. (2023) may already reflect early signs of habituation among technology-literate consumers. Longitudinal and panel designs are needed to separate novelty-driven skepticism from enduring trust deficits, and to determine whether the emotional content boundary condition identified by Kirk and Givi (2025) holds as general AI familiarity increases. This pattern of evidence leads to the expectation that as consumer exposure to AI-generated marketing content increases over time, the negative effect of AI disclosure on perceived authenticity will weaken through a habituation process, although the moderating role of content emotionality is likely to persist.

Second, cross-cultural variation remains severely underexplored. Only one study in the corpus, a cross-national analysis (Nguyen et al., 2026), explicitly examines cultural variation in consumer responses to AI-generated advertising. Given that AI adoption rates, regulatory environments, and cultural orientations toward technology differ markedly across markets, cross-cultural research represents a critical gap. Hofstede's uncertainty avoidance dimension offers a theoretically grounded starting point, as consumers in high uncertainty-avoidance societies may be more sensitive to the ambiguity that machine authorship introduces into brand communications. Comparative studies spanning Western, East Asian, and Global South markets would substantially advance the field. Taken together, the available evidence supports the expectation that the negative effect of AI-generated marketing content on consumer trust is stronger in high uncertainty-avoidance cultures than in low uncertainty-avoidance cultures, and that this relationship is mediated by differential persuasion knowledge activation.

Third, modality-specific effects represent a significant gap. The reviewed studies examine text-based, image-based, and mixed content, yet no study directly compares modalities within a single experimental design. Visual content may trigger uncanny valley reactions (Mori, 1970/2012), producing subtle perceptual discomfort arising from near-but-not-quite-human imagery that intensifies distrust beyond what text-based findings would predict. As AI-generated video (deepfakes, synthetic spokespersons) proliferates in advertising, understanding modality-specific trust dynamics becomes increasingly urgent. Zhang & Hur (2025) provide initial evidence that AI-generated images reduce authenticity perceptions, but direct text-versus-image-versus-video comparisons remain absent from

the literature. On this basis, there is reason to expect that AI-generated visual content (images and video) will elicit stronger negative trust responses than AI-generated textual content, because visual modalities activate uncanny valley perceptions and carry higher authenticity expectations.

Fourth, disclosure framing warrants systematic experimental investigation. The disclosure paradox documented across the corpus (consumers want transparency yet penalize brands that provide it) calls for research that moves beyond the binary of disclosure versus nondisclosure. Preliminary evidence that varying levels of AI involvement elicit different consumer reactions (Liu, Lian, & Osman, 2025a) suggests that the framing of "AI-generated" versus "human-created" may be insufficiently nuanced. Co-creation narratives that position AI as a tool under human editorial direction could preserve the transparency benefits of disclosure while maintaining attributions of human intentionality and oversight. Additionally, the emerging regulatory environment creates a natural distinction between mandatory and voluntary disclosure that has not been tested empirically. These considerations suggest that disclosure framing effects warrant direct experimental comparison, with the interaction between disclosure voluntariness (mandatory versus voluntary) and disclosure format (label-only versus brand-voice) representing a high-priority research design.

Fifth, the near-total reliance on stated intentions rather than observed behaviors represents a significant methodological limitation. Virtually all dependent variables in the reviewed corpus are stated intentions rather than observed behaviors. No study reports actual purchase data, real click-through rates, or observed engagement metrics in response to disclosed AI content. This limitation is important because the gap between stated intentions and actual behavior is well-documented in consumer research, and there are reasons to expect this gap to be especially pronounced for evaluations of AI-generated content. Field experiments on e-commerce platforms, A/B tests with actual engagement metrics, and conjoint analyses with revealed-preference validation are needed to determine whether the negative effects documented in the experimental literature translate to real marketplace behavior. The available evidence suggests that stated intention measures may overestimate the actual behavioral impact of AI disclosure, and field experiments with real purchase data are needed to test this expectation.

7. Limitations and Conclusion

7.1 Limitations

Several limitations constrain the conclusions that may be drawn from this review. The search strategy relied on two databases, Scopus and Google Scholar, and the inclusion of Web of Science, PsycINFO, or EBSCO might have surfaced additional relevant studies. The restriction to English-language publications excludes scholarship from markets where AI marketing adoption is particularly advanced, including China and South Korea, where considerable research is published in local languages. Given that 91% of the included studies were published between 2024 and 2026, the evidence base reflects a rapidly evolving field; the conclusions presented here may require revision as the literature matures and replication studies accumulate. Publication bias represents a further concern, as journals may disproportionately publish studies reporting significant effects, and null findings such as those of Kirkby et al. (2023) may be underrepresented. Finally, the thematic coding process, while structured and systematic, inevitably involves interpretive judgment, and alternative

categorization schemes could yield different emphases.

7.2 Conclusion

This systematic review set out to answer how AI-generated marketing content affects consumer trust-related responses and under what conditions those effects manifest. Across 35 empirical and conceptual studies, the evidence points to a central finding: AI-generated content poses a genuine and measurable challenge to consumer trust, operating primarily through the erosion of perceived authenticity. Disclosure of AI authorship activates persuasion knowledge, which diminishes credibility, brand attitudes, and behavioral intentions – yet these effects are not universal and not fixed. Content type, consumer AI literacy, cultural context, anthropomorphism cues, and disclosure framing each moderate the strength and direction of the relationship. The integrative framework advanced in this review positions perceived authenticity as the core construct and identifies moral disgust as a parallel affective pathway that warrants sustained theoretical attention. The five research propositions chart a course toward longitudinal, cross-cultural, modality-comparative, and behaviorally grounded research that the field presently lacks. As generative AI transitions from a competitive novelty to an operational default in marketing communications, the question confronting both scholars and practitioners is no longer whether AI-generated content affects trust, but how organizations can preserve authenticity in an age when the author may not be human at all.

References

1. Ali, L., Ali, F., Abdalla, M. J., & Alotaibi, S. (2025). Beyond the hype: Evaluating the impact of generative AI on brand authenticity, image, and consumer behavior in the restaurant industry. *International Journal of Hospitality Management*, *131*, 104318.

2. Amos, C., & Zhang, L. (2024). Consumer reactions to perceived undisclosed ChatGPT usage in an online review context. *Telematics and Informatics*, *93*, 102163. https://doi.org/10.1016/j.tele.2024.102163

3. Arango, L., Singaraju, S. P., & Niininen, O. (2023). Consumer responses to AI-generated charitable giving ads. *Journal of Advertising*, *52*(4), 486–503. https://doi.org/10.1080/00913367.2023.2183285

4. Bruns, J. D., & Meissner, M. (2024). Do you create your content yourself? Using generative artificial intelligence for social media content creation diminishes perceived brand authenticity. *Journal of Retailing and Consumer Services*, *79*, 103790. https://doi.org/10.1016/j.jretconser.2024.103790

5. Bui, H. T. (2025). Examining the effect of AI advertising involvement disclosure on advertising value and purchase intentions. *Journal of Research in Interactive Marketing*. https://doi.org/10.1108/JRIM-02-2025-0066

6. Chaudhuri, A., & Holbrook, M. B. (2001). The chain of effects from brand trust and brand affect to brand performance: The role of brand loyalty. *Journal of Marketing*, *65*(2), 81–93.

7. Chen, Y., Wang, H., Hill, S. R., & Li, B. (2024). Consumer attitudes toward AI-generated ads: Appeal types, self-efficacy and AI's social role. *Journal of Business Research*, *185*, 114867. https://doi.org/10.1016/j.jbusres.2024.114867

8. Cicek, M., Gursoy, D., & Lu, L. (2024). Adverse impacts of revealing the presence of 'Artificial Intelligence (AI)' technology in product and service descriptions on purchase intentions. *Journal of Hospitality Marketing & Management*, *34*(1), 1–23. https://doi.org/10.1080/19368623.2024.2368040

9. Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of Marketing Science*, *48*(1), 24–42.

10. Delgado-Ballester, E. (2004). Applicability of a brand trust scale across product categories. *European Journal of Marketing*, *38*(5/6), 573–592.

11. Du Plessis, C. (2025). Ethical requirements for generative AI in brand content creation: A qualitative comparative analysis. *Frontiers in Communication*, *10*, 1523077. https://doi.org/10.3389/fcomm.2025.1523077

12. Ford, J., Jain, V., Wadhwani, K., & Gupta, D. G. (2023). AI advertising: An overview and guidelines. *Journal of Business Research*, *166*, 114124. https://doi.org/10.1016/j.jbusres.2023.114124

13. Friestad, M., & Wright, P. (1994). The persuasion knowledge model: How people cope with persuasion attempts. *Journal of Consumer Research*, *21*(1), 1–31.

14. Grigsby, J. L., Michelsen, M., & Zamudio, C. (2025). Service ads in the era of generative AI: Disclosures, trust, and intangibility. *Journal of Retailing and Consumer Services*, *84*, 104231. https://doi.org/10.1016/j.jretconser.2025.104231

15. Hardcastle, K., Vorster, L., & Brown, D. M. (2025). Understanding customer responses to AI-driven personalized journeys: Impacts on the customer experience. *Journal of Advertising*. https://doi.org/10.1080/00913367.2025.2460985

16. Israfilzade, K. (2025). AI-generated versus human-created advertising: Effects on consumer trust and purchase intent. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, *20*(4), 1301–1337. https://doi.org/10.24136/eq.4038

17. Jiang, S., Zheng, W., & Kong, H. (2025). Trust or skepticism? Unraveling the communication mechanisms of AIGC advertisements on consumer responses. *Journal of Theoretical and Applied Electronic Commerce Research*, *20*(4), 339. https://doi.org/10.3390/jtaer20040339

18. Jung, T., Koghut, M., Lee, E., & Kwon, O. (2025). Artificial creativity in luxury advertising: How trust and perceived humanness drive consumer response to AI-generated content. *Journal of Retailing and Consumer Services*, *87*, 104403. https://doi.org/10.1016/j.jretconser.2025.104403

19. Kim, J. (2025). Advertising in the age of agentic AI: Call for research. *Journal of Interactive Advertising*. https://doi.org/10.1080/15252019.2025.2557107

20. Kirk, C. P., & Givi, J. (2025). The AI-authorship effect: Understanding authenticity, moral disgust, and consumer responses to AI-generated marketing communications. *Journal of Business Research*, *186*, 114984. https://doi.org/10.1016/j.jbusres.2024.114984

21. Kirkby, A., Baumgarth, C., & Henseler, J. (2023). To disclose or not disclose, is no longer the question – effect of AI-disclosed brand voice on brand authenticity and attitude. *Journal of Product & Brand Management*, *32*(7), 1108–1122. https://doi.org/10.1108/JPBM-02-2022-3864

22. Koning, B., & Voorveld, H. A. M. (2025). Disclaimer! This content is AI-generated: How AI-disclosures influence trust in advertisements and organizations. *Journal of Interactive Advertising*. https://doi.org/10.1080/15252019.2025.2554149

23. Lee, G., & Kim, H.-Y. (2024). Human vs. AI: The battle for authenticity in fashion design and consumer response. *Journal of Retailing and Consumer Services*, *77*, 103690. https://doi.org/10.1016/j.jretconser.2023.103690

24. Li, T.-G., Zhang, C.-B., Chang, Y., & Zheng, W. (2024). The impact of AI identity disclosure on consumer unethical behavior: A social judgment perspective. *Journal of Retailing and Consumer Services*, *76*, 103606. https://doi.org/10.1016/j.jretconser.2023.103606

25. Liu, Q., Lian, Z., & Osman, L. H. (2025a). Can artificial intelligence-generated sponsored vlogs trigger online shopping? *Journal of Promotion Management*. https://doi.org/10.1080/10496491.2025.2513323

26. Liu, Q., Osman, L. H., Lian, Z., Ab Hamid, S. N., & Che Wel, C. A. (2025b). From perception to purchase: How AI literacy shapes consumer decisions in AI-generated sponsored vlogs. *Journal of Theoretical and Applied Electronic Commerce Research*, *20*(4), 302. https://doi.org/10.3390/jtaer20040302

27. Morhart, F., Malar, L., Guevremont, A., Girardin, F., & Grohmann, B. (2015). Brand authenticity: An integrative framework and measurement scale. *Journal of Consumer Psychology*, *25*(2), 200–218.

28. Mori, M. (2012). The uncanny valley (K. F. MacDorman & N. Kageki, Trans.). *IEEE Robotics & Automation Magazine*, *19*(2), 98–100. (Original work published 1970)

29. Nguyen, K. M., Phan, T. M., Tran, Y. N. N., Nguyen, A. T., Nguyen, T. L. N., Hoang, G. H., Tran, T. T., & Nguyen, N. T. (2026). Evaluating the efficacy of AI-generated advertising: A cross-national analysis of customer responses on brand perceptions. *Journal of Global Scholars of Marketing Science*. https://doi.org/10.1080/21639159.2026.2617659

30. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hrobjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71.

31. Pierre, L. (2025). Exploring how brand use of artificial intelligence influences consumer perceptions and behavioral intentions through general AI attitudes and brand transparency. *Journal of Current Issues & Research in Advertising*. https://doi.org/10.1080/10641734.2025.2596011

32. Qiu, L., Wang, Y., Zeng, Y., & Cong, R. (2025). Artificial intelligence disclosure in cause-related marketing: A persuasion knowledge perspective. *Journal of Theoretical and Applied Electronic Commerce Research*, *20*(3), 193. https://doi.org/10.3390/jtaer20030193

33. Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339.

34. Su, B.-C. (2025). Navigating the new frontier: The role of AI-driven virtual influencers in consumer engagement. *AI Magazine*, *46*(2), e70012. https://doi.org/10.1002/aaai.70012

35. Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge by means of systematic review. *British Journal of Management*, *14*(3), 207–222.

36. Van Noort, G., Himelboim, I., Martin, J., & Collinger, T. (2020). Introducing a model of automated brand-generated content in an era of computational advertising. *Journal of Advertising*, *49*(4), 411–427. https://doi.org/10.1080/00913367.2020.1795954

37. Wang, Y., Sauka, K., & Situmeang, F. B. I. (2025). Anthropomorphism and transparency interplay on consumer behaviour in generative AI-driven marketing communication. *Journal of Consumer Marketing*, *42*(4), 512–536.

38. Wortel, C., Vanwesenbeeck, I., & Tomas, F. (2024). Made with artificial intelligence: The effect of artificial intelligence disclosures in Instagram advertisements on consumer attitudes. *Emerging Media*, *2*(3), 547–570. https://doi.org/10.1177/27523543241292096

39. Wu, L., Dodoo, N. A., & Wen, T. J. (2024). Disclosing AI's involvement in advertising to consumers: A task-dependent perspective. *Journal of Advertising*, *54*(1), 20–38. https://doi.org/10.1080/00913367.2024.2309929

40. Yu, T., Pan, Y., & Jang, W. (2025). Modeling consumer reactions to AI-generated content on e-commerce platforms: A trust-risk dual pathway framework. *Journal of Theoretical and Applied Electronic Commerce Research*, *20*(4), 257.

41. Zhang, L., & Hur, C. (2025). The impact of generative AI images on consumer attitudes in advertising. *Administrative Sciences*, *15*(10), 395.