JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Human–AI Collaboration in Digital Advertising: Effects on Perceived Authenticity, Consumer Trust, and Purchase Intentions

[1]K. Nageswara Rao,

[1]Lovely professional university,
[1]Computer science and engineering,
[1]144411, India.

Abstract : As generative artificial intelligence (AI) has become a key development in advertising production, a fundamental question for advertising scholarship is whether AI possession and collaboration boosts or undermines the outcomes of the advertising agency that aim to be shared with the consumer audience? This paper compiles and summarizes the still growing body of research on human–AI collaborative advertising to provide a comprehensive conceptual model for the relationship between human–AI collaboration level, perceived authenticity, consumer trust and purchase intention. Based on the existing literature of authenticity theory, the persuasion knowledge model, and trust-transfer logic, the paper proposes that the perceived authenticity and consumer trust act as sequential mediators between human–AI collaboration and purchase intention, and that the relationships between human–AI collaboration and purchase intention through the perceived authenticity and consumer trust both are moderated by AI-disclosure transparency, consumer AI literacy, and product involvement. A brief summary of the literature is presented, and 25 studies relevant to the hypotheses are grouped together from primarily 2002 to 2026. An empirical method based on a survey with validated multi-item scales is suggested in the future for empirical testing. The paper offers a theory-integrated framework, managerial implications for the disclosure practice of artificial intelligence, and an agenda for future research in this rapidly evolving field of the disclosure practice of AI.

***IndexTerms* - human–AI collaboration; generative AI advertising; perceived authenticity; consumer trust; purchase intention; AI disclosure; algorithm aversion**

I. INTRODUCTION

Within the blink of an eye, Generative AI has transitioned from an experimental novelty to a sought-after asset used in digital advertising, gaining significance in a relatively short time. Surveys indicate that nearly everyone in the marketing industry is currently incorporating AI tools into their content generation process, a choice attributed to the humanization of AI tools and the resulting efficiency gains and scalable content production. Surveys also show that the overwhelming majority of marketers are already integrating AI tools into their content creation processes, citing efficiency gains and scalable content production. According to Deloitte Digital's 2025 survey on U.S. consumers, a significant portion of brands already used generative AI in their marketing operations, with the survey's adopters stating that the use of the technology in their content production processes had led to a reduction in costs. And many of the adopters stated that they generate higher revenues compared to non-adopters, often at a faster rate. Meanwhile, a considerable share of U.S. adults is uncomfortable with AI-driven ads, with considerable demographic and ad content variation at play when it comes to comfort [32].

The conflict is at the heart of a rapidly emerging body of marketing research. GenAI can improve relevance by personalizing advertising based on user preferences, lifestyles, and cultural signals derived from data, on the other hand [3]. Consumers tend to feel that the image was AI-generated rather than human rather than the actual image of the advertising material, and thus perceive it as less authentic, less credible, and sometimes less convincing [2] [4] [22] [29]. Brüns and Meißner [2] argue that, for example, if social media content is created using generative AI and its authentic content is made clear, this has an adverse effect on the perceived brand authenticity. Israfilzade [1] demonstrates that consumers rated ads with the human-generated branding higher than ads branding them as AI-generated and that this impression deficit is larger for high-involvement products (electronics) than for low-involvement products (snacks).

But the wrote story isn't about an either-or proposition of the kind “AI bad, human good.” A stream of research parallel to the research mentioned above highlights that while AI may be able to reduce the authenticity risks, the negative results can be alleviated or even reversed by incorporating human creative input, and that this synergy must be shown to customers [4], [18]. Haupt, Freidank and Haas [4] conclude that human-AI collaborative framing holds promise for brands to “escape” algorithm aversion when generating content, as it keeps consumers' assessment of the authenticity of the message intact. Likewise, AI-generated advertisements have humans working “gatekeepers of humanity,” visibly involved in the creative process, establishing “perceptions of creativity and authenticity” despite the fact that much of the work is accomplished by the AI [18]. This implies that the perceived and extent of human involvement in an advertising pipeline enabled by AI (and not just the presence or absence of AI) is the more relevant variable from a theoretical and managerial perspective.

Although these advances were made, there are three areas in the literature that still need to be addressed. First, most existing research studies only examine two conditions: one being human created and the other being AI created without systematically modelling the combination of the two as a new condition whereby human interaction is visible and has a distinct psychological hallmark [1, 22]. Second, important conditions that facilitate or hinder successful collaboration (disclosure format, product involvement and consumer AI literacy) have been studied separately rather than within a unified framework of moderation [1], [6], [7], [17]. Last, and not least, few papers link the authenticity-and-trust literature to the larger nomological network of relationship marketing theory (commitment–trust theory) [9] or technology-adoption theory (TAM) [13, 16] to narrow the scope of applicability of their findings beyond the advertising context.

This paper fills these gaps by proposing an integrative conceptual model that suggests that the level of HAC has an impact on three parallel and independent psychological processes (perceived authenticity, perceived creativity/competence and AI-induced anxiety/algorithm aversion) that impact consumer trust and ultimately purchase intention. Consumer AI-literacy, product involvement, and disclosure transparency as mediated by AI are modelled. Moderating these mediated relationships are: disclosure transparency, product involvement, and consumer AI-literacy. In doing so the paper has three contributions: (a) it compiles a topic that seemed disjoint and quite fast changing in a single testable research framework; (b) it transcends the "either/and" dichotomy between "human" and "AI" by creating an open concept of human–AI collaboration; and (c) it provides a concrete falsifiable research design that further empirical studies can apply, including studies conducted by the authors and their research group (planned and ongoing) with their additional data collection.

II. Literature Review and Theoretical Background

2.1 Generative AI in Marketing and Advertising

Generative AI is one of the biggest technological changes that marketing has witnessed since the introduction of programmatic advertising [5]. Kshetri et al. [5] outline how various aspects of the marketing function can be leveraged using generative AI, from creating ideas and copy to making personalised creative products, and highlight some open research questions regarding authenticity, bias and disclosure. Huang and Rust [9] made a previous move, developing a strategic framework that divides AI into mechanical, thinking, and feeling types and assert that feeling-intelligence tasks those involving empathy, humour, and emotional nuances, which lie at the heart of persuasive advertising have proven difficult to execute with the authenticity necessary to convince readers or viewers. It also presages many of the current debates about authenticity of advertising, much of which will be a “feeling AI” job, and where AI-generated content is most likely to come across as hollow or generic [3].

2.2 Perceived Authenticity as a Consumer-Evaluative Construct

Authenticity has long been theorized in consumer research as judged (that is, as a construction carried out by consumers rather than a property possessed by the object, or process existed prior to consumer's evaluation) [12, 13] meant by consumers when judging whether an object is authentic and whether an evaluation is sincere. Grayson and Martinec [12] note that there are two ways of being “authentic”: a copy is “iconic” if it is able to resemble something authentic, and “indexical” if it makes a true and real link to its origin. It is important to note this distinction, when applied to advertising produced via AI, because it can become iconic (looks "like a real human ad") but lacks indexical authenticity (can't truly be claimed to be "by a human" on disclosure) and negative reactions from consumers appears to stem from this indexical gap [1], [2], [6]. Empirical studies in recent years have validated that disclosing AI involvement when it wasn't intentional can cause consumers to doubt the authenticity of the message, and this skepticism is heightened for more emotional types of disclosure, like charity or cause advertising [3].

2.3 Algorithm Aversion and Algorithm Appreciation

Another set of theory comes from the field of algorithm aversion, in which users have been observed to favor human judgments over algorithmic judgments, even when it is clear that the algorithm is more accurate, especially when the algorithm makes a mistake [10]. This is apparent in the field of creativity and communication as well: In a widely cited field experiment, Luo et al. [10] observed that customers were far less likely to be interested in buying goods after hearing a chatbot (as opposed to a human) providing service, unless the service of the AI was hidden. But when it comes to tasks where the goal is measurable and quantitative, a competing stream of thought is those conditions where people are demonstrating equal if not heightened trust in algorithmic judgment over human judgment, “algorithm appreciation” [15]. Jin, et al. [21] bring this together by presenting evidence that algorithm aversion and appreciation co-exist and that the former will be triggered by a task structure being "unstructured," or "subjective," and the latter by being "structured," or "data-driven. From this task-conditional point of view, advertising that is creatively well-gotten a matter of mood, where it is not obvious what is required and where emotions cannot be ignored, this follows our suggestion that the complete automation of advertising will evoke a sense of coldness and will not be accepted, that in turn may be offset by the introduction into the perception of “the human warmth” discourse this accompanying invisible participation in collaboration with humans – by tangible human involvement.

2.4 Disclosure, Transparency, and the Persuasion-Knowledge Paradox

One of the most widely researched moderators in this literature is the nature and extent to which the AI involvement is disclosed to consumers, with an actual paradox emerging here. At the other extreme, transparency regarding the use of AI is an ethical requirement, and now a regulatory one too: one such example is the EU's upcoming AI-content labelling guidelines [4]! However, when consumers are given information about a disclosure, it may also elicit their skepticism and create an impression of lack of human involvement, causing consumers to see advertisers as less genuine even if the message itself is not modified [1, 6, 8]. Pfeuffer and Huh [6] demonstrate that distinct message formats for disclosure result in distinct outcomes in terms of trust and attitudes, and Pfeuffer et al. [7] bolster this notion by exposing participants to AI-assisted image manipulation and discovering that its disclosure is mediated by persuasion-knowledge activation, thereby dampening the impact of disclosure on trust. It has also been shown that revealing the source of messages (namely the AI-powered generator) affects how messages are judged, regardless of its quality of message content [8]. Collectively the literature indicates disclosure cannot be viewed just as a dichotomous honesty variable; it is more nuanced, a design feature with measurable downstream trust repercussions, and one that designers of Brands must calibrate, not just consider an afterthought.

2.5 Human–AI Collaboration as a Distinct Condition

The current and most conceptually cutting-edge of this literature addresses human-AI collaboration as being different from either extreme. While generative rewriting has attracted widespread interest, collaborative content creation has been proposed by Haupt et al. [4] as a way of avoiding algorithm aversion where knowledge that a human is "in the loop" significantly reduces the degradation of message credibility judgments when compared to purely automated content. Similarly as with verbal ads, visual and character-based ads with AI-generated components and realistic human characters fare better on the authenticity and creativity dimensions, and are judged differently than all-fake AI character ads, according to research [18]. Comparable studies on advertisement generation in AI vs. human vs. hybrid (semi-AI) confirm that the most problematic effects of AI-only creation are shared with hybrid (semi-AI) ads [22] and hybrid ads are still rated as being somewhere between human and AI creation along the "realness" and effort dimensions [1]. This cross-over of results in the three independent studies is the empirical support for considering human–AI collaboration as a separate third condition in our model, as opposed to some kind of middle ground on one spectrum between humans and AI.

2.6 Trust and Purchase Intention as Downstream Outcomes

Consumer trust serves as a middle variable across the marketing field between advertising evaluation and purchase. Building on the general attitude–behavior–purchase–affection–loyalty framework, Morgan and Hunt's Commitment–Trust Theory (13) suggests that, beyond attitude, trust is the more immediate motivator of long-term effects and a connection to increased purchase and loyalty behavior; that is, trust lessens perceived relational risk and uncertainty, which in AI-mediated situations is exacerbated by the "black box" in algorithmic content production [14]. Building on this, McKnight et al. [14] break down the concept of trust in technology-mediated exchange into three components – competence, benevolence and integrity – which are of direct relevance to assessing the trustworthiness of an AI system (or a brand's AI product). Several studies from the 2024–2026 period also empirically back these claims attitude to AI and major purchase intention constructs of AI-related trust have comparable or larger effect sizes than standard attitude measures [19]–[20], [24]–25]. This enables us to resituate trust as key mediator between advertising production method and purchase intention in our model, not attitude toward the advertising itself.

2.7 Boundary Conditions: Product Involvement and Consumer AI Literacy

Product involvement, defined as feelings of relevance and risk around a product purchase, has proven to be a powerful moderator of the effects of AI disclosure. In line with elaboration-likelihood reasoning (that is, consumers elaborate on high-involvement versus low-involvement products, as well as evaluating source credibility in greater depth), Razak SM and Israfilzade [1] report that the penalties for AI disclosure are significantly larger for products that consumers invest a lot of trouble processing (high-involvement) vs. those that take little effort to process (low-involvement), such as a snack vs. a laptop. The use of consumer AI literacy follows a similar paradigm yet is different. A common misconception was that as people become more familiar with the tools, essays, and suggestions of AI, they would become more comfortable with AI being involved in advertising, however, data indicates people who consider themselves to be more knowledgeable about AI would be more critical of the presence of AI in ads, not less critical [3]. This dual role of literacy is an important modification for an iterative model which is incorporated as a moderator to be hypothesized contextual.

2.8 Summary of Key Studies

Table 1 synthesizes the studies most central to the proposed model, organized by their primary focus and key finding.

Table 1. Summary of Key Studies Underpinning the Conceptual Model

#	Author(s) & Year	Focus	Key Finding
1	Israfilzade (2025) [1]	AI vs. human-created ads; disclosure × involvement	Human-labeled ads receive higher trust/purchase intent; effect stronger for high-involvement products
2	Brüns & Meißner (2024) [2]	Generative AI disclosure in social content	Disclosure of AI use diminishes perceived brand authenticity
3	Haupt, Freidank & Haas (2024) [4]	Human–AI collaboration & algorithm aversion	Collaborative framing preserves message credibility vs. full automation
4	Kshetri et al. (2023) [5]	Generative AI in marketing (review)	Maps applications; flags authenticity/disclosure as open problems
5	Pfeuffer & Huh (2021) [6]	Sponsorship/AI disclosure formats	Disclosure format shapes trust and attitude outcomes
6	Pfeuffer et al. (2025) [7]	AI image-manipulation disclosure	Persuasion knowledge mediates disclosure → trust effect
7	Huang & Rust (2021) [9]	Strategic AI-in-marketing framework	"Feeling AI" tasks (incl. advertising) hardest to automate convincingly
8	Luo et al. (2019) [10]	Chatbot disclosure & purchases	Disclosed AI agents reduce purchase likelihood (algorithm aversion)

9	Jung, Koghut, Lee & Kwon (2025) [11]	AI-generated luxury advertising imagery	Perceived artificial creativity drives trust and purchase intention
10	Grayson & Martinec (2004) [12]	Theory of consumer authenticity	Distinguishes indexical vs. iconic authenticity
11	Morgan & Hunt (1994) [13]	Commitment–Trust Theory	Trust as central relational mediator of exchange outcomes
12	McKnight et al. (2002) [14]	Trust in e-commerce	Trust decomposed into competence, benevolence, integrity
13	Logg, Minson & Moore (2019) [15]	Algorithm appreciation	Consumers can prefer algorithmic over human judgment in some tasks
14	Madathil et al. (2025) [18]	Human–AI collaborative ad characters	Human presence anchors perceived authenticity/creativity
15	Jin et al. (2025) [21]	Algorithm aversion vs. appreciation	Task structure determines which effect dominates

2.9 Recent Developments (2024–2026): Generative AI, ChatGPT, AI Influencers, and AI-Generated Content in Practice

Beyond the foundational sources reviewed above, a fast-growing body of 2024–2026 scholarship and industry evidence speaks directly to four developments central to the present model: the accelerating diffusion of generative AI across the marketing function, the emergence of ChatGPT and other conversational agents as advertising channels in their own right, the rise of AI-driven virtual influencers, and the proliferation of wholly AI-generated brand content. Recent bibliometric and systematic reviews confirm that generative-AI-in-marketing research has grown sharply since 2024, with hundreds of studies now indexed across Scopus and Web of Science and a consistent conclusion that efficiency and personalization gains are accompanied by unresolved questions of authenticity, bias, and disclosure [26]. Subsequent systematic literature reviews covering 2024–2026 publications reinforce this pattern, showing that generative AI is increasingly embedded not just in content production but across the innovation, customer-engagement, and strategic-planning layers of the firm, while cautioning that governance and ethical frameworks have not kept pace with adoption [27], [28].

These dynamics are visible in practice. Nike's Athlete Imagined Revolution (A.I.R.) programme used generative AI models trained on athlete preference data to co-create footwear concepts with athletes ahead of the 2024 Paris Olympics, and the brand has since applied a similar hybrid workflow generating thousands of AI images that are then refined by human directors and VFX teams to campaigns such as its Travis Scott collaboration and subsequent 2025 creative work. Commentators contrasted this visibly human-supervised approach favourably with brands whose AI use appeared unsupervised, suggesting that the degree and visibility of human oversight, not the mere presence of AI, shaped audience reception an observation directly consistent with H1 and H9. Amazon offers a complementary example of generative AI's reach beyond creative production: its Rufus shopping assistant and AI-driven recommendation systems use generative and predictive models to personalize discovery and messaging at scale, illustrating that the marketing applications of generative AI extend well beyond advertising imagery into the broader customer journey envisioned by Kshetri et al. [5].

ChatGPT in Advertising has moved from a research question to a live commercial channel. OpenAI began piloting sponsored placements inside ChatGPT in late 2025 and reported expanding the programme through 2026 while stating that early pilot data showed no measurable decline in its own trust metrics [30]. Independent survey evidence complicates that picture: a January 2026 Ipsos survey found that 63% of US adults said ads appearing in AI search or chat results would make them trust the output less, and a separate Partnercentric survey found 57% of respondents said ads inside AI chatbots would reduce their trust in the advertised brand itself [29]. This gap between platform-reported neutrality and consumer-reported skepticism is a live, real-world instance of the disclosure–trust tension formalized in H8: low-salience, algorithmically-selected ad placements may minimize short-term measured disruption while still carrying a latent trust cost that only shows up in direct consumer self-report.

AI Influencers extend the same disclosure and authenticity dynamics to persona-based marketing. Ujjainwala [31] reports that AI-generated influencer content significantly reduces perceived authenticity and brand trust relative to content from human influencers, that explicit disclosure of AI generation intensifies this effect, and that the penalty is moderated by consumers' digital literacy a close empirical parallel to H8 and H10. A systematic review of 60 studies on virtual influencers similarly identifies trust and disclosure transparency as the two variables most consistently linked to whether a virtual influencer strengthens or undermines brand relationships, while noting that human-like, interactive personas fare better than overtly synthetic ones [32]. Sephora illustrates a lower-risk variant of this strategy: rather than deploying a fully synthetic spokesperson, the brand layers AI tools its Virtual Artist augmented-reality try-on feature and Smart Skin Scan diagnostic underneath disclosed human beauty advisors and customer reviews, keeping the AI in a visibly assistive role for a high-involvement product category where H9 predicts disclosure penalties are largest.

AI-Generated Content, when produced with minimal visible human framing, illustrates the downside risk the model anticipates. Coca-Cola's AI-generated Christmas advertisements, released in both 2024 and 2025 with only limited public evidence of human creative direction, were widely criticized as "soulless" and lacking the emotional warmth of the brand's earlier campaigns, with independent sentiment tracking showing a measurable decline in positive audience reaction after each launch. Because Coca-Cola's holiday advertising is unusually nostalgic and emotionally loaded, the backlash is consistent with H1 and H5: fully automated production is predicted to be most damaging precisely in unstructured, emotionally significant categories, echoing earlier findings that undisclosed or unmediated generative-AI use erodes perceived brand authenticity [2] and that hybrid human–AI production is rated closer to human-only work on authenticity and effort than to AI-only work [20], [22]. Table 3 summarizes how these four brand cases map onto the collaboration-level and trust constructs proposed in Section 3.

Table 2. Illustrative 2024–2026 Brand Cases Mapped to the Proposed Model

Brand	AI Application	Human–AI Collaboration Level	Trust-Relevant Outcome / Model Link
Nike	Generative co-design with athletes (A.I.R.); AI-assisted imagery refined by human directors/VFX	High – visible human creative direction and oversight	Favourable reception; consistent with H1/H9 (visible collaboration preserves authenticity)
Amazon	Generative/predictive AI in the Rufus shopping assistant and recommendation engine	Moderate – AI-driven personalization disclosed as a shopping aid	Supports H4 by pairing AI competence with a transparent, assistive framing
Sephora	Virtual Artist AR try-on and Smart Skin Scan layered under human advisors and reviews	Moderate-High – AI positioned as an aid, not a replacement, in a high-involvement category	Consistent with H9; disclosed, assistive AI mitigates disclosure penalties
Coca-Cola	Largely AI-generated holiday advertisements (2024, 2025) with limited visible human framing	Low – minimal disclosed human creative involvement	Backlash and declining sentiment; consistent with H1/H5 (algorithm aversion in emotionally-loaded categories)

III. Conceptual Framework and Hypotheses Development

3.1 Model Overview

Based on the literature reviewed, we suggest that the independent variable of interest is the level of Human-AI collaboration in advertising production (production without human, production with some human-AI collaboration, and production entirely by AI). It "works" via three mediating pathways: perceived authenticity, perceived creativity and competence, and AI-induced anxiety/algorithm aversion, and these predicted consumer trust, which subsequently predicted purchase intention. The following three variables are considered as mediators in the paths: AI-disclosure transparency, product involvement, and consumer knowledge about AI. The whole model is shown in figure 1.

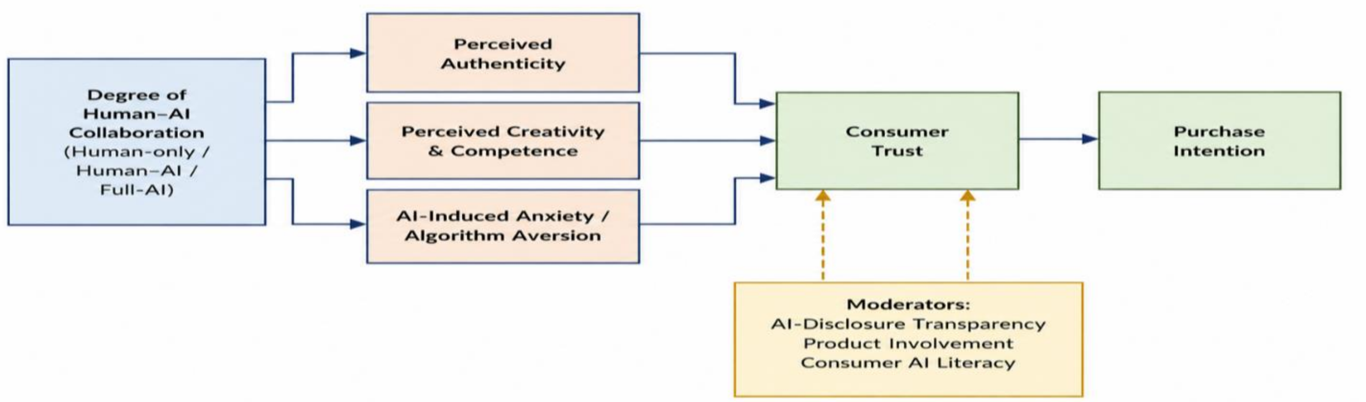

Fig.1. Proposed Conceptual Model of Human–AI Collaboration in Digital Advertising

3.2 Hypotheses

H1. The degree of human–AI collaboration has a curvilinear relationship with perceived authenticity, such that human–AI collaborative advertising is rated as more authentic than fully AI-generated advertising, and not significantly different from (or only modestly lower than) fully human-created advertising [1], [4], [18], [22].

H2. Perceived authenticity is positively associated with consumer trust [1], [2], [12].

H3. The degree of human–AI collaboration is positively associated with perceived creativity and competence, with human–AI collaborative advertising rated as most creative, reflecting the combination of human intentionality and AI-enabled novelty [9], [11].

H4. Perceived creativity and competence is positively associated with consumer trust [9], [11].

H5. Fully AI-generated advertising elicits significantly higher AI-induced anxiety/algorithm aversion than human-only or human–AI collaborative advertising [10], [15], [21].

H6. AI-induced anxiety/algorithm aversion is negatively associated with consumer trust [10], [21].

H7. Consumer trust is positively associated with purchase intention [13], [19], [20], [24], [25].

H8. AI-disclosure transparency moderates the authenticity–trust relationship, such that the positive effect of perceived authenticity on trust is weaker when AI involvement is explicitly and prominently disclosed than when it is disclosed more subtly or contextually [1], [6], [7].

H9. Product involvement moderates the strength of the human–AI-collaboration effects on trust, such that the authenticity/creativity benefits of visible human involvement are stronger for high-involvement product categories than for low-involvement categories [1].

H10. Consumer AI literacy moderates the disclosure–trust relationship such that higher-literacy consumers are more sensitive (not less) to AI involvement, holding brands to stricter transparency expectations, thereby attenuating the benefits of disclosure on trust for this segment [3].

IV. Proposed Methodology

Because the model integrates constructs validated in prior work but has not, to our knowledge, been tested as an integrated whole, we propose a two-phase mixed-methods design suitable for empirical validation in a subsequent study.

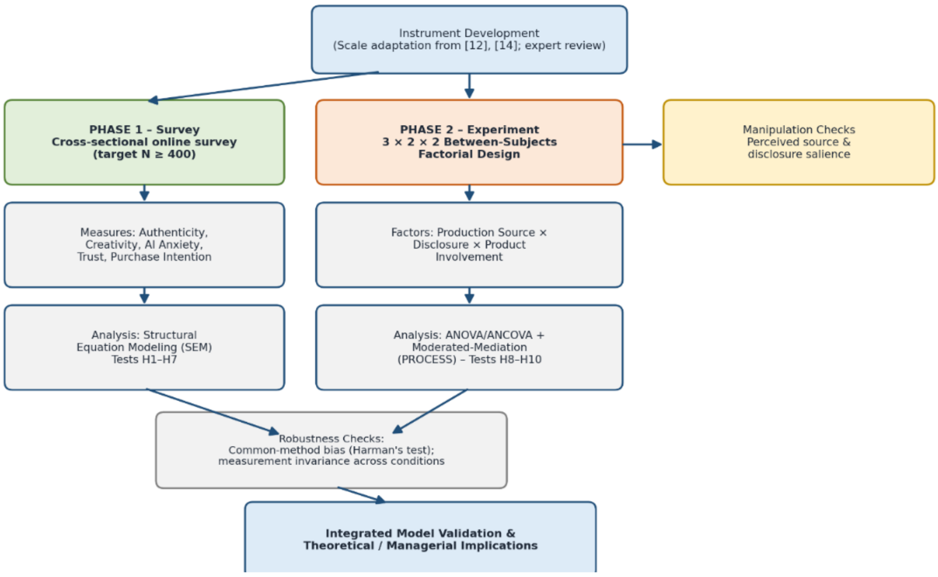

Fig.2. Proposed Two-Phase Research Design and Analytical Workflow

4.1 Phase 1: Cross-Sectional Survey

The quantitative survey, an adult consumer survey (N ≥ 400 persons powered for structural equation modeling for the number of latent constructs hypothesized in the study), would be conducted online and would include a demographically diverse sample with various age, gender, and AI familiarity categories. Existing scales for each domain validated were adapted: authenticity items based on Grayson and Martinec's use of the indexical/iconic paradigm [12], trust items modified from a scale developed by McKnight et al. [14] on the competence–benevolence–integrity typology, and purchase intention items adapted from standard multi-item marketing scales. Proposed constructs, sources, and illustrative measurement items for the set of questions created for the survey instrument are presented in Table 2.

Table 3. Proposed Constructs and Measurement Sources

Construct	Type	Illustrative Item	Adapted From
Perceived Authenticity	Mediator	“This advertisement feels genuine and sincere.”	[12]
Perceived Creativity & Competence	Mediator	“This advertisement shows original, skillful creative work.”	[9], [11]
AI-Induced Anxiety	Mediator	“Knowing AI was involved in this ad makes me uneasy.”	[10], [21]
Consumer Trust	DV1 / Mediator	“I believe this brand is honest about how it creates content.”	[14]
Purchase Intention	DV2	“I would consider buying this product after seeing this ad.”	Standard PI scale
AI-Disclosure Transparency	Moderator (manipulated)	Prominent label vs. subtle footnote vs. no disclosure	[1], [6]
Product Involvement	Moderator (manip./measured)	High-involvement (electronics) vs. low-involvement (snacks)	[1]
Consumer AI Literacy	Moderator (measured)	“I understand how generative AI tools create content.”	Adapted general-literacy scales

4.2 Phase 2: Between-Subjects Experiment

To demonstrate efficacy, a 3 (human-only, collaborative, AI-generated) × 2 (prominent, subtle) × 2 (high or low product engagement) between subject's factorial design is suggested following the design approach in similar recent studies [1], [18], [22]. Subjects would be shown only one of each of the matched ads under conditions, while the variable which could be product content, picture quality, or message content would differ. Before interpreting the outcomes on the downstream measures, manipulation checks would confirm perceived production source and disclosure salience would be used.

4.3 Analytical Strategy

The survey results would be analyzed by using structural equation modeling (SEM) with a view to testing the full mediation model (H1–H7) and ANOVA/ANCOVA would be applied to examine the experimental hypotheses (H8–H10), following the practice illustrated in the cited literature [1], [4], [17], [25]. The Harman's single factor test and marker-variable approach would be used to assess for common-method bias, and measurement invariance for the three conditions of production would be verified prior to comparing the paths of the structures.

V. Discussion

The suggested model changes the discussion that has frequently been articulated as an either/or proposition "AI or humans in ads"? into the more actionable question as to "how much." and "how visible." should be humans in ads and how should the collaboration be communicated? This paradigm shift aligns with the evolving direction of the most recent body of literature where the participation of the human and the artificial become more often than not not a middle ground, but rather a different strategic approach [4], [18]. This rephrasing has a number of implications.

First, from an ethical standpoint, and from a commercial one, too, all of this will not do in the unstructured and emotional categories: when an ad is fully automated there are no elements that consumers rely on for authenticity; they know what to expect, and thus they are commercially at risk as well. Secondly, the era of complete transparency of AI use is becoming ever more unfeasible due to increased regulations on disclosure [4] and heightened consumer awareness or knowledge of AI [3], as the model predicts. Third, and most novel, the model derives an interior optimum: while visible, well-communicated collaboration between human and AI humans may perform best with neither of the two extremes surprisingly often closer to the human-only end on both efficiency and trust, as reflected in quite recent empirical results where interaction conditions fall into the middle zone between human-only and human–AI extremes on the joint criteria of authenticity and effort perception scales [1, 22].

The possible moderators make a great deal to this picture. The involvement moderator indicates that full automation is likely to be the most dangerous solution for people with high-involvement/high-risk products, but that it can be a safer solution among people with low-involvement products when there is transparency around what it entails. This means that the full-automation approach may not be suitable for brands selling products with high involvement and a high risk of harm, such as financial products, healthcare, durable products; whereas it is more appropriate for brands in low-involvement products segments because there is for them more leeway to opt for full automation while not triggering dramatic trust penalties. The AI-literacy moderator (H10) notes that a key managerial bias is that over time AI-knowledgeable consumers will grow more acceptant of AI made ads. The literature discussed here indicates rather that only the level of literacy increases the expectations for transparency, but not resistance to AI itself [3].

VI. Theoretical and Managerial Implications

6.1 Theoretical Contributions

This paper makes three contributions to the theory. The first being, the extension of two theories: the Commitment–Trust Theory (13) and trust-typology researches (14) in the context of AI-mediated production of an ad, where the production process itself of an ad is treated as a trust-relevant cue, and it is not only its message content. Second, it reflects the ongoing literature on authenticity [12] by distinguishing the notion of human–AI collaboration as a new distinct category, instead of using "AI-generated" as one generic category. Third, introducing the previously disjointed algorithm-aversion/appreciation literature [10] [15] [21] into the advertising-authenticity literature, it suggests that the advertising domain has a task structure that is very different from other tasks in which an appreciation effect has been documented, and that results in such positive effects being very rare and thus overshadowed.

6.2 Managerial Implications

The model provides for practitioners a set of practical principles. Great brands should think of generative AI as a production assistant instead of a production team. especially when it comes to high involvement products and sensitive topics in campaigns [1], [3], [4]. For disclosure where necessary and/or desired, disclosure language and placement should be carefully designed, and label descriptors that label the content as "AI-generated" can lead to an inappropriately high level of persuasion-knowledge doubt about the content relative to more detailed, contextually-rich disclosure of the human involvement in the process [6, 7, 8]. Finally, there are visible cues that indicate human oversight roles, such as creative director, backstage supporting roles, or brand storytelling to highlight the human-AI collaboration effort, that marketing organizations should invest in. These cues are surrogates for preserving the cues of authenticity that are eliminated in fully automated pipelines [4] [18]. Finally, the brand impact seems to be a moderating effect on the impact of AI-disclosure penalties, indicating that established brands might have a bit more breathing room to try out creative advertising strategies with AI than lesser-known or less trusted brands [17].

6.3 Human–AI Collaboration as a Trust-Building Strategy

The cases examined in Section 2.9 share a practical solution to the aforementioned starting question of this paper: brands can boost consumer trust by creating the relationship between human and AI production and communicating it honestly rather than choosing sides between one or the other. The model and the brand evidence indicate the existence of four managerial levers which directly follow the model.

First, do not place only an approved human role at the point of final decision, but a named one that is visible at the point of judgement. While the model predicts convention (H1, H3) keeps athletes and human creative directors visible to the final outcome (H1, H3), it also saves on time, facilitates more lapses to produce ads - and improves the quality of the results. Brands should be wary of relying on generative AI as a type of "production team" for a high-involvement and emotionally-charged campaign, much like a paper-based expert influence would be subject to (as described above in Section 6.2), and instead think of it as an expert "assistant" that a named human, materially and meaningfully, tweaks and shapes its output.

Second, change disclosure, don't rely on the silence or "AI-generated" label. Measured trust disruption drops can be achieved using low salience placements, and no drop in trust could happen without activation of persuasion-knowledge and/or skepticism (H8) weighed by the evidence in the ChatGPT-advertising domain in Section 2.9. A trustful outcome is gained from the detailed and contextually rich disclosure, describing what the AI did, what a human did (H&M and Sephora are working on something similar with an AI-assisted image or diagnostic), compared to a generic AI label.

Thirdly, correspond the involvement of the products with collaboration. Brands that sell to higher involvement segments will be heralded to having human oversight on their brand website, while brands with lower feeling of involvement can trust in lightly-supervised automation like Amazon does in low-stakes, clearly-publicized personalization's.

Fourth, don't assume AI consumers will be lax, they won't, they will be more critical. As stated in H10 and influencer literature above [31, 32] brands with AI-knowledgeable people must be more transparent, and, not more tolerant over time with regard to undocumented use of AI. Instead, companies should continue to invest in longer-lasting disclosure practices (and think about whether it's a good idea to as is sometimes suggested assume that increasing familiarity with AI will diminish the need for disclosure in the future), rather than relying on disclosure that will only be used when it is convenient. Together, these four levers illustrate how an open approach to human-AI collaboration, as they all point to the same thing, is the strategic lever soon to become brands' most powerful weapon to turn the efficiency gained with generative AI into consumer trust, instead of losing it.

VII. Limitations and Future Research

This conceptual paper is limited in that there is not yet a body of work or evidence collected that this has been tested as an integrated whole, the hypotheses presented are therefore based on converging evidence similar to many independent studies, but in the end, are Phase 1 and Phase 2 of the proposed methodology, that this work will be carried out in the future? Further empirical research is needed with a wider scope of product categories, cultural contexts, and advertising formats (static image or text-based advertising, video or audio formats, and interactive/conversational formats like AI chatbots), as most of the studies conducted so far are focused on these latter formats. Longitudinal designs are also required as exposure to an AI technique once is not reflective of the effect it may have, as consumers will become familiar with the brand's use through repeated interactions. More generally, cross-cultural replication is justified: many of the mechanisms outlined here (especially algorithm aversion versus appreciation) are hypothesized to be culturally dependent and most studies to date are based on Western or East Asian samples, with few studies on other regions [21, 25].

VIII. Conclusion

Our main managerial and theoretic challenge is transitioning from "should we use AI" to "how can we integrate AI with human effort wisely to meet consumer trust demands? The primary managerial and theoretical questions are not simply transitioning from "should we use AI" to "how can we combine human and artificial intelligence wisely to meet consumer trust demands?" This paper brings together the existing, yet fragmented, literature into an integrative conceptual model to capture the relationship between perceived authenticity, perceived creativity, and AI-induced anxiety due to degree of human–AI collaboration, which subsequently affects consumer trust and purchase intention, as the degree of disclosure transparency, product involvement and consumer AI literacy act as moderators. Give human/AI collaboration, not complete automation or complete manual production, the strategic optimal point for many advertising situations provided that it is approached with a humanistic and articulate manner, not hidden or too simply presented as AI. Empirical testing of this model as described in the proposed method could be an opportune and significant future marketing research path to follow.

References

[1] 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

[2] Brüns, 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

[3] Universidade Católica Portuguesa (2025). The impact of AI-generated advertisement on consumer trust and emotional brand connection. Institutional repository working paper.

[4] Haupt, M., Freidank, J., & Haas, A. (2024). Consumer responses to human–AI collaboration at organizational frontlines: Strategies to escape algorithm aversion in content creation. Review of Managerial Science, 19, 377–413. https://doi.org/10.1007/s11846-024-00748-y

[5] Kshetri, N., Dwivedi, Y. K., Davenport, T. H., & Panteli, N. (2023). Generative artificial intelligence in marketing: Applications, opportunities, challenges, and research agenda. International Journal of Information Management, 75, 102716.

[6] Pfeuffer, A., & Huh, J. (2021). Effects of different sponsorship disclosure message types on consumers' trust and attitudes. International Journal of Advertising, 40(1), 49–80. https://doi.org/10.1080/02650487.2020.1807229

[7] Pfeuffer, A., Hatfield, H. R., Evans, N., & Kim, J. (2025). Illegally beautiful? The role of trust and persuasion knowledge in online image manipulation disclosure effects. International Journal of Advertising, 44(4), 696–717.

[8] Lim, S., & Schmälzle, R. (2024). The effect of source disclosure on evaluation of AI-generated messages: A two-part study. Computers in Human Behavior: Artificial Humans, 2(1), 100058.

[9] Huang, M.-H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. Journal of the Academy of Marketing Science, 49, 30–50. https://doi.org/10.1007/s11747-020-00749-9

[10] Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. Marketing Science, 38(6), 937–947. https://doi.org/10.1287/mksc.2019.1192

[11] Jung, T., Koghut, M., Lee, E., & Kwon, O. (2025). The impact of AI-generated marketing imagery on consumer trust and purchase intentions: Examining the effect of human-AI assisted images on marketing in luxury advertising. Journal of Retailing and Consumer Services, 87.

[12] Grayson, K., & Martinec, R. (2004). Consumer perceptions of iconicity and authenticity in retail store settings. Journal of Consumer Research, 31(2), 296–312. https://doi.org/10.1086/422109

[13] Morgan, R. M., & Hunt, S. D. (1994). The commitment-trust theory of relationship marketing. Journal of Marketing, 58(3), 20–38. https://doi.org/10.2307/1252308

[14] McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. Information Systems Research, 13(3), 334–359. https://doi.org/10.1287/isre.13.3.334.81

[15] Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. Organizational Behavior and Human Decision Processes, 151, 90–103. https://doi.org/10.1016/j.obhdp.2018.12.005

[16] Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340.

[17] Illescas-Manzano, M., Martínez-Puertas, S., Cardoso, P. R., & Segovia-López, C. (2025). Use of online shop chatbots: How trust in seller moderates brand preference and purchase intention. In Disruptions, Diversity, and Ethics in Marketing (ICoAMP 2024). Springer Proceedings in Business and Economics. https://doi.org/10.1007/978-3-031-70488-8_10

[18] Madathil, J. C., et al. (2025). Generative AI advertisements and human–AI collaboration: The role of humans as gatekeepers of humanity. Journal of Retailing and Consumer Services.

[19] Kim, S., et al. (2026). How AI involvement affects curiosity and content engagement. Journal of Business Research.

[20] Lee, G., et al. (2024). Human vs. AI in fashion design. Journal of Retailing and Consumer Services. https://doi.org/10.1016/j.jretconser.2023.103512

[21] Jin, Y., et al. (2025). The influence of algorithm aversion and algorithm appreciation among consumers in unstructured tasks. International Journal of Consumer Studies.

[22] Kučinskas, N. G. (2025). Revealing AI involvement in ad creation: Effects on authenticity, brand perceptions and consumer intentions. Working paper / preprint.

[23] Deloitte Digital. (2025). Human-AI marketing collaboration: GenAI in marketing operations research report. Deloitte Digital (based on Lawless Research survey of 1,000 U.S. consumers).

[24] Nuremberg Institute for Market Decisions (NIM). Buder, F., et al. (2024). Marketers' generative AI adoption survey. NIM Publications.

[25] EMARKETER. (2025). Many US adults avoid, distrust brands that use AI ads. Industry report, based on Attest survey data.

[26] Prasanna, B. (2025). Generative AI in marketing: Foundations, trends, and future research propositions. Human Behavior and Emerging Technologies, 2025, 5542513. https://doi.org/10.1155/hbe2/5542513

[27] El Oualfi, A., & El Hamzaoui, M. (2026). Reimagining marketing through generative AI: A systematic literature review (SLR). In N. Aboutabit & I. Hafidi (Eds.), Advances in Machine Intelligence and Computer Science Applications (ICMICSA 2025), Lecture Notes in Networks and Systems (Vol. 1882). Springer. https://doi.org/10.1007/978-3-032-19963-8_28

[28] Mridha, M. F., Hossen, M. J., et al. (2026). A systematic literature review on the impact of generative AI in digital marketing: Advancements, opportunities, and challenges. Computers, Materials & Continua, 87(3). https://doi.org/10.32604/cmc.2026.071029

[29] EMARKETER. (2026). FAQ on ChatGPT advertising: Formats, costs, and early strategies to win. Industry report.

[30] OpenAI. (2026). Testing ads in ChatGPT. Company product update.

[31] Ujjainwala, F. J. (2025). Influence of AI-generated influencer content on brand trust and authenticity perceptions. Journal of Marketing & Social Research, 2(9), 256–262.

[32] Hewapathirana, I. U., & Perera, N. (2024). Navigating the age of AI influence: A systematic literature review of trust, engagement, efficacy and ethical concerns of virtual influencers in social media. Journal of Infrastructure, Policy and Development, 8(8), 6352. https://doi.org/10.24294/jipd.v8i8.6352