

Article

AI-Generated vs. Human-Created Sustainable Advertising: Effects of Source Disclosure and Environmental Claim Strength on Perceived Greenwashing and Green Trust Among Generation Z Consumers

Khalil Israfilzade 

Communication Department, Faculty of Business, Kauno Kolegija Higher Education Institution, 50468 Kaunas, Lithuania; khalil.israfilzade@go.kauko.lt

Abstract

The rapid integration of generative artificial intelligence into digital advertising, combined with growing consumer concern about greenwashing, has created a dual credibility burden in which sustainability messages must be evaluated for both environmental truthfulness and authenticity of authorship—a challenge intensified by emerging AI disclosure regulations such as Article 52 of the EU AI Act. This study investigates how ad source (AI-generated vs. human-created), source label (labelled as AI vs. labelled as human), and environmental claim strength (vague vs. strong) jointly influence perceived greenwashing and green trust among Generation Z consumers. A $2 \times 2 \times 2$ mixed factorial experiment was conducted with 154 undergraduate participants randomly assigned to one of four source–label conditions and exposed to both vague and strong environmental claims; data were analysed through one-way ANOVAs with Tukey HSD post hoc tests and paired-samples *t*-tests. All eight hypotheses were supported: AI attribution consistently elevated greenwashing perceptions and depressed green trust, while strong claims significantly reduced greenwashing perceptions and elevated trust within both correctly disclosed conditions. Most notably, the disclosed label—rather than the actual generative source—emerged as the dominant psychological cue, with human content mislabelled as AI suffering the same credibility penalty as genuine AI content. The findings reveal a transparency paradox with significant implications for AI disclosure regulation and sustainable marketing practice.

Keywords: sustainable digital marketing; artificial intelligence; generative AI; greenwashing; green trust; environmental claims; green consumer; green marketing; Generation Z



Academic Editor: Jiang Wu

Received: 6 May 2026

Revised: 21 June 2026

Accepted: 3 July 2026

Published: 16 July 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The rapid integration of generative artificial intelligence (AI) into digital marketing has fundamentally transformed how advertising is produced and personalised, with AI now capable of generating ad copy, visual assets, and entire campaign narratives at a scale and speed that would have been inconceivable a decade ago [1–4]. At the same time, sustainability has become a central concern in consumer markets, particularly among younger generations, with environmental claims now appearing in a substantial share of advertising messages across virtually every product category. These two developments—the rise of generative AI and the proliferation of sustainability communication—have intersected in ways that are only beginning to be understood empirically. When a brand

uses generative AI to produce a sustainability advertisement, the resulting message carries a dual credibility burden: consumers must evaluate both the truth of the environmental claim and the authenticity of the medium through which it is delivered. The question of how consumers respond to this dual burden—and whether mandated disclosure of AI involvement helps or hinders their evaluation—has emerged as a pressing concern for marketing scholars, regulators, and practitioners alike.

The existing literature offers strong but partially conflicting guidance on how consumers process such messages. On one hand, research on greenwashing consistently demonstrates that vague, symbolic environmental claims (e.g., “eco-friendly,” “natural”) elevate consumer scepticism and erode brand trust, while substantive, evidence-based claims that provide quantified, verifiable environmental benchmarks reduce scepticism and build green trust through credibility signalling [5–9]. On the other hand, an emerging body of research on AI-generated content documents a robust “algorithm discount”—a tendency for consumers to systematically distrust outputs attributed to AI systems, even when the content is functionally indistinguishable from human-produced equivalents [10–12]. These two studies suggest competing predictions when combined: substantive environmental claims should build trust regardless of source, yet AI attribution should depress trust regardless of claim strength. Adding a further layer of complexity, regulatory frameworks such as Article 52 of the EU AI Act [13] are increasingly mandating explicit disclosure of AI-generated content on the assumption that transparency protects consumer interests—yet empirical evidence indicates that AI labelling can paradoxically reduce credibility rather than enhance it, a phenomenon described in the literature as the *transparency paradox* [14,15]. Whether substantive claims can overcome the AI labelling penalty, whether the actual source or the disclosed label is the operative psychological cue, and whether mandated transparency genuinely serves consumer welfare in the sustainability context remain open empirical questions.

The main aim of this study is to investigate how the *source of a sustainable advertisement* (AI-generated vs. human-created), the *disclosed source label* (labelled as AI vs. labelled as human), and the strength of its *environmental claim* (vague vs. strong) jointly influence *perceived greenwashing* and *green trust* among Generation Z consumers. Specifically, the study pursues three interrelated objectives: *first*, to determine whether perceived greenwashing and green trust differ significantly across the four configurations of ad source and source label under both vague- and strong-claim conditions; *second*, to test whether environmental claim strength independently shapes consumer perceptions when source disclosure is accurate; and *third*, to examine whether the actual generative source or the disclosed label is the operative psychological cue driving consumer responses, with direct implications for the design of AI transparency policy.

This study addresses these questions through a 2 (*Ad Source*: AI-generated vs. human-created) $\times$ 2 (*Source Label*: labelled as AI vs. labelled as human) $\times$ 2 (*Claim Strength*: vague vs. strong) mixed factorial experiment conducted with 154 Generation Z consumers. Generation Z was selected as the target population because they represent the first cohort to grow up entirely within an algorithmic media ecosystem and are simultaneously the most environmentally attuned consumer segment, making them the most theoretically informative audience for examining responses to AI-mediated sustainability advertising. The study tests eight hypotheses examining (i) whether perceived greenwashing and green trust differ across the four source–label configurations under each level of claim strength and (ii) whether environmental claim strength independently shapes consumer perceptions when source disclosure is accurate.

The principal findings of the study are that all eight hypotheses receive empirical support: AI attribution—whether truthful or merely labelled—consistently elevates perceived

greenwashing and depresses green trust, and within both correctly disclosed conditions, vague claims produce significantly higher greenwashing and lower trust than strong claims. The most distinctive finding is that the disclosed *label*, rather than the actual generative source, is the dominant cue shaping consumer perceptions: human-created advertisements mislabelled as AI suffer the same credibility penalty as genuinely AI-generated content, while AI-generated advertisements concealed behind a human label perform comparably to the human control. This pattern reveals a transparency paradox with significant implications for emerging *AI disclosure regulation*, for the design of corporate sustainability communication, and for the broader theoretical literature on consumer responses to AI-mediated persuasion.

2. Literature Review and Hypothesis Development

2.1. AI-Generated Content in Marketing

The pace at which generative AI has reshaped marketing practice is difficult to overstate. What once required dedicated creative teams and extended production timelines can now be prototyped, iterated, and deployed at a fraction of the cost and time—a shift that has touched everything from digital advertising to personalised customer journeys across virtually every industry sector. Across all industries, businesses have reported significant improvements in terms of efficiency and scale. Efficiency gains have been real and well-documented. Yet these same gains have raised uncomfortable questions about authenticity, accountability, and the ethics of automated persuasion that the field is still working through.

Generative AI allows the automation of ad copy, visuals, and videos, leading to better conversion rates and campaign performance while significantly lowering human effort [3,4,16,17]. It enables hyper-personalisation at scale, tailoring messages to individual profiles and matching or exceeding the measured effectiveness—such as engagement and conversion—of human-written advertisements [18,19]. Additionally, it reduces production costs [20]. It also allows for SEO optimisation, but it requires human involvement to avoid generic, “sterile” content and errors. Undoubtedly, it has become such an essential part of the modern marketing field [2,21,22] that generative AI applications are becoming more and more integrated into digital channels such as social media, email, search, and chatbots [16].

Nonetheless, significant challenges have been reported. Reed and colleagues [23] highlight that bias plays a crucial role in the discussions about transparency and authenticity. Borimnejad & Borimnejad [24] suggest that although sentiment analysis predominantly reflects positive academic attitudes regarding the efficiency gains of AI, the same body of academic work registers increasing concern about ethical risks, including algorithmic bias, privacy issues, and the erosion of trust.

2.1.1. Role of Generative AI in Advertising

Generative AI redefined advertising by using big language models and text-to-image diffusion frameworks; marketing specialists can make scripts, storyboards, and images that are in line with brands and often have the same effect as human-written content [1,3,16,25–27]. According to empirical studies, this AI-based personalisation has been found to significantly increase consumer engagement and click-through rates compared to generic advertising approaches [18,25,28]. These findings, however, concern the persuasive performance of AI-assisted personalisation rather than the effect of disclosure: in most cases AI involvement was not revealed to consumers, and where it was [18], the outcome of interest remained persuasive impact rather than the greenwashing and trust perceptions examined here. How disclosing AI authorship shapes consumer evaluation thus remains an open question, which the present study addresses.

While AI-driven campaigns are frequently regarded as effective and productive, their success is heavily influenced by perceived creativity and the quality of human–AI collaboration. AI-generated advertisements can be as coherent and persuasive as human-authored ads [29,30]. However, the strategic success of these tools is limited; it remains difficult to generate truly novel ideas without human strategic input, and poor AI outputs can receive strongly negative consumer reactions when quality expectations are not met [31]. As a result, marketers increasingly see GenAI as a dependable tool for ensuring consistency across narratives and storyboards, rather than a replacement for human creativity.

2.1.2. Consumer Perceptions of AI vs. Human Content

Consumer ratings of AI-generated versus human-created content reveal an ongoing preference for human creativity, particularly concerning perceived credibility and trust. Empirical research consistently shows an “*algorithm discount*,” in which labelling content as AI-generated reduces perceived accuracy, sharing intentions, and overall credibility [10,11,32]. This effect is especially noticeable in high-involvement contexts like advertising and news, where consumers associate AI with decreased transparency and trust [12]. Notably, this dynamic does not always reflect a straightforward distrust of machines. When content quality is controlled for, the deficit attributed to AI authorship often looks less like rejection and more like a baseline *favouritism* toward human effort—a preference that persists even when the outputs are functionally indistinguishable [12].

AI as an author has a significant impact on perceived authenticity and emotional resonance. These are key drivers of consumer engagement [33]. Knowing that AI had a hand in creating something can quietly deflate how genuine a brand feels, and that matters—because authenticity is one of the stronger predictors of whether consumers will recommend a brand, stick with it, or feel anything positive toward it at all [34–36]. Yet some experimental studies have shown that AI-generated copy may cause emotional responses similar to those elicited by human-created content [37]. Nevertheless, such findings remain limited in their generalisability and should be interpreted with appropriate caution.

This does not, however, indicate a dead end for AI in content production. The more instructive finding from the literature is that consumer responses are highly malleable depending on how AI involvement is framed and disclosed [35,38]. *Trust recovers*—sometimes substantially—when audiences understand that a human was guiding the process. Organisations use control mechanisms to maintain authenticity and optimise algorithmic efficiency by framing AI as a collaborative tool rather than an autonomous creator [35,39]. Transparency about how human strategists and AI generators work helps consumers, minimising algorithm aversion without affecting engagement [38,40].

2.2. Greenwashing and Sustainable Advertising

Greenwashing—the practice of making false, exaggerated, or unsupported environmental claims—has become a key issue in sustainable and digital advertising. With brands highlighting more and more environmental benefits, consumers are finding it difficult to judge whether they are honest [41–44]. Empirical evidence from multiple systematic reviews confirms that greenwashing reduces consumer trust and increases scepticism. Persakis et al. [7] examined 419 peer-reviewed articles published between 2004 and 2024 and concluded that greenwashing “*pervasively affects marketing ethics and consumer scepticism*.” These findings are supported by Huang et al. [44], who examined 180 academic articles and observed unclear definitions and measurements of greenwashing while consistently documenting its negative impact on consumer perceptions. The construct consists of several dimensions: false or misleading claims, lack of evidence or unverifiability, vagueness and uncertainty, and irrelevance or inflated claims of minor actions [5,41,45]. These issues

get worse online because ESG-labelled and social media ads use persuasion, influencers, and algorithms together. This makes it harder to spot false sustainability claims as they become more widespread [43,46,47].

The distinction between *symbolic (vague)* and *substantive (strong)* environmental claims is critical for understanding how greenwashing works and why certain claims lead to consumer scepticism. Symbolic claims are defined by vague, unsubstantiated language (e.g., “eco-friendly”, “all natural”) that gives an impression of environmental responsibility without providing real proof [5,7,48,49]. Persakis et al. [7] define this as “green embellishment,” which uses “vague or unsubstantiated terms like ‘natural’ or ‘eco-friendly’ to create a halo effect, where consumers perceive the company as comprehensively sustainable regardless of actual impact.”

The gap between these two types is not merely rhetorical; it is epistemological. What makes symbolic claims particularly stubborn is that they are designed, whether intentionally or not, to exploit a genuine limitation in how most people process environmental information. It is possible to emphasise that *information asymmetry* is the primary reason for this difficulty [50]. Consumers generally lack both the technical knowledge and cognitive capacity to analyse complex environmental claims, leaving vague language effectively unchecked. However, for knowledgeable consumers, vague claims can fail by increasing green scepticism and decreasing willingness to pay, whereas specific evidence-based claims reduce scepticism [42,43,51].

The effects of greenwashing on consumer psychology and behaviour are widely reported. Prior cases of greenwashing are a major source of *green scepticism*, making consumers sceptical of future sustainability claims [41,52,53]. Perceived greenwashing undermines brand trust, credibility, and corporate reputation, potentially surpassing any short-term image gains from “*looking green*” [54–56]. This scepticism is especially evident among Generation Z consumers, with 50% expressing that businesses and brands are greenwashing their efforts to be more environmentally responsible [57]. Repeated exposure to misleading claims leads to “*green fatigue*” and “*green cynicism*” [58]. This affects entire industries and reduces the effectiveness of genuine green marketing efforts. Behavioural outcomes include decreased purchase intention, reduced willingness to pay, and increased boycott intentions, particularly in high-impact industries like airlines [42,51,59].

Greenwashing’s impact is increased by the unique challenges that digital marketing environments present. According to research, over 70% of green claims on social media are misleading in some samples [47]. This suggests that digital channels not only transmit greenwashing more efficiently but also actively change its character, making it more natural, difficult to pin down, and thus harder to regulate.

2.3. Green Trust in Sustainable Marketing

Green trust, defined as consumers’ confidence in a brand’s environmental performance, plays a key role in sustainable marketing because it mediates the relationship between sustainability claims and consumer purchasing behaviour [8,60]. Because environmental attributes are largely unobservable and difficult for consumers to independently verify, green trust serves as the critical psychological bridge that connects green marketing signals to behavioural outcomes such as purchase intention, loyalty, and positive word-of-mouth [61,62]. We can see it in a variety of product and service contexts, including food, tourism, fast-moving consumer goods (FMCG), and digital commerce [60–64].

One of the most fundamental conditions that contribute to the importance of trust in this field is the reality of *information asymmetry*. There is a need for consumers to rely on stated claims because they are unable to directly observe or verify a company’s environmental performance. As a result, trust is an essential cognitive and affective bridge [56].

A meta-analysis that analysed 79 empirical publications confirmed that both cognitive and affective antecedents significantly predict green trust, which drives behavioural outcomes [65]. According to Riva et al. [66] and Khan et al. [63], the formation of green trust is fundamentally dependent on perceived credibility and transparency. Consumers interpret environmental claims as signals that need to be evaluated for reliability before they can be used to inform purchase decisions.

Empirical research has extensively examined the antecedents of green trust, showing that perceived credibility and authenticity are the primary drivers. Chen and Chang [8] indicated that green perceived quality has a positive impact on green trust, while green perceived risk has a negative impact, with green satisfaction acting as a partial mediator. Following studies expanded on this model, demonstrating that environmental friendliness influences green trust both directly and indirectly by green satisfaction and perceived quality, establishing a multi-path causal architecture for trust development [39,65].

These findings are consistent with signalling theory, which holds that firms' environmental actions, qualifications, and transparent disclosures provide signals that reduce information asymmetry and demonstrate expertise and ecological responsibility [9,50,63,66]. When consumers believe that green marketing signals are consistent, verifiable, and aligned with actual corporate practices, they attribute genuine environmental commitment to the brand [46,67]. And this commitment encourages green trust. On the other hand, inconsistencies or unsupported claims provoke scepticism and diminish trust. According to research from 2020 [9], both experimental and survey evidence demonstrate that green marketing communication that successfully enhances consumers' perceptions of company expertise resulted in significant trust increases, particularly among younger consumers.

The mediating role of green trust between greenwashing perceptions and consumer behaviour has been robustly established across multiple studies. Guerreiro and Pacheco [60] surveyed 302 participants and found that greenwashing expectations have no direct effect on green purchase decisions; instead, green trust and green word-of-mouth fully mediate this relationship. Similarly, De Sio et al. [61] demonstrated that trust fully mediates the relationship between green advertising scepticism and intention to buy green food among 410 Italian consumers, explaining 23% of the variability in purchase intention.

Credible key opinion leaders (KOLs) who truly promote real environmental actions boost green brand trust beyond direct brand communication, while vague or exaggerated language creates cynicism and greenwashing perceptions among tech-savvy consumers, particularly Generation Z [68,69]. These findings demonstrate that simply avoiding greenwashing is not enough to motivate consumers to make environmentally conscious purchases. Businesses need to actively establish green trust through credible communication and make certain that positive word-of-mouth promotes consumer trust [60].

Taken together, the existing literature demonstrates that green trust is neither automatically given nor permanently lost. It changes based on the situation and is shaped by how well the credibility of the message source, the tone of communication, and the evidence used to support it all fit together. Also, consumers must not feel like they are being lied to, which is a perception that sustainable marketers must actively foster and constantly maintain in order to keep customers' trust and behaviour engaged.

2.4. Source Disclosure and Transparency

The rapid growth of generative AI in digital marketing has led to regulatory and ethical debates over source disclosure, with policymakers increasingly requiring explicit labelling of AI-generated content to protect consumers from deception and manipulation. The European Union's AI Act [13] exemplifies this trend, establishing transparency obligations that require providers to inform users when they are interacting with AI systems

or consuming synthetic media [13,70]. Ethical frameworks also underline the “right to know” as a fundamental component of reliable AI marketing, claiming that disclosure protects against deepfakes, ensures accountability, and protects consumer privacy [71,72]. However, empirical evidence shows a significant disconnect between regulatory intent and consumer response: disclosing AI authorship consistently lowers perceived credibility, ad attitudes, and purchase intentions, especially for high-involvement products [12,14,73]. This pattern suggests that, while transparency mandates are morally compelling, they may unintentionally limit engagement with acceptable AI-enhanced content without improving the quality of the decision-relevant information available to consumers—the label signals who produced the message but adds nothing to the substance or verifiability of the environmental claim itself.

The *transparency paradox* theoretically explains this tension, which happens when increased disclosure activates psychological mechanisms that weaken rather than increase trust. According to research based on the persuasion knowledge model, AI labelling raises consumers’ awareness of persuasive intent, leading to cognitive defences and scepticism about message credibility [14,74,75]. *Algorithm aversion*, or the tendency to distrust automated outputs, creates these effects, as disclosures enhance perceptions of reduced authenticity and emotional resonance [76]. Similarly, Israfilzade [12] found that transparently labelled AI-generated advertisements have lower trust and purchase intent ratings, with the negative effect magnified for high-involvement products. Koning and Voorveld [15] expanded on these findings, demonstrating that AI disclosures reduce trust not only in the advertisements themselves but also in the organisations behind them, implying that the reputational costs of transparency go beyond immediate message evaluation and harm long-term brand equity.

Disclosure, on the other hand, can produce positive trust outcomes when combined with visible human-oversight indicators, ethical representation, and substantive consumer knowledge about AI capabilities—conditions that shift responsibility from algorithmic manipulation to institutional accountability [15,35,77,78]. This paradox has particularly serious implications for sustainability communication, in which authenticity and emotional resonance are essential psychological foundations for behavioural change. Disclosing AI involvement in environmental or prosocial messaging presents a double transparency responsibility, requiring consumers to trust both the environmental claims and the AI-generated content through which they are delivered [79,80].

Taken together, these studies position the present study at the intersection of four conversations that have so far run largely in parallel. Research on AI disclosure outside sustainability documents a consistent credibility penalty for content labelled as AI-generated [9,11,12,14] yet leaves open whether that penalty holds in a moral domain where authenticity is the message’s substance rather than a peripheral attribute. The source credibility tradition explains why attributed expertise and trustworthiness shape persuasion [81], but it treats the source as given rather than as a disclosed, potentially inaccurate label. Green scepticism and green trust studies establish that claim verifiability drives sustainability evaluations [7,8,49,60], while typically holding the human origin of the message constant. And work on consumer responses to automated communication shows that algorithm aversion is malleable and context-dependent [12,35,82,83], without extending to disclosed AI authorship in persuasive environmental appeals. By holding message content constant while logically varying the actual source and disclosed label, the current study connects these sections and determines whether the credibility associated with AI comes from the technology itself or the act of disclosing it.

Therefore, the growing body of academic opinion does not suggest that transparency should be abandoned as a moral principle [14,15,72,77,80]. Instead, it suggests that it

should be strategically implemented. This can be done by combining clear AI labelling with human-oversight signals, ethical communication frameworks, and AI literacy programmes that are meant to stop defensive scepticism.

2.5. Hypothesis Development

The theoretical foundations established in the preceding sections touch on a central empirical question: whether the source of a sustainable advertisement—whether generated by AI or crafted by human marketing professionals—interacts with source disclosure and environmental claim strength to shape consumer perceptions of greenwashing and green trust. Four interconnected theoretical streams inform the formulation of testable hypotheses.

Building on prior research, several theoretical perspectives jointly explain consumer responses to AI-generated sustainability advertising. First, the algorithm aversion literature [12,32,82,83] suggests that consumers systematically discount outputs attributed to algorithmic systems, particularly in high-involvement contexts such as sustainability, where authenticity and moral credibility are essential, and any perception of automated mediation may weaken trust. Second, the persuasion knowledge model [14,75,84] posits that disclosing AI authorship activates consumers' persuasion coping mechanisms, increasing scepticism as individuals become more alert to potential manipulation and interpret environmental claims as strategically optimised rather than genuinely reflective of corporate practices. Third, the greenwashing literature [5,7,49,85,86] differentiates between vague, symbolic claims that heighten scepticism due to lack of verifiability and substantive, evidence-based claims that reduce perceived deception and foster green trust by providing concrete evaluative benchmarks. Finally, the transparency paradox in AI marketing [12,15] highlights a critical tension whereby regulatory requirements for AI disclosure—intended to enhance transparency—can paradoxically diminish credibility and purchase intentions, especially in sustainability contexts where consumers must simultaneously assess both the environmental claims and the credibility of the AI-mediated source.

These streams converge on two complementary lenses—*source credibility theory* and *signalling theory*—that together explain why the *disclosure* of authorship, rather than authorship itself, governs evaluation. Source credibility is a multidimensional, receiver-side attribution whose most enduring core dimensions are perceived expertise and trustworthiness; higher credibility generally predicts greater persuasion and attitude change, though the effect is conditional rather than absolute—it can facilitate, inhibit, or have no systematic effect depending on the situation, and it matters most when receivers lack prior attitudes and cannot rely on message content to guide judgement [87–90]. Because credibility is inferred from available cues, the disclosed AI-versus-human label can shape it independently of the source's true nature. Signalling theory complements this by treating the environmental claim itself as a credibility signal whose value depends on its verifiability and cost [9,51,63,66]: vague claims are cheap, unverifiable signals, whereas substantive claims are costly, verifiable ones. Together the two lenses identify two partly independent inputs to judgement—a *source signal* (the label) and a *content signal* (claim strength).

Together, these theoretical frameworks support the expectation that responses to sustainable advertising will vary meaningfully across experimental conditions defined by ad generation source (AI vs. human-created), source label (labelled as AI-generated vs. labelled as human-created), and environmental claim strength (vague/symbolic vs. strong/substantive). The study operationalises these three dimensions in a mixed factorial design and examines their effects on two established outcome variables: perceived greenwashing and green trust. The following subsections develop and formally state the study's eight hypotheses.

2.5.1. Cross-Condition Differences in Perceived Greenwashing and Green Trust (H1–H4)

Hypotheses H1 through H4 address the central comparative question of this study: whether perceived greenwashing and green trust differ significantly across experimental conditions that simultaneously vary ad generation source and source labelling, under each level of claim strength. These hypotheses do not posit simple main effects attributable to a single factor in isolation; rather, they recognise that the joint configuration of source and label constitutes the primary unit of psychological meaning for the consumer. The consumer encounters not merely an advertisement generated by AI or by a human, but a configuration—an advertisement presented with a specific transparency disclosure that may or may not correspond to its true origin.

This configurational logic is theoretically motivated by the interplay between *source cues* and *label cues* in consumer information processing. Research on source effects on persuasion [81] has established that attribution of message origin substantially affects how claims are evaluated. When the attributed source is an AI system, consumers exhibit heightened scepticism due to algorithm aversion [76,82,83] and reduced perceptions of authenticity and emotional investment [34,35]. When the attributed source is a human team, consumers apply social trust heuristics and attribute genuine intentionality to environmental claims, thereby reducing greenwashing suspicion and facilitating trust formation.

Under *vague environmental claims* (H1 and H2), the information asymmetry inherent in symbolic language [5,50,91] is expected to interact with source attribution to produce pronounced differences across conditions, as vague claims provide no cognitive anchors to constrain consumer interpretation and thereby leave evaluations disproportionately shaped by source cues; AI-labelled advertisements should elicit heightened greenwashing perceptions because algorithm aversion compounds the inherent unverifiability of symbolic claims, while human labelling is expected to attenuate such perceptions by activating attributions of genuine personal commitment and moral accountability. Under *strong environmental claims* (H3 and H4), although substantive and specific claims reduce information asymmetry and provide verifiable benchmarks that should constrain greenwashing perceptions across all conditions [7,49], the transparency paradox predicts that AI labelling continues to impose a credibility penalty regardless of message quality, since AI disclosures activate persuasion knowledge mechanisms independently of claim substance and lead consumers to attribute strong claims in AI-labelled advertisements to algorithmic optimisation rather than authentic environmental performance [14,15,48,75]; consistent with algorithm aversion research showing that consumers resist updating source-based evaluations even when message content warrants it [12,82,83], significant differences across the four conditions are therefore expected under strong claims as well, with AI-labelled advertisements continuing to attract higher greenwashing perceptions and lower green trust relative to their human-labelled counterparts. Based on this reasoning, the following hypotheses are proposed:

H1. *There are significant differences in perceived greenwashing across advertisement conditions combining ad generation source (AI vs. human) and ad labelling (labelled as AI vs. labelled as human) when vague (symbolic) environmental claims are provided.*

H2. *There are significant differences in green trust across advertisement conditions combining ad generation source (AI vs. human) and ad labelling (labelled as AI vs. labelled as human) when vague (symbolic) environmental claims are provided.*

H3. *There are significant differences in perceived greenwashing across advertisement conditions combining ad generation source (AI vs. human) and ad labelling (labelled as AI vs. labelled as human) when strong (substantive) environmental claims are provided.*

H4. *There are significant differences in green trust across advertisement conditions combining ad generation source (AI vs. human) and ad labelling (labelled as AI vs. labelled as human) when strong (substantive) environmental claims are provided.*

2.5.2. The Effect of Claim Strength Within Correctly Disclosed Conditions (H5–H8)

Hypotheses H1 through H4 establish whether perceived greenwashing and green trust differ across the four experimental conditions at each level of claim strength. H5 through H8 address a related but conceptually distinct question: within conditions where the source label accurately reflects the advertisement's true generative origin—that is, AI-generated advertisements correctly labelled as AI-generated, and human-created advertisements correctly labelled as human-created—*does environmental claim strength alone produce a significant difference in consumer perceptions?* In other words, when there is no deception about the source, *does it still matter whether the environmental claim is vague or strong?*

This distinction is theoretically important. H1–H4 treat the four experimental configurations as the unit of comparison, meaning that both source and label vary simultaneously across conditions. H5–H8 hold the source–label relationship constant—fixing it at accurate disclosure—and examine the independent signal value of the claim content. The two sets of hypotheses are therefore not redundant: a significant overall pattern across conditions (H1–H4) does not, by itself, confirm that claim strength drives perceptions within any specific condition. It is entirely plausible that cross-condition differences are driven primarily by the label, while claim strength plays a negligible role once source attribution is held constant. Testing H5–H8 allows the study to determine whether claim type carries meaningful explanatory weight even after the source-disclosure dynamic is controlled for.

The theoretical basis for expecting a claim-strength effect under correctly disclosed conditions draws on the greenwashing literature and signalling theory. Vague, symbolic claims exploit information asymmetry by providing no verifiable basis for evaluation and are consistently identified as the primary linguistic trigger of greenwashing perceptions across consumer segments [5,7,8,55,92]. Conversely, strong, substantive claims function as credibility signals that reduce information asymmetry, support green trust formation, and enhance perceived corporate environmental expertise—effects that are particularly pronounced among environmentally literate younger consumers [8,9,66,93].

These claim-strength mechanisms are expected to operate within both the AI-disclosed and the human-disclosed correctly labelled conditions, though potentially with different magnitudes. Algorithm aversion suggests that substantive claims in AI-labelled advertisements may be interpreted as algorithmically optimised for persuasion rather than grounded in genuine corporate commitment, partially attenuating the trust-building potential of strong claims [10,12,82]. For human-disclosed advertisements, no such dampening is anticipated, as human authorship implies moral agency that amplifies the credibility signal value of substantive claims [34,35].

Importantly, however, H5–H8 do not formally test this difference in magnitude. They are stated as condition-specific directional hypotheses rather than moderation claims, which is consistent with both the study's design and the primary inferential interest: not whether the claim effect is stronger in one condition than another, but whether it exists and is significant in each correctly disclosed condition independently.

On the basis of this reasoning, the following hypotheses are formally stated:

H5. *For human-created advertisements with correct source disclosure, vague (symbolic) environmental claims will lead to higher perceived greenwashing than strong (substantive) claims.*

H6. *For human-created advertisements with correct source disclosure, vague (symbolic) environmental claims will lead to lower green trust than strong (substantive) claims.*

H7. *For AI-generated advertisements with correct source disclosure, vague (symbolic) environmental claims will lead to lower green trust than strong (substantive) claims.*

H8. *For AI-generated advertisements with correct source disclosure, vague (symbolic) environmental claims will lead to higher perceived greenwashing than strong (substantive) claims.*

Taken together, Hypotheses H1–H8 constitute a coherent theoretical framework for empirically examining the complex interplay of ad source, source disclosure, and environmental claim strength in sustainable advertising.

3. Methodology

3.1. Research Design

This study employed a 2 (*Ad Source*: AI-generated vs. human-created) $\times$ 2 (*Source Label*: labelled as AI-generated vs. labelled as human-created) $\times$ 2 (*Claim Strength*: vague vs. strong) mixed factorial experimental design. The first two factors—ad source and source label—were manipulated between subjects, meaning each participant was randomly assigned to one of four experimental groups defined by their combination: (1) human-created advertisement labelled as human-created (veridical human condition; control); (2) human-created advertisement labelled as AI-generated (mislabelled human condition); (3) AI-generated advertisement labelled as AI-generated (veridical AI condition); and (4) AI-generated advertisement labelled as human-created (mislabelled AI condition). The third factor—claim strength—was manipulated within subjects, such that all participants evaluated both a vague and a strong version of the advertisement assigned to their condition. This within-subject treatment of claim strength maximises statistical power for detecting claim-level effects while preserving the between-subject integrity of the source and label manipulations.

The between-subject design for source and label was chosen for both practical and theoretical reasons. Exposing participants to advertisements with different sources or labels within the same session would jeopardise demand characteristics, as the combination of AI- and human-attributed content would reveal the study's focus on source attribution and most likely misrepresent natural evaluative responses. The within-subject design for claim strength was considered appropriate because the vague and strong advertisements depicted different products and visual contexts (a water bottle and an electric scooter, respectively), reducing the risk of direct comparison effects while retaining the efficiency benefits of repeated measurement. This design yields eight distinct experimental cells, enabling the study to examine both cross-condition differences in consumer perceptions at each level of claim strength (H1–H4) and claim-strength effects within correctly disclosed conditions (H5–H8). All dependent variables—perceived greenwashing and green trust—were measured on five-item, seven-point Likert scales following each advertisement exposure, producing two measurement occasions per participant.

3.2. Stimuli Development

Two pairs of advertisements were developed for use in the experiment, each pair consisting of a vague-claim and a strong-claim version. The advertisements depicted distinct product categories—a reusable plastic water bottle for the vague-claim condition and an electric scooter for the strong-claim condition—both of which carry established sustainability associations and are familiar consumer product categories for Generation Z audiences. The use of different product categories across the two claim-strength conditions was a deliberate methodological choice: because claim strength was manipulated within subjects, using the same product for both vague and strong versions would have risked

carry-over effects whereby participants' evaluations of the second advertisement were systematically contaminated by their memory of the first.

The two claim-strength versions were operationalised against explicit, theory-derived criteria [9,49,60]. A claim was classified as *vague* if it used non-specific, unverifiable language (e.g., “eco-friendly,” “natural”) that conveyed an environmental orientation without any quantifiable referent, and as *strong* if it stated a specific, quantified, verifiable benefit (e.g., a numerical reduction in emissions or material use). The two levels differ in their degree of specificity, quantifiability and verifiability. The environmental theme, visual format and length of message were controlled. The precise claim language for each condition is provided in Appendix A.

Each advertisement existed in two source versions—AI-generated and human-created—yielding four unique visual stimuli in total. The AI-generated versions were produced using generative image models within ChatGPT Image 1.5, with carefully developed prompts specifying composition, visual elements, typography, lighting, and stylistic attributes to ensure professional advertising quality comparable to the human-created counterparts (see Appendix A, Table A1, for full prompts and stimuli examples). It is important to note that identical prompts do not guarantee identical outputs due to the inherent stochastic variability of generative models.

The next step was to identify advertisements created by humans. To reduce differences in product categories and appeal, Google Reverse Image was used to identify similar products. Following the selection of human-generated advertisements, all brand-related labels were removed to reduce respondents' potential biases toward specific brands.

Source disclosure was operationalised through a standardised transparency notice appended to the top of each advertisement. Participants in AI-labelled conditions saw the statement “*Transparency Notice: This advertisement and its environmental data were generated using Generative AI.*” Participants in human-labelled conditions saw “*Transparency Notice: This advertisement and its environmental data were created by a Professional Marketing Team.*” This manipulation was designed to reflect the type of source disclosure that is increasingly mandated by emerging regulatory frameworks, including Article 52 of the EU AI Act [13], while remaining ecologically plausible within the context of contemporary digital advertising practice.

The complete set of stimuli—including the word-for-word generative AI prompts, the final advertisement images for both product categories, and the exact wording of the source-disclosure notices used in each condition—is provided in Appendix A (Table A1) to support full transparency and reproducibility.

3.3. Sample and Data Collection

The target population for this study was Generation Z consumers, operationalised as undergraduate university students, for both theoretical and practical reasons. Generation Z represents the first cohort to grow up within an algorithmic media ecosystem in which AI-curated content, personalised advertising, and digital sustainability communication are ambient features of daily life, making them the most theoretically relevant audience for a study examining responses to AI-generated sustainable advertising.

Data were collected from undergraduate students enrolled at ADA University in Baku, Azerbaijan, over five consecutive days, from 2 to 6 February 2026. Participation was entirely voluntary, and the final sample comprised 154 respondents from the Generation Z demographic.

Participants were randomly assigned to one of four experimental conditions using Nimble Links, a randomisation tool that automatically directed each respondent to one of four unique survey URLs upon access, ensuring assignment without researcher in-

volvement and thereby eliminating assignment bias. To further protect internal validity, a double-blind design was employed in which neither participants nor those administering the study were aware of the specific condition to which respondents were being directed. Condition sizes ranged from 34 to 46 participants, with female respondents accounting for 69 participants (44.8%) and male respondents for 85 (55.2%). The full distribution of participants across experimental conditions by gender is presented in Table 1, confirming broadly proportional gender representation across all four groups.

Table 1. Distribution of participants across experimental conditions by gender.

Gender	Condition	<i>n</i>	% of Total	Cumulative %
Female	Human-created ad, labelled as human-created (control)	18	11.7 %	11.7 %
	Human-created ad, labelled as AI-generated	19	12.3 %	24.0 %
	AI-generated ad, labelled as AI-generated	12	7.8 %	31.8 %
	AI-generated ad, labelled as human-created	20	13.0 %	44.8 %
Male	Human-created ad, labelled as human-created (control)	20	13.0 %	57.8 %
	Human-created ad, labelled as AI-generated	17	11.0 %	68.8 %
	AI-generated ad, labelled as AI-generated	22	14.3 %	83.1 %
	AI-generated ad, labelled as human-created	26	16.9 %	100.0 %
Grand Total		154		

As shown in Table 1, participants were evenly distributed across the four experimental conditions, supporting balanced statistical power for between-group comparisons. This demographic homogeneity enhances internal validity by reducing variance attributable to age, academic experience, or cultural exposure, while the gender balance mitigates potential confounding effects of sex-based differences in sustainability attitudes. All participants provided informed consent prior to engagement.

As a further check on the comparability of the randomly assigned groups, participants' general importance attached to environmental sustainability in their purchasing decisions was measured prior to stimulus exposure using a single seven-point item. Baseline importance was uniformly high and closely comparable across the four conditions (veridical human control $M = 4.79$, $SD = 1.49$; human-created/AI-labelled $M = 4.56$, $SD = 1.68$; veridical AI $M = 4.59$, $SD = 1.62$; mislabelled AI $M = 4.63$, $SD = 1.60$), and a Welch's one-way ANOVA confirmed that the conditions did not differ significantly on this measure, $F(3, 81.2) = 0.17$, $p = 0.920$. This baseline equivalence indicates that the differences subsequently observed in perceived greenwashing and green trust cannot be attributed to pre-existing variation in environmental concern across the experimental groups.

To verify that gender was balanced across the experimental conditions, a Pearson chi-square test of independence was conducted on the gender $\times$ condition contingency table. The distribution of male and female participants did not differ significantly across the four conditions, $\chi^2(3, N = 154) = 2.30$, $p = 0.512$, Cramér's $V = 0.12$, confirming that gender was independent of experimental assignment and therefore unlikely to confound the observed effects.

3.4. Measures

This study employed a survey-based approach, collecting self-reported data through a structured online questionnaire. Validated measurement scales from the established literature were adapted to ensure construct reliability, content validity, and comparability with prior research in sustainable marketing and consumer psychology. All items were rated on a 7-point Likert scale (1 = *Strongly Disagree*; 7 = *Strongly Agree*). A full overview of

measurement items, item codes and factor loadings is presented in Tables A2 and A3 (see Appendix B).

Perceived greenwashing was measured using a five-item scale (GW1–GW5) assessing the degree to which participants perceived the advertisement’s environmental claims as exaggerated, misleading, or strategically deployed for marketing purposes rather than grounded in genuine environmental performance. The scale was primarily adapted from Chen and Chang [8], whose foundational conceptualisation of greenwashing operationalises the construct as consumers’ perception that a brand makes false, misleading, or unsubstantiated environmental claims. Additional item formulations were informed by Schmuck et al. [5], whose affect–reason–involvement framework identifies misleading green advertising as a distinct psychological experience involving both cognitive evaluation of claim validity and affective responses to perceived deception, and by Ha [94], whose mediation model of greenwashing and green brand equity employs analogous scale items to capture perceived environmental dishonesty.

Green trust was measured using a five-item scale (GT1–GT5) assessing participants’ confidence in the brand’s environmental claims, perceived genuineness of environmental commitment, and overall sense of reliability in the sustainability information presented. The scale was primarily adapted from Chen [92] and Chen and Chang [8], whose pioneering conceptualisation defines green trust as consumers’ willingness to depend on a product, service, or brand based on the belief that its environmental claims are credible and benevolent and fulfil stated commitments. Subsequent item formulations drew on Guerreiro and Pacheco [60], who operationalise green trust within a mediation model linking greenwashing perceptions to purchase intentions, and on Ha [94], who examines green trust as a mediating mechanism between greenwashing and brand equity outcomes.

The psychometric properties of both scales were evaluated across all experimental conditions prior to hypothesis testing. Confirmatory factor analysis confirmed that each scale loaded on a single latent factor, with all item loadings exceeding the 0.50 threshold conventionally required for acceptable convergent validity (see Appendix B for full factor loadings).

3.5. Reliability and Validity Assessment

The reliability and validity of both measurement scales were assessed across all eight experimental conditions—four under vague claims and four under strong claims—prior to hypothesis testing. Internal consistency was evaluated using Cronbach’s alpha, with all values exceeding the widely accepted 0.70 threshold. For perceived greenwashing, alpha ranged from 0.784 to 0.866 across the vague-claim conditions and from 0.784 to 0.861 across the strong-claim conditions. For green trust, alpha ranged from 0.839 to 0.910 under vague claims and from 0.863 to 0.914 under strong claims, indicating strong and consistent reliability across both constructs and all experimental groups (see Appendix B). Composite reliability (CR) values corroborated these findings, exceeding 0.70 in all conditions for both scales. Average variance extracted (AVE) values met or approached the conventional 0.50 threshold across conditions, with stronger convergence observed under strong-claim conditions, collectively supporting the convergent validity of both constructs.

Construct validity was further evaluated through confirmatory factor analysis (CFA) conducted separately for each experimental condition and each claim-strength context. Model fit was assessed using a standard battery of indices: the chi-square to degrees of freedom ratio (χ^2/df), the Comparative Fit Index (CFI), the Tucker–Lewis Index (TLI), and the Root Mean Square Error of Approximation (RMSEA). Under the vague-claim conditions, fit was acceptable to good across conditions, with χ^2/df ratios ranging from 1.210 to 1.659, CFI values between 0.852 and 0.952, TLI values between 0.804 and 0.936,

and RMSEA values between 0.078 and 0.132. While RMSEA values in Conditions 1 and 2 slightly exceeded the preferred 0.08 threshold, the remaining indices remained within acceptable ranges, and the overall pattern is consistent with the relatively small sample sizes per condition, which are known to inflate RMSEA estimates. Under the strong-claim conditions, model fit was markedly stronger and more uniform across all four groups, with χ^2/df ratios between 1.010 and 1.508, CFI values between 0.929 and 0.997, TLI values between 0.906 and 0.996, and RMSEA values between 0.018 and 0.105, indicating good to excellent fit in all conditions.

Taken together, the reliability and validity evidence support the appropriateness of both scales for use in this experimental context.

3.6. Data Analysis Strategy

To test *Hypotheses H1 through H4*, one-way analyses of variance (ANOVAs) were conducted separately for *perceived greenwashing* and *green trust* under each claim-strength condition, treating the four experimental groups—defined by the combination of ad source and source label—as the independent variable. Where ANOVA results indicated significant overall differences, Tukey’s Honest Significant Difference (HSD) post hoc tests were applied to identify which specific pairwise comparisons drove the observed effects, while controlling the familywise error rate. Condition-specific comparisons were employed rather than a full factorial modelling approach because the study’s primary inferential interest lies in identifying meaningful differences across practically defined advertising configurations—each representing a distinct real-world combination of source origin and transparency disclosure—rather than in estimating isolated main effects or interaction terms that would obscure these configurations. To test *Hypotheses H5 through H8*, paired-samples *t*-tests were conducted within the two correctly disclosed conditions—the veridical AI condition and the veridical human condition—comparing perceived greenwashing and green trust scores between the vague- and strong-claim exposures, exploiting the within-subject nature of the claim-strength manipulation. The results of all analyses are reported in the following section.

4. Results

4.1. Hypothesis Testing (H1–H4)

Hypotheses 1 through 4 were evaluated using one-way ANOVAs followed by Tukey’s Honestly Significant Difference (HSD) post hoc tests, as pre-specified in the data analysis strategy. The omnibus tests examined whether perceived greenwashing and green trust differed significantly across the four source–label configurations under each claim-strength condition.

4.1.1. Vague-Claim Conditions (H1 and H2)

H1 and H2 predicted significant differences in perceived greenwashing and green trust across the four experimental conditions when vague, symbolic environmental claims were presented. Both hypotheses were supported. The one-way ANOVA for perceived greenwashing under vague claims was significant, $F(3, 82.5) = 7.19, p < 0.001, \eta^2 = 0.12$, with condition means revealing a pronounced pattern consistent with the theoretical predictions (Table 2).

The control condition again returned the lowest greenwashing mean ($M = 4.07, SD = 1.067$), while the AI-labelled human condition ($M = 4.84, SD = 0.925$) and the veridical AI condition ($M = 5.05, SD = 1.082$) returned substantially higher scores. The mislabelled AI condition ($M = 4.19, SD = 1.359$) remained closer to the control, suggesting that concealing AI origin behind a human label partially attenuated greenwashing perceptions even when

claims were vague. For green trust under vague claims, the ANOVA was similarly significant, $F(3, 82.8) = 6.09, p < 0.001, \eta^2 = 0.10$. The control condition recorded the highest trust level ($M = 4.43, SD = 1.169$), while the veridical AI condition ($M = 3.58, SD = 0.900$) and the AI-labelled human condition ($M = 3.66, SD = 0.821$) exhibited the sharpest trust reductions, with the mislabelled AI condition ($M = 4.29, SD = 1.405$) again performing comparably to the control. *H1 and H2 are therefore both supported.*

Table 2. One-way ANOVA results for perceived greenwashing and green trust under vague environmental claims.

Variable/Condition	<i>n</i>	Mean	SD	SE	ANOVA F	df1	df2	<i>p</i> -Value
<i>Greenwashing</i>								
Human-created ad, labelled as human-created (control)	38	4.07	1.067	0.173	7.19	3	82.5	<0.001
Human-created ad, labelled as AI-generated	36	4.84	0.925	0.154				
AI-generated ad, labelled as AI-generated	34	5.05	1.082	0.186				
AI-generated ad, labelled as human-created	46	4.19	1.359	0.200				
<i>Green Trust</i>								
Human-created ad, labelled as human-created (control)	38	4.43	1.169	0.190	6.09	3	82.8	<0.001
Human-created ad, labelled as AI-generated	36	3.66	0.821	0.137				
AI-generated ad, labelled as AI-generated	34	3.58	0.900	0.154				
AI-generated ad, labelled as human-created	46	4.29	1.405	0.207				

4.1.2. Strong-Claim Conditions (H3 and H4)

H3 and H4 predicted significant differences in perceived greenwashing and green trust, respectively, across the four experimental conditions when strong, substantive environmental claims were presented. Both hypotheses were supported (Table 3). A one-way ANOVA revealed a significant overall difference in perceived greenwashing across the four conditions, $F(3, 81.8) = 6.11, p < 0.001, \eta^2 = 0.11$.

Table 3. One-way ANOVA results for perceived greenwashing and green trust under strong environmental claims.

Variable/Condition	N	Mean	SD	SE	ANOVA F	df1	df2	p-Value
Greenwashing								
Human-created ad, labelled as human-created (Control)	38	3.06	0.996	0.162	6.11	3	81.8	<0.001
Human-created ad, labelled as AI-generated	36	3.74	0.868	0.145				
AI-generated ad, labelled as AI-generated	34	4.02	0.964	0.165				
AI-generated ad, labelled as human-created	46	3.60	1.012	0.149				
Green Trust								
Human-created ad, labelled as human-created (Control)	38	5.29	1.221	0.198	4.24	3	82	0.008
Human-created ad, labelled as AI-generated	36	5.13	1.12	0.187				
AI-generated ad, labelled as AI-generated	34	4.35	1.179	0.202				
AI-generated ad, labelled as human-created	46	4.80	1.285	0.190				

Condition means indicated a clear pattern of elevation in greenwashing perceptions as AI involvement—whether real or attributed—entered the advertising context. The veridical human condition (human ad, labelled human) returned the lowest greenwashing score ($M = 3.06, SD = 0.996$), followed by the AI-labelled human condition ($M = 3.74, SD = 0.868$), the mislabelled AI condition ($M = 3.60, SD = 1.012$), and the veridical AI condition ($M = 4.02, SD = 0.964$). For green trust under strong claims, the ANOVA result was likewise significant, $F(3, 82) = 4.24, p = 0.008, \eta^2 = 0.08$. The control condition exhibited the highest green trust ($M = 5.29, SD = 1.221$), followed by the AI-labelled human condition ($M = 5.13, SD = 1.120$), the mislabelled AI condition ($M = 4.80, SD = 1.285$), and the veridical AI condition ($M = 4.35, SD = 1.179$), confirming that AI source attribution—particularly when correctly disclosed—

meaningfully suppressed trust even in the presence of specific, verifiable environmental claims. *H3 and H4 are therefore both supported.*

4.1.3. Post Hoc Analysis

Tukey's HSD post hoc tests were conducted to identify the specific pairwise comparisons driving the significant ANOVA effects (see Appendix C). The results, illustrated in Figures 1 and 2 for the vague- and strong-claim conditions respectively, reveal a coherent and theoretically meaningful pattern of differences centred on two key contrasts: the effect of attributing AI authorship to an advertisement—whether truthfully or through mislabelling—relative to the veridical human baseline, and the effect of correct source disclosure on perceptions within AI-labelled conditions.

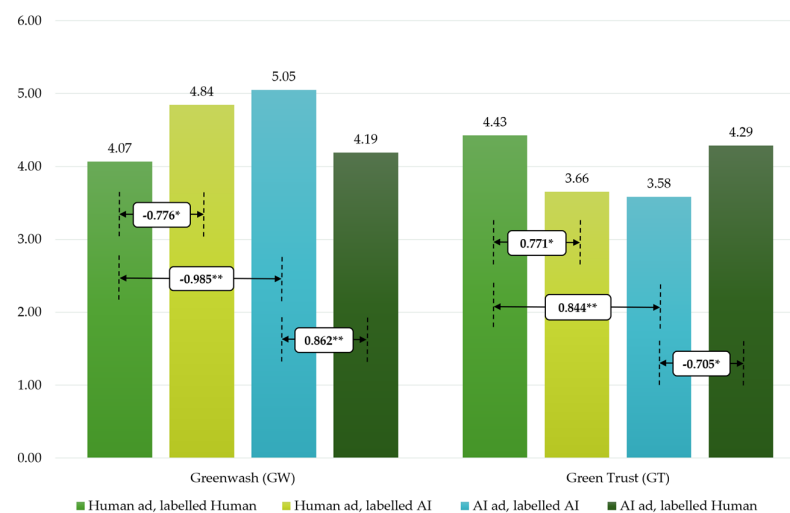


Figure 1. Mean perceived greenwashing and green trust scores across four experimental conditions under vague environmental claims. Note. * $p < 0.05$, ** $p < 0.01$.

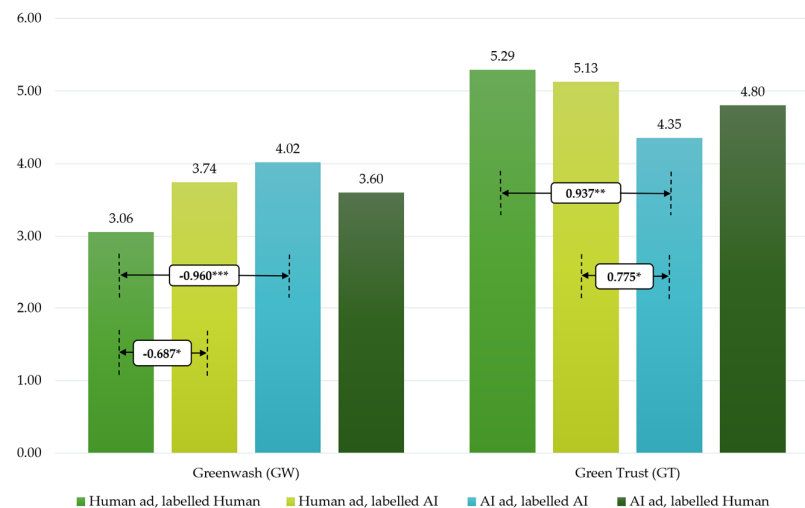


Figure 2. Mean perceived greenwashing and green trust scores across four experimental conditions under strong environmental claims. Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Under vague claims (Figure 1), the post hoc pattern was more extensive and the effect sizes larger, consistent with the theoretical expectation that vague claims leave consumer interpretation more exposed to source cues.

For perceived greenwashing, the control differed significantly from both the AI-labelled human condition (Mean $\bar{D} = -0.776$, $p = 0.020$) and the veridical AI condition (Mean $\bar{D} = -0.985$, $p = 0.002$), again demonstrating that AI attribution—regardless of

the actual source—substantially elevated greenwashing perceptions when environmental claims provided no verifiable informational anchor. Notably, the veridical AI condition also differed significantly from the mislabelled AI condition (Mean $\bar{D} = 0.862$, $p = 0.006$), revealing that correctly disclosing AI authorship on a vague-claim advertisement produced significantly higher greenwashing suspicion than when concealing that same AI origin behind a human label—a finding with direct implications for transparency policy. For green trust under vague claims, the control differed significantly from the AI-labelled human condition (Mean $\bar{D} = 0.771$, $p = 0.019$) and the veridical AI condition (Mean $\bar{D} = 0.844$, $p = 0.010$), while the veridical AI condition also differed significantly from the mislabelled AI condition (Mean $\bar{D} = -0.705$, $p = 0.032$). Together, these results confirm that under vague claims, AI disclosure—correct or mislabelled—is the primary driver of trust erosion, and that the label, rather than the actual generative source, constitutes the operative psychological cue shaping consumer perceptions.

Under strong claims (Figure 2), Tukey comparisons confirmed that the veridical human condition differed significantly from both the AI-labelled human condition (Mean $\bar{D} = -0.687$, $p = 0.014$) and the veridical AI condition (Mean $\bar{D} = -0.960$, $p < 0.001$) on perceived greenwashing, indicating that AI attribution—whether real or merely labelled—elevated greenwashing perceptions even when the advertisement presented quantified, verifiable environmental claims. The mislabelled AI condition did not differ significantly from the control ($p = 0.052$), suggesting that concealing an AI-generated advertisement behind a human label effectively neutralised greenwashing suspicion under strong claims.

For green trust under strong claims, the most substantial pairwise differences emerged between the control and the veridical AI condition (Mean $\bar{D} = 0.937$, $p = 0.007$) and between the AI-labelled human condition and the veridical AI condition (Mean $\bar{D} = 0.775$, $p = 0.040$), confirming that correct AI disclosure imposed the greatest trust penalty, and that this penalty persisted even relative to conditions where the advertisement was human-created but attributed to AI.

4.2. Paired-Samples *t*-Test Results (H5–H8)

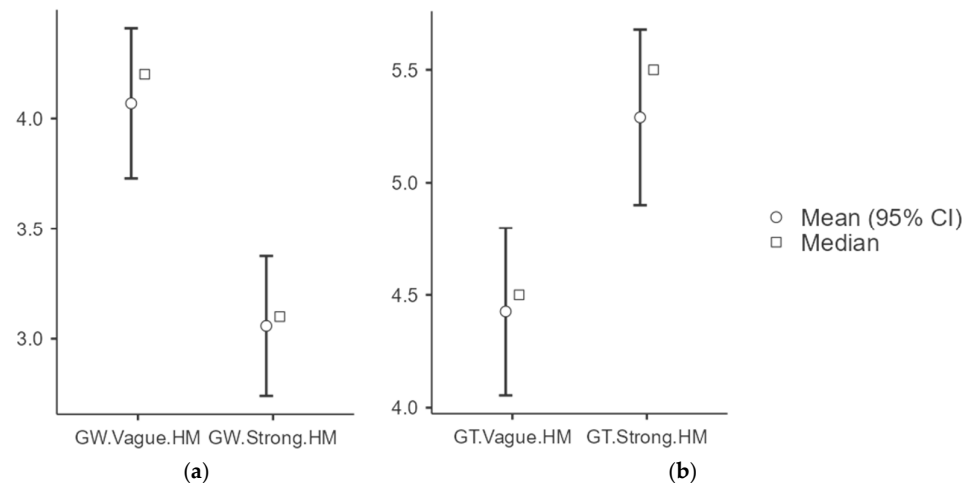
To test Hypotheses H5 through H8, paired-samples *t*-tests were conducted within the two correctly disclosed conditions—the veridical human condition (human ad, labelled human) and the veridical AI condition (AI ad, labelled AI)—comparing perceived greenwashing and green trust scores between the vague- and strong-claim exposures. These analyses address whether claim strength alone produces a significant shift in consumer perceptions when source disclosure is accurate, independently of the cross-condition source and label dynamics examined in H1–H4.

H5 and H6—Human-Created Advertisements with Correct Disclosure. H5 predicted that for human-created advertisements correctly labelled as human-created, vague claims would elicit significantly higher perceived greenwashing than strong claims, and H6 predicted the corresponding suppression of green trust under vague claims. Both hypotheses were supported (Table 4). Participants in the human-created, human-labelled condition reported substantially higher greenwashing perceptions under vague claims ($M = 4.07$) than under strong claims ($M = 3.06$), with a mean difference of 1.011 ($SE = 0.155$; 95% CI [0.697, 1.324]), $t(37) = 6.52$, $p < 0.001$, *Cohen's d* = 1.058—a very large effect.

The corresponding green trust results showed the predicted reverse pattern: trust was considerably lower under vague claims ($M = 4.43$) than under strong claims ($M = 5.29$), with a mean difference of -0.863 ($SE = 0.232$; 95% CI [-1.332 , -0.394]), $t(37) = -3.73$, $p < 0.001$, *Cohen's d* = -0.605 , a medium effect. As illustrated in Figure 3, the 95% confidence intervals for the vague- and strong-claim conditions are fully separated for both outcomes.

Table 4. Paired-samples *t*-test: vague vs. strong claims under correct human disclosure.

Human-Created Ad, Labelled as Human-Created		Statistic	df	<i>p</i>	Mean $\bar{D}$	SE Difference	95% Conf. Interval		Effect Size
							Lower	Upper	<i>Cohen's d</i>
GW_Vague	GW_Strong	6.52	37	<0.001	1.011	0.155	0.697	1.324	1.058
GT_Vague	GT_Strong	−3.73	37	<0.001	−0.863	0.232	−1.332	−0.394	−0.605

**Figure 3.** Paired comparisons of (a) perceived greenwashing and (b) green trust by claim strength: human-created advertisement with correct human disclosure.

Together, H5 and H6 demonstrate that within the human-disclosed context, claim strength operates as a robust independent driver of consumer perceptions: substantive, verifiable environmental claims meaningfully reduce greenwashing suspicion and elevate trust even when source attribution is unambiguous, and consumers know the advertisement was produced by human professionals.

H7 and H8—AI-Generated Advertisements with Correct Disclosure. H7 predicted that for AI-generated advertisements correctly labelled as AI-generated, vague claims would produce significantly lower green trust than strong claims, and H8 predicted the corresponding elevation of perceived greenwashing under vague claims. Both hypotheses were supported (Table 5). Participants who viewed the AI-generated advertisement correctly labelled as AI-generated reported markedly lower trust under vague claims ($M = 3.58$) than under strong claims ($M = 4.35$), with a mean difference of -0.771 ($SE = 0.174$; 95% CI $[-1.125, -0.416]$), $t(33) = -4.42$, $p < 0.001$, *Cohen's d* = -0.758 —a medium-to-large effect.

Table 5. Paired-samples *t*-test: vague vs. strong claims under correct AI disclosure.

AI Ad, Labelled AI		Statistic	df	<i>p</i>	Mean $\bar{D}$	SE Difference	95% Conf. Interval		Effect Size
							Lower	Upper	<i>Cohen's d</i>
GW_Vague	GW_Strong	4.83	33	<0.001	1.035	0.214	0.599	1.471	0.828
GT_Strong	GT_Vague	−4.42	33	<0.001	−0.771	0.174	−1.125	−0.416	−0.758

The corresponding greenwashing results showed substantially higher perceptions under vague claims ($M = 5.05$) than under strong claims ($M = 4.02$), with a mean difference of 1.035 ($SE = 0.214$; 95% CI $[0.599, 1.471]$), $t(33) = 4.83$, $p < 0.001$, *Cohen's d* = 0.828 , a large effect. Figure 4 confirms clearly non-overlapping confidence intervals for both outcomes within this condition.

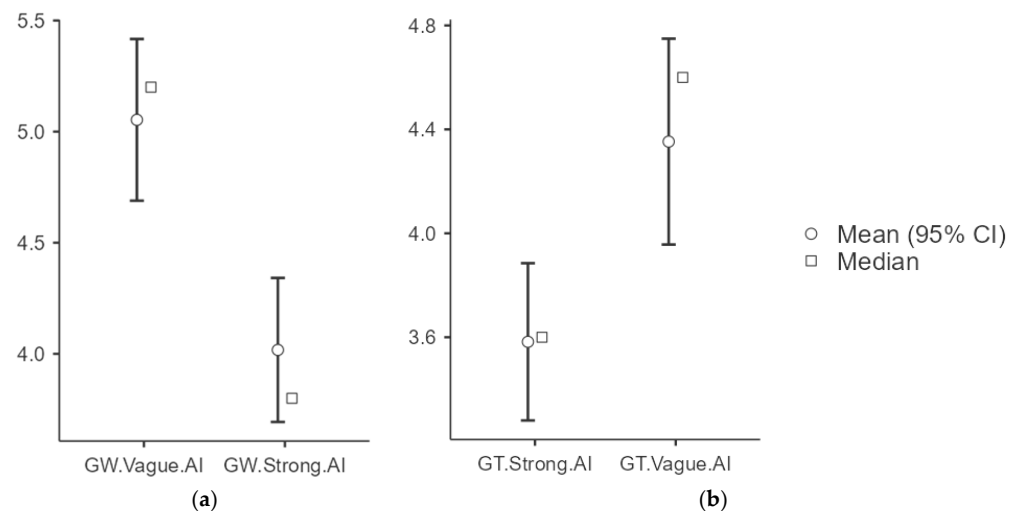


Figure 4. Paired comparisons of (a) perceived greenwashing and (b) green trust by claim strength: AI-generated advertisements with correct AI disclosure.

These results establish that even under conditions of correct AI disclosure—where algorithm aversion theoretically dampens trust formation and elevates baseline scepticism—claim strength continues to produce meaningful and statistically significant shifts in consumer perceptions. Notably, however, the absolute levels of greenwashing remain higher and the absolute levels of green trust remain lower in the veridical AI condition than in the veridical human condition at both claim-strength levels (vague greenwash: 5.05 vs. 4.07; strong greenwash: 4.02 vs. 3.06; vague trust: 3.58 vs. 4.43; strong trust: 4.35 vs. 5.29), indicating that while substantive claims significantly attenuate the negative effects of AI disclosure within this condition, they do not fully eliminate the persistent baseline penalty associated with transparent AI attribution—a finding revisited in the discussion.

4.3. Summary of Hypothesis Testing

Across the eight hypotheses tested, the results provide consistent and statistically robust support for the study's theoretical framework. Hypotheses H1 through H4 confirmed that perceived greenwashing and green trust differ significantly across the four experimental configurations of ad source and source label under both vague- and strong-claim conditions, with all four ANOVAs returning significant overall effects. Hypotheses H5 through H8 further confirmed that within both correctly disclosed conditions—human-disclosed and AI-disclosed—vague environmental claims produce significantly higher greenwashing perceptions and significantly lower green trust than strong claims, with all four paired-sample comparisons yielding significant results and medium-to-large effect sizes. All eight hypotheses were therefore supported. A consolidated overview of the hypothesis testing outcomes is presented in Table 6.

Table 6. Summary of hypothesis testing results.

Hypothesis	Result	Supported
H1—Perceived greenwashing significantly differs across ad source $\times$ label conditions under vague (symbolic) claims.	$F(3, 82.5) = 7.19, p < 0.001$	Yes
H2—Green trust significantly differs across ad source $\times$ label conditions under vague (symbolic) claims.	$F(3, 82.8) = 6.09, p < 0.001$	Yes
H3—Perceived greenwashing significantly differs across ad source $\times$ label conditions under strong (substantive) claims.	$F(3, 81.8) = 6.11, p < 0.001$	Yes

Table 6. Cont.

Hypothesis	Result	Supported
H4—Green trust significantly differs across ad source $\times$ label conditions under strong (substantive) claims.	$F(3, 82) = 4.24, p = 0.008$	Yes
H5—For human-created ads with correct disclosure, vague claims lead to higher perceived greenwashing than strong claims.	$t(37) = 6.52, p < 0.001, d = 1.058$	Yes
H6—For human-created ads with correct disclosure, vague claims lead to lower green trust than strong claims.	$t(37) = -3.73, p < 0.001, d = -0.605$	Yes
H7—For AI-generated ads with correct disclosure, vague claims lead to lower green trust than strong claims.	$t(33) = -4.42, p < 0.001, d = -0.758$	Yes
H8—For AI-generated ads with correct disclosure, vague claims lead to higher perceived greenwashing than strong claims.	$t(33) = 4.83, p < 0.001, d = 0.828$	Yes

5. Discussion

5.1. Summary of Key Findings

This study examined how AI-generated versus human-created sustainable advertising, in combination with source disclosure and environmental claim strength, influences perceived greenwashing and green trust among Generation Z consumers. Three core findings emerge. First, AI involvement in sustainability communication—whether real or merely attributed through labelling—consistently elevated perceived greenwashing and depressed green trust relative to the veridical human control condition, with the AI-disclosed condition recording the highest greenwashing scores ($M = 5.05$ under vague claims; $M = 4.02$ under strong claims) and the lowest green trust scores ($M = 3.58$ under vague claims; $M = 4.35$ under strong claims). Second, vague environmental claims produced markedly stronger negative effects than strong claims across all four experimental configurations, with within-condition effect sizes ranging from *Cohen's d* = 0.605 to 1.058.

The third and most striking finding is that the *label*—rather than the actual generative source—is the dominant psychological cue shaping consumer perceptions. Human-created advertisements mislabelled as AI-generated produced greenwashing and trust outcomes statistically indistinguishable from those of veridical AI advertisements, while AI-generated advertisements mislabelled as human-created performed comparably to the veridical human control. What consumers believe about an advertisement's source—not what is objectively true about its origin—drives their evaluative response, with profound implications for transparency policy in AI-mediated marketing.

5.2. Theoretical Implications

The findings make three interrelated theoretical contributions. First, the consistent depression of green trust in AI-attributed conditions extends the documented phenomenon of *algorithm aversion* [12,75,76,82,83] into the moral domain of sustainability communication, suggesting that AI scepticism is not merely cognitive but morally inflected: consumers interpret AI involvement in sustainability messaging as a marker of reduced authentic commitment, even when the underlying claims are substantive. Second, the elevation of greenwashing perceptions even under strong, quantified claims when AI labelling is present provides direct empirical support for the *persuasion knowledge model* as applied to AI-mediated communication [14,74,75,84], demonstrating that AI disclosure activates persuasion coping mechanisms that operate independently of message content—consumers attribute even substantive claims to algorithmic optimisation rather than genuine corporate commitment.

Third, the within-condition results for H5–H8 reinforce *signalling theory* in green marketing [9,63,66,94] while clarifying its boundary conditions. Within both correctly disclosed conditions, strong claims significantly reduced greenwashing and elevated trust, confirming that substantive environmental information functions as a credibility signal. However, the absolute trust levels achieved within the AI-disclosed condition under strong claims ($M = 4.35$) remained considerably below those in the human-disclosed condition under strong claims ($M = 5.29$), indicating that signal credibility is itself moderated by source attribution: the same quantified environmental commitment is interpreted as a more credible signal when attributed to a human team than to an AI system.

5.3. The Transparency Paradox: A Core Contribution

The most theoretically and practically significant finding concerns the *transparency paradox* in AI-mediated sustainability communication. While regulatory frameworks such as Article 52 of the EU AI Act [13] mandate explicit AI disclosure on the assumption that transparency protects consumer interests, the present results reveal a counterintuitive consequence: AI labelling itself, regardless of whether it accurately reflects the advertisement's true origin, significantly damages consumer perceptions and reduces brand credibility. Under vague claims, the mislabelled human condition recorded greenwashing $M = 4.84$ and trust $M = 3.66$ —statistically indistinguishable from the veridical AI condition's $M = 5.05$ and $M = 3.58$ —meaning that genuinely human-produced sustainability content suffers the same credibility penalty as algorithmic content the moment it carries an AI label.

Why should AI disclosure act as such a powerful negative heuristic? The label seems to be a secondary clue: instead of focusing on the message's content, readers assume that AI authorship means strategic optimisation, less authenticity, and a higher chance of greenwashing. This inference is based on morals instead of logic, so it is especially harmful in sustainability, where sincerity of intent is what makes something valuable. The penalty persists even with strong, quantified claims, which suggests that the label makes consumers view AI-generated sustainability messages as manipulative once activated. Verifiable content can partially redeem an AI-labelled message, suggesting that consumers withhold rather than permanently reject credibility. The heuristic applies to disclosure itself, not content, so mandatory AI labelling risks penalising transparency unless paired with credibility-restoring signals discussed in Section 5.5.

Conversely, AI-generated advertisements mislabelled as human-created performed comparably to the veridical human control across both outcomes and both claim conditions, with the mislabelled AI condition under vague claims yielding greenwashing $M = 4.19$ and trust $M = 4.29$ against the control's $M = 4.07$ and $M = 4.43$. This demonstrates that AI-produced content, when concealed behind a human label, escapes the credibility penalty associated with transparent AI disclosure entirely. The implication is that the regulatory premise underlying mandated AI disclosure is partially inverted in the sustainability context: disclosure does not enable better information processing but instead activates source-based scepticism, creating a perverse incentive structure in which the most ethical brands—those that disclose honestly—bear a credibility cost that brands willing to obscure AI involvement avoid. Resolving this paradox requires moving beyond binary disclosure mandates toward frameworks that pair AI labelling with credibility-restoring signals, such as third-party verification or visible human oversight.

5.4. Comparison with Prior Literature

The findings converge with several established lines of research while extending and complicating prior conclusions. The negative effect of AI labelling on perceived credibility aligns with Israfilzade [12], Baek et al. [14], and Koning and Voorveld [15], who collectively

document that AI disclosures reduce trust in advertising and corporate communication. The present study extends these findings by demonstrating that the credibility penalty applies not only to advertisement evaluations but specifically to perceptions of environmental claim authenticity—a domain where authenticity is constitutive of the message’s purpose. The “algorithm discount” identified by Altay and Gilardi [10] and Lee et al. [11] is here shown to operate in the morally laden context of sustainability, where its consequences are arguably more severe than in the news or general persuasion contexts where it has previously been documented. Similarly, the claim-strength findings reinforce Schmuck et al.’s [5,91], Pham and Barretta’s [49] and Persakis et al.’s [7] characterisation of symbolic environmental language as a primary driver of consumer scepticism, while the trust-enhancing effect of strong claims under correct disclosure aligns with the signalling-theoretic predictions of Chen and Chang [8] and Waites et al. [9].

The most distinctive contribution of this study lies in its empirical demonstration that the *label*, not the actual source, is the operative psychological cue. Existing research has tended to treat source and disclosure as either equivalent or causally linked, but the mislabelling conditions in this study reveal that consumer perceptions track the disclosed source independently of the actual generative origin. This finding extends the persuasion knowledge framework [14,46,84] by showing that the activation of persuasion coping is triggered by the *cue* of AI disclosure rather than by any property of the content itself, and it raises empirical questions for the regulatory literature on AI transparency that prior work has not directly addressed.

5.5. Implications for Digital Marketing Policy

The findings carry significant implications for the design of AI disclosure regulation in digital advertising. The dominant regulatory paradigm—exemplified by Article 52 of the EU AI Act and emerging national frameworks [13,72]—assumes that mandated AI disclosure protects consumer interests by enabling informed evaluation of automated content. The present study challenges this assumption: because AI labels appear to activate generalised source-based scepticism rather than enable nuanced evaluation, simple binary disclosure mandates may inadvertently undermine the very consumer protections they are designed to provide, suppressing engagement with legitimate environmental commitments and creating perverse incentives for non-compliance among brands willing to obscure AI involvement.

Policy frameworks would benefit from moving beyond binary disclosure toward graduated and context-sensitive transparency requirements. AI disclosure could be paired with mandated indicators of human oversight—for example, third-party certification of environmental claims or explicit identification of human strategic input—since the literature suggests that disclosure combined with credibility-restoring signals can preserve trust where bare disclosure does not [35,65,77]. Differentiated disclosure regimes for high-stakes communication domains such as sustainability, public health, and political messaging may also be warranted, alongside investment in consumer-facing AI literacy programmes that can shift the interpretive lens through which AI labels are processed. There is additionally a regulatory case for substantiation requirements specifically for AI-generated sustainability claims—requiring documentation of quantified benchmarks or a third-party audit before such claims may circulate—which would address greenwashing and AI transparency concerns simultaneously.

An additional imbalance needs consideration. At the moment, disclosure rules only apply to content made by AI, not content made by humans. Given that a human attribute consistently increased credibility in this study (H4), this imbalance suggests that a truthful “human-made” or “human-supervised” signal could also act as a positive trust cue. This is

something that should be looked at by regulators and ethical experts, as these signals can only protect consumers if they are true and verifiable.

5.6. Implications for Practitioners

For marketing practitioners, the findings carry three actionable implications. First, brands should categorically avoid vague, symbolic environmental claims, regardless of whether the content is AI-generated or human-created. Across every experimental condition, vague claims produced higher greenwashing perceptions and lower green trust than strong claims, with effect sizes consistently in the medium-to-large range. The strategic intuition that vague claims are commercially safer because they are less assertive is contradicted by the data: among environmentally literate Generation Z consumers, vague language reads as evasive rather than modest and is interpreted as evidence of insincere commitment rather than cautious accuracy.

Second, brands should approach AI labelling with strategic care, pairing any AI disclosure with explicit signals of human oversight—statements identifying the human strategic team that approved the content, the third-party audit that verified its environmental claims, or the editorial process that shaped its messaging—in order to shift the credibility narrative from “this content was produced by an algorithm” to “this content was produced collaboratively by a human team supported by AI tools” [35,38,45]. Third, brands seeking to maximise green trust should combine substantive claims with credible source signals: the single highest-trust configuration in this study was the human-created, human-labelled, strong-claim condition ($M = 5.29$), which achieved substantially higher trust than any AI-attributed configuration. While this does not imply that brands should avoid AI tools entirely—the efficiency advantages are well-documented—it does suggest that high-stakes claim-bearing content should be reserved for human-led or visibly human-supervised production processes, while lower-stakes promotional content carries less reputational risk from AI deployment.

These implications translate into distinct priorities for different actors. For *sustainable brand managers*, the immediate priority is to audit existing campaigns for vague environmental language and replace it with quantified, independently verifiable claims, treating substantive specificity as the primary lever of green trust. For *digital advertisers*, the actionable rule is to never deploy an AI label in isolation: any disclosure should be accompanied, within the same creative unit, by a visible human-oversight or third-party-verification cue that reframes the content as human-supervised rather than autonomously generated. For *organisations using generative AI in ESG communication*, the findings recommend a tiered content strategy—reserving high-stakes, claim-bearing sustainability messages for human-led or visibly human-supervised production, while confining unsupervised AI generation to lower-stakes promotional material. For *policymakers* implementing AI disclosure regulation, the corresponding design recommendations are developed in Section 5.5. Across all four audiences, the central managerial challenge is to balance transparency against trust restoration: because bare disclosure imposes a credibility penalty, the practical question is not *whether* to disclose but *how* to couple honest disclosure with credibility-restoring signals so that transparency and trust become complementary rather than competing objectives.

5.7. Limitations and Future Research

This study has several limitations that bound its conclusions. First, the analytical strategy intentionally privileged condition-specific comparisons (one-way ANOVA, Tukey HSD, paired *t*-tests) over a full factorial ANOVA model with formal interaction tests. While this approach was theoretically motivated by the study’s interest in practically meaningful advertising configurations, it does not formally estimate the higher-order interaction effects

between source, label, and claim strength that a $2 \times 2 \times 2$ factorial model would produce; future research should complement the present design with formal factorial modelling and structural equation modelling (SEM) approaches that can simultaneously estimate the contributions of source, label, and claim strength while accounting for measurement error in the latent constructs. Second, the study focused exclusively on Generation Z consumers sampled from a single university, which—while theoretically defensible given Generation Z's distinctive relationship with algorithmic media—limits generalisation to other demographic groups whose AI literacy and environmental attitudes differ; cross-generational research would clarify whether the transparency paradox documented here generalises beyond this cohort.

In addition, participants were recruited through convenience sampling at a single institution, and although random assignment produced groups that were equivalent on baseline environmental importance and balanced on gender, the study did not measure other individual-difference variables—notably prior familiarity with generative AI and general digital literacy—that may plausibly moderate responses to AI disclosure. Future research should adopt probability-based or multi-site quota sampling and incorporate these variables as measured moderators to establish the boundary conditions of the present effects.

Next, two related features of the claim-strength manipulation warrant interpretive caution. First, no pre-tested manipulation-check items confirmed that participants consciously perceived the intended difference in claim specificity; although the large, consistent claim-strength effects offer indirect evidence, future work should include validated checks, ideally in a separate pre-test. Second, because claim strength was manipulated within subjects using two different products in a fixed order (a vague-claim water bottle followed by a strong-claim electric scooter), claim strength, product category, and presentation order are confounded in the within-subject comparisons (H5–H8), which may therefore partly reflect product or sequence differences. Relatedly, the study operationalised source disclosure as a binary AI-versus-human label; graduated labels distinguishing fully AI-generated from AI-assisted or human-supervised content represent an important direction for future research. Critically, this does not affect the study's primary contribution: the dominance of the disclosed label over the actual source rests on the between-subject comparisons (H1–H4) in which product category and presentation order are held constant within each claim-strength level.

Lastly, the study employed a controlled experimental setting in which participants evaluated single advertisements in a survey environment, maximising internal validity but not capturing the full ecology of digital advertising exposure within social feeds and across multiple touchpoints over time; field experimental research embedding the manipulations within real-world digital environments would strengthen external validity and enable measurement of behavioural rather than purely attitudinal outcomes. Finally, the cultural context of the study represents a specific configuration of norms regarding sustainability, brand communication, and AI familiarity, and cross-cultural replication comparing high-AI-adoption Western markets, emerging Asian markets, and Global South contexts would clarify whether the transparency paradox is a generalisable feature of human–AI communication or culturally moderated, with significant implications for the international harmonisation of AI transparency regulation.

6. Conclusions

This study set out to investigate how AI-generated versus human-created sustainable advertising, in combination with source disclosure and environmental claim strength, influences perceived greenwashing and green trust among Generation Z consumers. Across

eight hypotheses, the results demonstrate that consumer responses to sustainability advertising are jointly shaped by source attribution, transparency disclosure, and claim content, but that the disclosed source label—rather than the actual generative origin of the advertisement—operates as the dominant psychological cue. AI-attributed advertisements consistently elevated perceived greenwashing and depressed green trust relative to human-attributed counterparts, and this penalty applied equally to genuinely AI-generated content and to human-created content mislabelled as AI. Conversely, AI-generated content concealed behind a human label produced consumer responses comparable to those elicited by the veridical human control, indicating that the credibility cost of AI involvement is triggered by the act of disclosure itself rather than by any inherent property of the AI-produced content.

The study's findings on environmental claim strength reinforce and extend established greenwashing scholarship. Within both correctly disclosed conditions—human-disclosed and AI-disclosed—vague, symbolic environmental claims produced significantly higher greenwashing perceptions and significantly lower green trust than strong, substantive claims, with medium-to-large effect sizes confirming that claim content carries independent and substantial explanatory weight. Yet the study also reveals an important boundary condition: even under correctly disclosed AI conditions accompanied by quantified, verifiable environmental claims, absolute green trust levels remained considerably below those achieved in the human-disclosed condition with strong claims. This indicates that substantive claim content, while necessary, is not sufficient to fully neutralise the credibility penalty associated with transparent AI attribution. Strong claims significantly mitigate but do not eliminate the AI disclosure effect.

The most consequential contribution of this study is the empirical articulation of the *transparency paradox* in AI-mediated sustainability communication. Regulatory frameworks worldwide are moving toward mandatory disclosure of AI-generated content on the assumption that transparency protects consumer interests by enabling informed evaluation. The present findings suggest that, in the sustainability context, this assumption is partially inverted: AI labels appear to activate generalised source-based scepticism rather than enable nuanced evaluation of underlying content, and they impose this credibility cost regardless of whether the disclosed AI authorship is accurate. The result is a perverse incentive structure in which the most ethically transparent brands bear a credibility penalty that brands willing to obscure AI involvement avoid. Resolving this paradox will require regulatory frameworks that move beyond binary disclosure toward graduated approaches pairing AI labels with credibility-restoring signals such as third-party verification, visible human oversight, or substantiation requirements for AI-generated sustainability claims. For practitioners, the implications are clear: vague environmental claims should be categorically avoided regardless of source, AI labelling should never be deployed without accompanying signals of human strategic involvement, and high-stakes sustainability content carrying substantive environmental commitments should be reserved for human-led or visibly human-supervised production processes. While the study is bounded by its focus on Generation Z, its single-university sample, and its controlled experimental setting, the consistency and magnitude of the observed effects suggest that the transparency paradox documented here is a robust phenomenon with significant implications for the academic literature on consumer psychology in AI-mediated communication, for the international harmonisation of AI transparency regulation, and for the practical design of corporate sustainability communication in the era of generative AI.

Funding: This research received no external funding. The APC was funded by Kauno Kolegija Higher Education Institution, funding number: No. 1073, 9 June 2026.

Institutional Review Board Statement: Ethical review and approval were waived for this study, as it involved an anonymous, voluntary online questionnaire administered via Google Forms to adult participants, with no intervention, clinical procedures, or collection of biological samples. The project was considered to have no risk and involved only non-identifiable survey data. Participation was entirely voluntary, and data were collected only after participants provided informed consent. No direct or indirect identifiers were recorded, and all data were handled in accordance with the General Data Protection Regulation (GDPR).

Informed Consent Statement: All participants were provided with an informed consent information sheet prior to participation, outlining the purpose of the study, the voluntary nature of participation, the anonymity of responses, and the handling of data in accordance with the GDPR. Participants were informed that by clicking the survey link and proceeding to the questionnaire, they were providing their informed consent to take part in the study. Participation was entirely voluntary, and no personally identifying information was collected at any stage.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy and ethical restrictions, as participants provided informed consent for the use of their responses solely within the scope of this academic study and in accordance with the GDPR, EU 2016/679.

Acknowledgments: The author would like to express sincere gratitude to Aistė Lastauskaitė, Faculty of Business, Kauno Kolegija Higher Education Institution, for her invaluable support throughout this research. Heartfelt thanks are also extended to Viktorija Navickienė, Communication Department, Faculty of Business, Kauno Kolegija Higher Education Institution. The author further acknowledges the broader academic community of Kauno Kolegija Higher Education Institution for providing the supportive scholarly environment in which this work was completed.

Conflicts of Interest: The author declares no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

This appendix documents the stimulus materials developed for the four experimental conditions. For the vague-claim advertisement, the copy employed unsubstantiated symbolic language characteristic of green embellishment. The headline read “Pure water, cleaner planet,” accompanied by the tagline “Our eco-friendly bottle uses 25% recycled plastic. Together we can reduce waste.” While the reference to 25% recycled plastic content introduces a degree of specificity, the broader claim structure remains anchored in vague, non-verifiable sustainability language—“eco-friendly,” “cleaner planet,” “reduce waste”—with no quantified environmental benchmarks, independent verification, or science-based targets referenced. The strong-claim advertisement featured substantially more substantive environmental communication. The headline employed a deliberate rhetorical inversion—“Scooters don’t make a difference”. Critically, the advertisement incorporated three quantified, verifiable environmental performance claims presented as infographic statistics: “Just 20 g CO₂ per km,” “90% lower emissions than cars,” and “Save 1.2 tons of CO₂ per year.” This operationalisation aligns with established definitions of substantive environmental claims, which require specific, measurable, and independently evaluable information about environmental performance.

The four resulting advertisements—paired across claim strength and source attribution—are presented in Table A1 below.

Table A1. Advertisements used in the experiment for vague and strong environmental claims.

Vague Claim	Strong Claim
	
Human-generated ad	Human-generated ad
	
AI-generated ad *	AI-generated ad *
* Prompt was used to generate an AI image in ChatGPT. Professional eco-friendly product advertisement, centered composition of a transparent plastic water bottle floating slightly tilted, green cap, minimal label with a simple green leaf icon. Bright vibrant green gradient background with soft glow and particles. Surrounding elements include a small Earth globe held in a green hand, wind turbine, solar panel, glowing light bulb with leaf inside, and small green plants. Clean modern advertising style, sustainability theme, soft diffused lighting, high detail reflections on plastic, 3D render style, glossy finish, depth of field, subtle bokeh. Typography included: Top text: "Pure water, cleaner planet." Bottom text: "Our eco-friendly bottle uses 25% recycled plastic. Together we can reduce waste." Minimalist layout, balanced composition, fresh and futuristic look, 4K, commercial design, brandless.	* Prompt was used to generate an AI image in ChatGPT. Realistic outdoor billboard advertisement mockup on a modern city street, large glass display case. Inside the billboard: a sleek green electric scooter positioned on the right side, studio lighting, sharp shadows, ultra-detailed. Background inside billboard is dark green textured gradient with soft lighting. Bold modern typography on the left: "Scooters don't make a difference." (with "don't" slightly scratched or crossed stylistically). Subtext: "Switch to electric and reduce your carbon footprint. Enjoy a cleaner, quieter commute." Infographic icons and stats: "Just 20 g CO₂ per km"; "90% lower emissions than cars"; "Save 1.2 tons of CO₂ per year" Bottom line: "Join the movement for a greener future." Photorealistic style, high contrast, cinematic lighting, urban environment reflections on glass, premium advertising design, 8K resolution.

Note. The AI-generated ads used the two prompts mentioned above, although providing the same prompt does not guarantee similar outputs due to the intrinsic diversity of generative models.

Appendix B

Table A2. Factor loadings, reliability, and validity metrics for perceived greenwashing and green trust across four vague environmental claims conditions.

Construct	Code	Items	Factor Loading			
			Con. 1	Con. 2	Con. 3	Con. 4
Perceived Greenwashing	GW1	The environmental claims in this advertisement seem exaggerated.	0.712	0.601	0.742	0.820
	GW2	The advertisement makes the product appear more environmentally friendly than it really is.	0.639	0.674	0.871	0.569
	GW3	The environmental message in this advertisement seems misleading.	0.864	0.504	0.711	0.771
	GW4	The advertisement appears to use environmental claims mainly as a marketing tactic.	0.845	0.784	0.776	0.633
	GW5	The brand may be overstating its environmental responsibility.	0.694	0.855	0.521	0.727
		<i>Cronbach's α</i>	0.866	0.806	0.846	0.830
		<i>Composite Reliability (CR)</i>	0.868	0.819	0.850	0.833
		<i>Average Variance Extracted (AVE)</i>	0.571	0.483	0.538	0.504
Green Trust	GT1	I trust the environmental claims made by this brand.	0.688	0.755	0.719	0.829
	GT2	This brand appears genuinely committed to environmental sustainability.	0.672	0.702	0.721	0.716
	GT3	I believe the environmental information provided in this advertisement.	0.735	0.752	0.756	0.835
	GT4	The brand seems honest about its environmental impact.	0.769	0.692	0.803	0.894
	GT5	I feel confident that this product is environmentally responsible.	0.766	0.684	0.809	0.814
		<i>Cronbach's α</i>	0.848	0.839	0.873	0.910
		<i>Composite Reliability (CR)</i>	0.848	0.841	0.874	0.911
		<i>Average Variance Extracted (AVE)</i>	0.529	0.515	0.582	0.672

Table A3. Factor loadings, reliability, and validity metrics for perceived greenwashing and green trust across four strong environmental claims conditions.

Construct	Code	Items	Factor Loading			
			Con. 1	Con. 2	Con. 3	Con. 4
Perceived Greenwashing	GW1	The environmental claims in this advertisement seem exaggerated.	0.707	0.720	0.772	0.842
	GW2	The advertisement makes the product appear more environmentally friendly than it really is.	0.763	0.602	0.647	0.473
	GW3	The environmental message in this advertisement seems misleading.	0.779	0.717	0.738	0.655
	GW4	The advertisement appears to use environmental claims mainly as a marketing tactic.	0.649	0.635	0.856	0.727
	GW5	The brand may be overstating its environmental responsibility.	0.651	0.623	0.742	0.700
		<i>Cronbach's α</i>	0.833	0.784	0.861	0.808
		<i>Composite Reliability (CR)</i>	0.836	0.794	0.867	0.815
		<i>Average Variance Extracted (AVE)</i>	0.507	0.437	0.569	0.476
Green Trust	GT1	I trust the environmental claims made by this brand.	0.777	0.735	0.805	0.752
	GT2	This brand appears genuinely committed to environmental sustainability.	0.750	0.888	0.698	0.830
	GT3	I believe the environmental information provided in this advertisement.	0.807	0.484	0.847	0.881
	GT4	The brand seems honest about its environmental impact.	0.897	0.742	0.809	0.879
	GT5	I feel confident that this product is environmentally responsible.	0.741	0.862	0.850	0.820
		<i>Cronbach's α</i>	0.897	0.863	0.899	0.914
		<i>Composite Reliability (CR)</i>	0.896	0.865	0.901	0.919
		<i>Average Variance Extracted (AVE)</i>	0.634	0.571	0.646	0.695

Appendix C

Table A4. Tukey post hoc test results for vague environmental claims products.

Tukey Post Hoc Test—Greenwashing					
		Human ad, labelled human (control)	Human ad, labelled AI	AI ad, labelled AI	AI ad, labelled human
Human ad, labelled human (control)	Mean $\bar{D}$	—	−0.776 *	−0.985 **	−0.123
	<i>p</i> -value	—	0.020	0.002	0.961
Human ad, labelled AI	Mean $\bar{D}$	—	—	−0.208	0.653
	<i>p</i> -value	—	—	0.869	0.052
AI ad, labelled AI	Mean $\bar{D}$	—	—	—	0.862 **
	<i>p</i> -value	—	—	—	0.006
AI ad, labelled human	Mean $\bar{D}$	—	—	—	—
	<i>p</i> -value	—	—	—	—

Tukey Post Hoc Test—Green Trust					
		Human ad, labelled human (control)	Human ad, labelled AI	AI ad, labelled AI	AI ad, labelled human
Human ad, labelled human (control)	Mean $\bar{D}$	—	0.771 *	0.844 **	0.139
	<i>p</i> -value	—	0.019	0.010	0.942
Human ad, labelled AI	Mean $\bar{D}$	—	—	0.0732	−0.631
	<i>p</i> -value	—	—	0.993	0.060
AI ad, labelled AI	Mean $\bar{D}$	—	—	—	−0.705 *
	<i>p</i> -value	—	—	—	0.032
AI ad, labelled human	Mean $\bar{D}$	—	—	—	—
	<i>p</i> -value	—	—	—	—

Note. * $p < 0.05$, ** $p < 0.01$.

Table A5. Tukey post hoc test results for vague strong claims products.

Tukey Post Hoc Test—Greenwashing					
		Human ad, labelled human (control)	Human ad, labelled AI	AI ad, labelled AI	AI ad, labelled human
Human ad, labelled human (control)	Mean $\bar{D}$	—	−0.687 *	−0.960 ***	−0.546
	<i>p</i> -value	—	0.014	<0.001	0.052
Human ad, labelled AI	Mean $\bar{D}$	—	—	−0.273	0.14
	<i>p</i> -value	—	—	0.638	0.915
AI ad, labelled AI	Mean $\bar{D}$	—	—	—	0.413
	<i>p</i> -value	—	—	—	0.236
AI ad, labelled human	Mean $\bar{D}$	—	—	—	—
	<i>p</i> -value	—	—	—	—

Tukey Post Hoc Test—Green Trust					
		Human ad, labelled human (control)	Human ad, labelled AI	AI ad, labelled AI	AI ad, labelled human
Human ad, labelled human (control)	Mean $\bar{D}$	—	0.162	0.937 **	0.485
	<i>p</i> -value	—	0.939	0.007	0.263
Human ad, labelled AI	Mean $\bar{D}$	—	—	0.775 *	0.323
	<i>p</i> -value	—	—	0.040	0.627
AI ad, labelled AI	Mean $\bar{D}$	—	—	—	−0.451
	<i>p</i> -value	—	—	—	0.354
AI ad, labelled human	Mean $\bar{D}$	—	—	—	—
	<i>p</i> -value	—	—	—	—

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

References

1. Pilelienė, L.; Bogoyavlenska, Y. Artificial Intelligence in Influencer Marketing: Current Researchscape, Trends and Insights from a Bibliometric Review. *Equilib. Q. J. Econ. Econ. Policy* **2025**, *20*, 583–611. [\[CrossRef\]](#)
2. Israfilzade, K.; Sadili, N. Beyond Traditional AI: How Generative Models Are Transforming E-Commerce Ecosystems. In *Proceedings of the 2025 IEEE 19th International Conference on Application of Information and Communication Technologies (AICT), AI Ain, United Arab Emirates, 29–31 October 2025*; IEEE: New York, NY, USA, 2025; pp. 1–6.
3. Grewal, D.; Saturnino, C.B.; Davenport, T.; Guha, A. How Generative AI Is Shaping the Future of Marketing. *J. Acad. Mark. Sci.* **2025**, *53*, 702–722. [\[CrossRef\]](#)
4. Salih, S.; Husain, O.; Mohamed Almohamedh, R.; Tajelsier, H.; Hassan Abdalla Hashim, A.; Elshafie, H.; Motwakel, A. From Ideation to Execution: Unleashing the Power of Generative AI in Modern Digital Marketing and Customer Engagement—A Systematic Literature Review and Case Study. *Array* **2026**, *29*, 100630. [\[CrossRef\]](#)
5. Schmuck, D.; Matthes, J.; Naderer, B. Misleading Consumers with Green Advertising? An Affect–Reason–Involvement Account of Greenwashing Effects in Environmental Advertising. *J. Advert.* **2018**, *47*, 127–145. [\[CrossRef\]](#)
6. Zaid, A.A.; Bawaqni, S.; Shahwan, R.; Alnasr, F. Effects of Greenwashing on Green Purchase Intention: The Mediating Role of Green Skepticism, Green Brand Love and Green Brand Loyalty. *J. Foodserv. Bus. Res.* **2025**, *28*, 927–954. [\[CrossRef\]](#)
7. Persakis, A.; Nikolopoulos, T.; Negkakis, I.C.; Pavlopoulos, A. Greenwashing in Marketing: A Systematic Literature Review and Bibliometric Analysis. *Int. Rev. Public Nonprofit Mark.* **2025**, *22*, 957–992. [\[CrossRef\]](#)
8. Chen, Y.-S.; Chang, C.-H. Greenwash and Green Trust: The Mediation Effects of Green Consumer Confusion and Green Perceived Risk. *J. Bus. Ethics* **2013**, *114*, 489–500. [\[CrossRef\]](#)
9. Waites, S.F.; Stevens, J.L.; Hancock, T. Signaling Green: Investigating Signals of Expertise and Prosocial Orientation to Enhance Consumer Trust. *J. Consum. Behav.* **2020**, *19*, 632–644. [\[CrossRef\]](#)
10. Altay, S.; Gilardi, F. People Are Skeptical of Headlines Labeled as AI-Generated, Even If True or Human-Made, Because They Assume Full AI Automation. *PNAS Nexus* **2024**, *3*, pgae403. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Lee, D.C.; Jhang, J.; Baek, T.H. AI-Generated News Content: The Impact of AI Writer Identity and Perceived AI Human-Likeness. *Int. J. Hum.–Comput. Interact.* **2025**, *41*, 13862–13874. [\[CrossRef\]](#)
12. Israfilzade, K. AI-Generated versus Human-Created Advertising: Effects on Consumer Trust and Purchase Intent. *Equilib. Q. J. Econ. Econ. Policy* **2025**, *20*, 1301–1337. [\[CrossRef\]](#)
13. Smuha, N.A. European Parliament and Council. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence. *Off. J. Eur. Union* **2024**, 32024R1689. [\[CrossRef\]](#)
14. Baek, T.H.; Kim, J.; Kim, J.H. Effect of Disclosing AI-Generated Content on Prosocial Advertising Evaluation. *Int. J. Advert.* **2026**, *45*, 171–192. [\[CrossRef\]](#)
15. Koning, B.; Voorveld, H.A.M. Disclaimer! This Content Is AI-Generated: How AI-Disclosures Influence Trust in Advertisements and Organizations. *J. Interact. Advert.* **2025**, *25*, 240–253. [\[CrossRef\]](#)
16. Israfilzade, K.; Sadili, N. Beyond Interaction: Generative AI in Conversational Marketing—Foundations, Developments, and Future Directions. *JLECON* **2024**, *11*, 13–29. [\[CrossRef\]](#)
17. Gao, B.; Wang, Y.; Xie, H.; Hu, Y.; Hu, Y. Artificial Intelligence in Advertising: Advancements, Challenges, and Ethical Considerations in Targeting, Personalization, Content Creation, and Ad Optimization. *Sage Open* **2023**, *13*, 21582440231210759. [\[CrossRef\]](#)
18. Matz, S.C.; Teeny, J.D.; Vaid, S.S.; Peters, H.; Harari, G.M.; Cerf, M. The Potential of Generative AI for Personalized Persuasion at Scale. *Sci. Rep.* **2024**, *14*, 4692. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Meguellati, E.; Han, L.; Bernstein, A.; Sadiq, S.; Demartini, G. How Good Are LLMs in Generating Personalized Advertisements? In *Companion Proceedings of the ACM Web Conference 2024, Singapore, 13–17 May 2024*; ACM: New York, NY, USA, 2024; pp. 826–829.
20. Kshetri, N. Generative AI in Advertising. *IT Prof.* **2024**, *26*, 15–19. [\[CrossRef\]](#)
21. Gupta, R.; Nair, K.; Mishra, M.; Ibrahim, B.; Bhardwaj, S. Adoption and Impacts of Generative Artificial Intelligence: Theoretical Underpinnings and Research Agenda. *Int. J. Inf. Manag. Data Insights* **2024**, *4*, 100232. [\[CrossRef\]](#)
22. Aghaei, R.; Kiaei, A.A.; Boush, M.; Vahidi, J.; Zavvar, M.; Barzegar, Z.; Rofoosheh, M. Harnessing the Potential of Large Language Models in Modern Marketing Management: Applications, Future Directions, and Strategic Recommendations. *arXiv* **2025**, arXiv:2501.10685.
23. Reed, C.; Wynn, M.; Bown, R. Artificial Intelligence in Digital Marketing: Towards an Analytical Framework for Revealing and Mitigating Bias. *BDCC* **2025**, *9*, 40. [\[CrossRef\]](#)
24. Borimnejad, H.; Borimnejad, V. Emerging Trends Challenges and Research Opportunities in Artificial Intelligence Applications in Marketing. *Discov. Artif. Intell.* **2025**, *6*, 9. [\[CrossRef\]](#)
25. Israfilzade, K. The Role of Generative Artificial Intelligence in E-Commerce: Trends, Challenges and Opportunities. *EPES* **2025**, *42*, 1–20. [\[CrossRef\]](#)

26. Chaisatitkul, A.; Luangngamkhum, K.; Noulpum, K.; Kerdvibulvech, C. The Power of AI in Marketing: Enhancing Efficiency and Improving Customer Perception through AI-Generated Storyboards. *Int. J. Inf. Technol.* **2024**, *16*, 137–144. [\[CrossRef\]](#)
27. Zhang, L.; Hur, C. The Impact of Generative AI Images on Consumer Attitudes in Advertising. *Adm. Sci.* **2025**, *15*, 395. [\[CrossRef\]](#)
28. Israfilzade, K. Perspective Chapter: The Impact of Generative AI in Shaping Customer Engagement—Humanising Conversational Marketing. In *Business, Management and Economics*; Sabah, S., Buran, I., Eds.; IntechOpen: London, UK, 2025; Volume 39.
29. Wu, L.; Dodoo, N.A.; Wen, T.J. Disclosing AI's Involvement in Advertising to Consumers: A Task-Dependent Perspective. *J. Advert.* **2025**, *54*, 20–38. [\[CrossRef\]](#)
30. Hamouda, M.; Aissaoui, A. Cracking the Code: How Can AI-Generated Advertising Be Effective? Empirical Evidence on the Antecedents and Consequences of Advertising Value. *Eur. Bus. Rev.* **2026**, *38*, 224–243. [\[CrossRef\]](#)
31. Ryoo, Y.; Bakpayev, M.; Jeon, Y.A.; Kim, K.; Yoon, S. High Hopes, Hard Falls: Consumer Expectations and Reactions to AI-Human Collaboration in Advertising. *Int. J. Advert.* **2026**, *45*, 48–80. [\[CrossRef\]](#)
32. Rix, J.; Berger, B.; Hess, T.; Rzepka, C. The Algorithm Discount: Explaining Consumers' Valuation of Human- versus Algorithm-Created Digital Products. *J. Manag. Inf. Syst.* **2025**, *42*, 633–668. [\[CrossRef\]](#)
33. Israfilzade, K. Conversational Marketing as a Framework for Interaction with the Customer: Development & Validation of the Conversational Agent's Usage Scale. *JLECON* **2021**, *8*, 533–546. [\[CrossRef\]](#)
34. Brüns, J.D.; Meißner, M. Do You Create Your Content Yourself? Using Generative Artificial Intelligence for Social Media Content Creation Diminishes Perceived Brand Authenticity. *J. Retail. Consum. Serv.* **2024**, *79*, 103790. [\[CrossRef\]](#)
35. Haupt, M.; Freidank, J.; Haas, A. Consumer Responses to Human-AI Collaboration at Organizational Frontlines: Strategies to Escape Algorithm Aversion in Content Creation. *Rev. Manag. Sci.* **2025**, *19*, 377–413. [\[CrossRef\]](#)
36. Singh, D. The Role of AI-Generated Advertising Sentiments in Poaching Consumer Emotions: A Behavioural Marketing Approach. *IJSREM* **2025**, *09*, 1–9. [\[CrossRef\]](#)
37. Sallaku, K.; Macca, L.S.; Santoro, G.; Thu Phan, T. Advertising in the Digital Era: Can AI Evoke Emotions like Humans? *Bus. Process Manag. J.* **2025**, *31*, 1785–1811. [\[CrossRef\]](#)
38. Kubovics, M. Influencing the Perception and Acceptance of AI-Generated Marketing Content as Part of SGD. *J. Lifestyle SDGs Rev.* **2025**, *5*, e06929. [\[CrossRef\]](#)
39. Chen, Y.-S.; Lin, C.-L.; Chang, C.-H. The Influence of Greenwash on Green Word-of-Mouth (Green WOM): The Mediation Effects of Green Perceived Quality and Green Satisfaction. *Qual. Quant.* **2014**, *48*, 2411–2425. [\[CrossRef\]](#)
40. Rae, I. The Effects of Perceived AI Use on Content Perceptions. In *Proceedings of the CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 11–16 May 2024*; ACM: New York, NY, USA; pp. 1–14.
41. Dutta-Powell, R.; Rhee, J.J.; Wodak, S. Two Interventions for Mitigating the Harms of Greenwashing on Consumer Perceptions. *Bus. Strategy Environ.* **2024**, *33*, 882–903. [\[CrossRef\]](#)
42. Neureiter, A.; Matthes, J. Comparing the Effects of Greenwashing Claims in Environmental Airline Advertising: Perceived Greenwashing, Brand Evaluation, and Flight Shame. *Int. J. Advert.* **2023**, *42*, 461–487. [\[CrossRef\]](#)
43. Neureiter, A.; Stubenvoll, M.; Matthes, J. Is It Greenwashing? Environmental Compensation Claims in Advertising, Perceived Greenwashing, Political Consumerism, and Brand Outcomes. *J. Advert.* **2024**, *53*, 511–529. [\[CrossRef\]](#)
44. Huang, Z.; Shi, Y.; Jia, M. Greenwashing: A Systematic Literature Review. *Account. Financ.* **2025**, *65*, 857–882. [\[CrossRef\]](#)
45. Keilmann, J.; Koch, T. Little Green Lies: Greenwashing's Detrimental Impact on Corporate Perception and the Interplay of Perceived Motives, Skepticism and Authenticity. *Corp. Commun. Int. J.* **2026**, *31*, 259–276. [\[CrossRef\]](#)
46. Balaskas, S.; Stamatiou, I.; Komis, K.; Nikolopoulos, T. Perceptions of Greenwashing and Purchase Intentions: A Model of Gen Z Responses to ESG-Labeled Digital Advertising. *Risks* **2025**, *13*, 157. [\[CrossRef\]](#)
47. Kwon, K.; Lee, J.; Wang, C.; Diwanji, V.S. From Green Advertising to Greenwashing: Content Analysis of Global Corporations' Green Advertising on Social Media. *Int. J. Advert.* **2024**, *43*, 97–124. [\[CrossRef\]](#)
48. Orazi, D.C.; Chan, E.Y. "They Did Not Walk the Green Talk!:" How Information Specificity Influences Consumer Evaluations of Disconfirmed Environmental Claims. *J. Bus. Ethics* **2020**, *163*, 107–123. [\[CrossRef\]](#)
49. Pham, N.T.; Barretta, P.G. Green Marketing or Greenwashing: How Consumers Evaluate Environmental Ads. *JABE* **2024**, *26*, 19–33. [\[CrossRef\]](#)
50. Srisathan, W.A.; Naruetharadhol, P. Exploring Moral Hazard and Adverse Selection in the Context of Greenwashing and Organic Product Consumption. *J. Retail. Consum. Serv.* **2025**, *84*, 104203. [\[CrossRef\]](#)
51. Fernandes, J.; Segev, S.; Leopold, J.K. When Consumers Learn to Spot Deception in Advertising: Testing a Literacy Intervention to Combat Greenwashing. *Int. J. Advert.* **2020**, *39*, 1115–1149. [\[CrossRef\]](#)
52. Farooq, Y.; Wicaksono, H. Advancing on the Analysis of Causes and Consequences of Green Skepticism. *J. Clean. Prod.* **2021**, *320*, 128927. [\[CrossRef\]](#)
53. Rehman, A.U.; Kumar, S.; Alghafes, R.; Broccardo, L.; Patel, A.K. Role of Greenwashing in Influencing Brand Attitude and Consumption: Identifying Sustainable Business Strategies. *Bus. Strategy Environ.* **2025**, *34*, 6686–6703. [\[CrossRef\]](#)

54. Javed, U.; Rashid, M.A.; Hussain, G.; Shafique, S. Consumers' Perception of Green Brand Attributes and Its Outcomes: A Sustainable Perspective. *J. Environ. Plan. Manag.* **2024**, *67*, 1655–1677. [\[CrossRef\]](#)
55. Daou, L.; Sayegh, E.; Atallah, E.; Jabbour Al Maalouf, N.; Sarkis, N. Greenwashing as a Barrier to Sustainable Marketing: Expectation Disconfirmation, Confusion, and Brand–Consumer Relationships. *Sustainability* **2025**, *17*, 9979. [\[CrossRef\]](#)
56. Rahman, S.U.; Nguyen-Viet, B. Towards Sustainable Development: Coupling Green Marketing Strategies and Consumer Perceptions in Addressing Greenwashing. *Bus. Strategy Environ.* **2023**, *32*, 2420–2433. [\[CrossRef\]](#)
57. Capgemini Research Institute. A World in Balance: Heightened Sustainability Awareness yet Lagging Actions. 2023. Available online: <https://www.capgemini.com/insights/research-library/sustainability-trends-2023/> (accessed on 14 December 2025).
58. Strother, J.B.; Fazal, Z. Can Green Fatigue Hamper Sustainability Communication Efforts? In *Proceedings of the 2011 IEEE International Professional Communication Conference, Cincinnati, OH, USA, 17–19 October 2011*; IEEE: New York, NY, USA, 2011; pp. 1–6.
59. Nguyen, N.B.; Duong, T.D. Examining the Impact of Greenwashing on Customer Boycott Intentions: The Mediating Role of Green Confusion. *Discov. Sustain.* **2025**, *6*, 460. [\[CrossRef\]](#)
60. Guerreiro, J.; Pacheco, M. How Green Trust, Consumer Brand Engagement and Green Word-of-Mouth Mediate Purchasing Intentions. *Sustainability* **2021**, *13*, 7877. [\[CrossRef\]](#)
61. De Sio, S.; Zamagni, A.; Casu, G.; Gremigni, P. Green Trust as a Mediator in the Relationship between Green Advertising Skepticism, Environmental Knowledge, and Intention to Buy Green Food. *Int. J. Environ. Res. Public Health* **2022**, *19*, 16757. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Alhomaidd, A. Building Trust in Sustainable Journeys: The Interplay Between Green Marketing, Green Brand Trust, and Tourism Purchase Intentions. *Sustainability* **2025**, *17*, 8464. [\[CrossRef\]](#)
63. Khan, K.U.; Atlas, F.; Arshad, M.Z.; Akhtar, S.; Khan, F. Signaling Green: Impact of Green Product Attributes on Consumers Trust and the Mediating Role of Green Marketing. *Front. Psychol.* **2022**, *13*, 790272. [\[CrossRef\]](#) [\[PubMed\]](#)
64. Tan, Z.; Sadiq, B.; Bashir, T.; Mahmood, H.; Rasool, Y. Investigating the Impact of Green Marketing Components on Purchase Intention: The Mediating Role of Brand Image and Brand Trust. *Sustainability* **2022**, *14*, 5939. [\[CrossRef\]](#)
65. Chauhan, S.; Goyal, S. A Meta-Analysis of Antecedents and Consequences of Green Trust. *J. Consum. Mark.* **2024**, *41*, 459–473. [\[CrossRef\]](#)
66. Riva, F.; Magrizos, S.; Rizomyliotis, I.; Uddin, M.R. Beyond the Hype: Deciphering Brand Trust amid Sustainability Skepticism. *Bus. Strategy Environ.* **2024**, *33*, 6491–6506. [\[CrossRef\]](#)
67. Chang, J.-H.; Wang, S.-H.; Weng, H.-C. From Claims to Conversions: Exploring Sustainable Advertising, Brand Trust and Review Dynamics in Online Green Purchases. *Asia Pac. J. Mark. Logist.* **2025**, 1–21. [\[CrossRef\]](#)
68. Ramdhani, E.F.; Pradisti, L. Digital Sustainability Claims and Consumer Skepticism: A Study of Generation Z's Behavioral Responses to Msmes' Green Marketing on Social Media in Purwokerto, Indonesia. *ICSEMA* **2025**, *1*, 1282–1288. [\[CrossRef\]](#)
69. Tikaro, M. Building Trust in Green Brands: A Dual-Path Model of Consumer Perception and Social Media Communication. *SMP* **2025**, *1*, 1–10. [\[CrossRef\]](#)
70. El Ali, A.; Venkatraj, K.P.; Morosoli, S.; Naudts, L.; Helberger, N.; Cesar, P. Transparent AI Disclosure Obligations: Who, What, When, Where, Why, How. In *Proceedings of the CHI EA '24: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 11–16 May 2024*; ACM: New York, NY, USA, 2024; pp. 1–11.
71. Kim, T.W.; Jiang, L.; Duhachek, A.; Lee, H.; Garvey, A. Do You Mind If I Ask You a Personal Question? How AI Service Agents Alter Consumer Self-Disclosure. *J. Serv. Res.* **2022**, *25*, 649–666. [\[CrossRef\]](#)
72. Al-kfairy, M.; Mustafa, D.; Kshetri, N.; Insiew, M.; Alfandi, O. Ethical Challenges and Solutions of Generative AI: An Interdisciplinary Perspective. *Informatics* **2024**, *11*, 58. [\[CrossRef\]](#)
73. Lim, S.; Schmälzle, R. The Effect of Source Disclosure on Evaluation of AI-Generated Messages. *Comput. Hum. Behav. Artif. Hum.* **2024**, *2*, 100058. [\[CrossRef\]](#)
74. Beckert, J.; Koch, T.; Viererbl, B.; Schulz-Knappe, C. The Disclosure Paradox: How Persuasion Knowledge Mediates Disclosure Effects in Sponsored Media Content. *Int. J. Advert.* **2021**, *40*, 1160–1186. [\[CrossRef\]](#)
75. Qiu, X.; Wang, Y.; Zeng, Y.; Cong, R. Artificial Intelligence Disclosure in Cause-Related Marketing: A Persuasion Knowledge Perspective. *JTAER* **2025**, *20*, 193. [\[CrossRef\]](#)
76. Downen, T.; Kim, S.; Lee, L. Algorithm Aversion, Emotions, and Investor Reaction: Does Disclosing the Use of AI Influence Investment Decisions? *Int. J. Account. Inf. Syst.* **2024**, *52*, 100664. [\[CrossRef\]](#)
77. Feng, Y.; Kim, H.J. Decoding the Trust Matrix: Unraveling Key Predictors of Consumer Trust in AI-Generated Personalized Advertising. *J. Interact. Advert.* **2025**, *25*, 123–138. [\[CrossRef\]](#)
78. Waltenrath, A. Consumers' Ambiguous Perceptions of Advertising Disclosures in Influencer Marketing: Disentangling the Effects on Current and Future Social Media Engagement. *Electron. Mark.* **2024**, *34*, 8. [\[CrossRef\]](#)
79. Arango, L.; Singaraju, S.P.; Niininen, O. Consumer Responses to AI-Generated Charitable Giving Ads. *J. Advert.* **2023**, *52*, 486–503. [\[CrossRef\]](#)

80. Wang, Y.; Sauka, K.; Situmeang, F.B.I. Anthropomorphism and Transparency Interplay on Consumer Behaviour in Generative AI-Driven Marketing Communication. *J. Consum. Mark.* **2025**, *42*, 512–536. [[CrossRef](#)]
81. Petty, R.E.; Cacioppo, J.T.; Sedikides, C.; Strathman, A.J. Affect and Persuasion: A Contemporary Perspective. *Am. Behav. Sci.* **1988**, *31*, 355–371. [[CrossRef](#)]
82. Dietvorst, B.J.; Simmons, J.P.; Massey, C. Algorithm Aversion: People Erroneously Avoid Algorithms after Seeing Them Err. *J. Exp. Psychol. Gen.* **2015**, *144*, 114–126. [[CrossRef](#)] [[PubMed](#)]
83. Reich, T.; Kaju, A.; Maglio, S.J. How to Overcome Algorithm Aversion: Learning from Mistakes. *J. Consum. Psychol.* **2023**, *33*, 285–302. [[CrossRef](#)]
84. Friestad, M.; Wright, P. The Persuasion Knowledge Model: How People Cope with Persuasion Attempts. *J. Consum. Res.* **1994**, *21*, 1–31. [[CrossRef](#)] [[PubMed](#)]
85. Kazan, G.; Kocamış, T.U.; Güngör, A. SYMBOLIC OR SUBSTANTIVE? GREENWASHING IN CORPORATE SUSTAINABILITY REPORTS OF TURKISH FIRMS. *Veredas Direito* **2025**, *22*, e223441. [[CrossRef](#)]
86. Gu, V.X.; Matisoff, D.C. Symbolic or Substantive Action: Intent, Effort, and Results. *Bus. Strategy Environ.* **2026**, bse.70545. [[CrossRef](#)]
87. Flanagin, A.J.; Metzger, M.J. Source Credibility. In *The International Encyclopedia of Media Psychology*; Bulck, J., Ed.; Wiley: Hoboken, NJ, USA, 2020; pp. 1–5.
88. Hocevar, K.P.; Metzger, M.; Flanagin, A.J. Source Credibility, Expertise, and Trust in Health and Risk Messaging. In *Oxford Research Encyclopedia of Communication*; Powers, M., Ed.; Oxford University Press: New York, NY, USA, 2017.
89. Sternthal, B.; Phillips, L.W.; Dholakia, R. The Persuasive Effect of Source Credibility: A Situational Analysis. *Public Opin. Q.* **1978**, *42*, 285. [[CrossRef](#)] [[PubMed](#)]
90. Kumkale, G.T.; Albarracín, D.; Seignourel, P.J. The Effects of Source Credibility in the Presence or Absence of Prior Attitudes: Implications for the Design of Persuasive Communication Campaigns. *J. Appl. Soc. Psychol.* **2010**, *40*, 1325–1356. [[CrossRef](#)] [[PubMed](#)]
91. Schmuck, D.; Matthes, J.; Naderer, B.; Beaufort, M. The Effects of Environmental Brand Attributes and Nature Imagery in Green Advertising. *Environ. Commun.* **2018**, *12*, 414–429. [[CrossRef](#)]
92. Chen, Y.-S. The Drivers of Green Brand Equity: Green Brand Image, Green Satisfaction, and Green Trust. *J. Bus. Ethics* **2010**, *93*, 307–319. [[CrossRef](#)]
93. Iovino, R.; Testa, F.; Iraldo, F. Do Consumers Understand What Different Green Claims Actually Mean? An Experimental Approach in Italy. *J. Advert.* **2024**, *53*, 200–214. [[CrossRef](#)]
94. Ha, M.; Ngan, V.T.K.; Nguyen, P.N.D. Greenwash and Green Brand Equity: The Mediating Role of Green Brand Image, Green Satisfaction and Green Trust and the Moderating Role of Information and Knowledge. *Bus. Ethics Environ. Responsib.* **2022**, *31*, 904–922. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.