AI-GENERATED ADVERTISING

CONSUMER PERCEPTIONS, TRUST, AND ETHICAL GOVERNANCE

Preliminary Insights from the Turkish Market

CHAPTERS

EDITED BY

ASSO. PROF. DR.

Mehmet Özer DEMİR

Alanya Alaaddin Keykubat University Press : 008

ISBN: 978-605-74581-8-6

AI-Generated Advertising: Consumer Attitudes, Trust, and Ethical Governance – Preliminary Insights from the Turkish Market

INTRODUCTION

The rapid proliferation of generative artificial intelligence (AI) has fundamentally transformed the advertising landscape. From algorithmically written copy and synthetically generated visuals to personalized product recommendations and virtual influencers, AI now plays a central role in how brands create, target, and deliver persuasive messages to consumers. Industry projections suggest that by 2026, a substantial portion of digital advertising content will involve some form of AI generation (Jesus et al., 2026), driven by efficiency gains, creative augmentation, and the promise of hyper-personalization at scale. Yet this technological shift has introduced profound questions that extend far beyond marketing effectiveness. How do consumers evaluate advertisements they know – or suspect – were generated by a machine? Does the use of AI in advertising enhance or erode trust in the brands and manufacturers behind those ads? Can consumers calibrate their trust appropriately, neither over-relying on nor dismissing AI-generated persuasive content? And what ethical governance mechanisms are needed to ensure that AI-driven marketing aligns with consumer expectations, legal requirements, and societal values?

This book addresses these questions by bringing together five empirical studies that examine consumer responses to AI-generated advertising from multiple psychological, relational, and governance perspectives. Rather than treating AI-generated ads as a simple variation of traditional advertising, the volume conceptualizes algorithmic authorship as a trigger for multilevel evaluative processes – affecting brand attitudes, manufacturer perceptions, corporate reputation, affective trust, trust calibration, and ethical judgments. The chapters collectively advance a human-centered understanding of AI in marketing, emphasizing that consumers do not respond to AI content in isolation; rather, they evaluate the technological source, the organizational entities responsible for it, the transparency of its use, and the ethical framework within which it operates.

The Con□eptual Thread A□ross Chapters

Although each chapter tests a distinct theoretical model or develops a novel measurement instrument, the book is organized around a coherent conceptual thread: how consumers form, adjust, and govern their trust in AI-generated advertising across multiple levels of analysis.

Chapter 1 establishes the foundational cascade of consumer evaluations. Focusing on the relationships among brand attitude, attitude toward the manufacturer, and corporate reputation, the chapter demonstrates that favorable attitudes toward AI-generated brand advertisements positively influence manufacturer perceptions, which in turn shape broader corporate reputation. This hierarchical model positions manufacturer attitude as a key mediator, showing that the effects of AI advertising ripple upward from specific brand encounters to general organizational judgments.

Chapter 2 reverses the causal lens. Rather than examining how brand attitudes affect manufacturer perceptions, it investigates whether attitude toward the manufacturer (the company that adopts AI advertising) shapes brand attitude and, through it, affective brand trust – the emotional confidence that a brand cares about its customers and acts with integrity. The findings support full mediation: consumers who respect and trust manufacturers that use AI are more likely to develop positive brand attitudes, which then generate emotional trust. Together, Chapters 1 and 2 paint a reciprocal picture: brand and manufacturer evaluations influence each other, and both ultimately affect higher-order trust and reputation.

Chapter 3 introduces a concept imported from human-automation interaction – trust calibration – to the advertising domain. Recognizing that measuring trust levels is insufficient, the chapter develops the Trust Calibration toward AI Advertising (TCAIA) scale, along with a composite Calibration Index that captures the alignment between consumers' trust in AI ads and their perceived performance of those ads. Although the proposed four-dimensional structure did not achieve satisfactory confirmatory fit, the average Calibration Index of 0.906 suggests that Turkish consumers, on average, are reasonably well-calibrated. The chapter also reveals that perceived performance may be the dominant driver of trust calibration, with transparency and ethics playing supportive roles.

Chapter 4 shifts from trust calibration to the specific role of AI explanations. The Artificial Intelligence Explanation Effects Scale (AIDES) operationalizes four dimensions of how

consumers respond to explanations provided by AI systems: Perceived Transparency and Understanding (PTU), Perceived Credibility and Honesty (PCH), Ethical Fairness and Accountability (EFA), and Trust Calibration and Behavioral Response (TCB). While the measurement model showed mixed fit, the chapter provides a structured framework for empirically assessing the psychological effects of explainable AI (XAI) in marketing contexts, moving beyond purely technical evaluations of algorithmic transparency.

Chapter 5 addresses the broadest question: What constitutes ethical governance in AI-driven marketing? The Ethical Governance in AI Marketing (EG-AIM) scale operationalizes five dimensions – Accountability & Responsibility (AR), Transparency & Explainability (TE), Ethical Oversight & Compliance (EOC), Stakeholder Inclusiveness & Feedback (SIF), and Monitoring & Risk Management (MRM). The scale demonstrates satisfactory internal consistency and HTMT-based discriminant validity, although model fit indices fell below conventional thresholds and the Fornell-Larcker criterion was not fully satisfied. The chapter offers a theoretically grounded foundation for future research while cautioning against the uncritical use of the scale in its current form.

The Turkish Context and Its Limitations

All five chapters draw on data collected from Turkish consumers, with sample sizes ranging from 270 to 435 participants. Turkey represents a valuable context for studying AI-generated advertising: it has high digital and social media penetration, a young and digitally active population, and growing exposure to AI-generated content on platforms such as Instagram, TikTok, YouTube, and e-commerce sites. However, the Turkish context also imposes important limitations. Findings may not generalize directly to consumers in other cultural, regulatory, or economic environments where trust in AI, attitudes toward automation, and expectations of transparency differ. The book does not claim cross-cultural universality; rather, it offers a detailed empirical account from a specific market, inviting replication and extension elsewhere.

Who Should Read This Book

This volume is intended for several audiences. Academic researchers in marketing, advertising, consumer psychology, human-computer interaction, and AI ethics will find theoretically grounded models, preliminary scales, and a comprehensive review of the literature on trust, transparency, and governance in AI advertising. Marketing practitioners and brand managers will gain insights into how consumers evaluate AI-generated ads, why transparency matters

(even when it temporarily reduces persuasion), and how manufacturer reputation and corporate governance shape brand trust. Regulators and policymakers will find evidence relevant to disclosure requirements, ethical auditing, and consumer protection in algorithmic marketing. Finally, AI developers and advertising technology firms will encounter user-centered perspectives that can inform the design of explainable, accountable, and trustworthy AI systems for marketing applications.

Organization of the Book

Following this introduction, the book presents five self-contained chapters, each structured as a full empirical study with its own abstract, literature review, methodology, results, discussion, and references. Chapter 1 examines the cascade from brand attitude to manufacturer attitude to corporate reputation. Chapter 2 reverses the causal ordering to test manufacturer attitude → brand attitude → affective brand trust. Chapter 3 develops the Trust Calibration toward AI Advertising (TCAIA) scale and the Calibration Index. Chapter 4 introduces the Artificial Intelligence Explanation Effects Scale (AIDES). Chapter 5 presents the Ethical Governance in AI Marketing (EG-AIM) scale. A concluding chapter synthesizes key findings across studies, identifies patterns and unresolved questions, and proposes an agenda for future research on trust, calibration, and ethical governance in AI-generated advertising.

A Final Word Before Reading

The integration of generative AI into advertising is not a distant future – it is the present. As consumers encounter algorithmically authored content daily, understanding their psychological, relational, and ethical responses becomes a matter not only of academic inquiry but also of consumer protection, market integrity, and democratic accountability. This book does not offer final answers. It offers empirical foundations, conceptual frameworks, and preliminary measurement tools. It invites scholars, practitioners, and regulators to build on this work, to refine its instruments, to test its models across cultures, and to engage in the collective effort of shaping an AI-driven marketing ecosystem that is not only effective but also trustworthy, transparent, and fair.

CHAPTER 1

Attitude Toward the Manufacturer, Brand Attitude, and Affective Brand Trust in AI-Generated Advertising: A Mediation Study

Abdullah TÖRE

Abstract

This study investigates the relationship between consumer attitudes toward manufacturers that use AI-generated advertising, brand attitude, and affective brand trust, specifically testing whether brand attitude mediates the effect of manufacturer attitude on affective brand trust. Data were collected from Turkish consumers via an online survey using scales for manufacturer attitude, brand attitude, and affective brand trust. Confirmatory factor analysis confirmed good measurement model fit, and reliability was high. Structural equation modelling with bootstrapping revealed a significant indirect effect of manufacturer attitude on affective brand trust through brand attitude. Manufacturer attitude strongly predicted brand attitude, and brand attitude strongly predicted affective brand trust. The model explained a substantial portion of the variance in both brand attitude and affective brand trust. This study is the first to demonstrate that brand attitude fully mediates the relationship between manufacturer attitude and affective brand trust in the context of AI-generated advertising, indicating that consumers' trust in brands using AI is shaped not directly by their perception of the manufacturer but through their overall evaluation of the brand. Practical implications for AI advertising strategies are discussed.

1. Introduction

The rapid proliferation of artificial intelligence (AI) in the advertising ecosystem has fundamentally transformed how brands create, target, and deliver persuasive messages. From programmatic ad placements and personalized product recommendations to fully generative content produced by large language models and image synthesizers (e.g., ChatGPT, DALL-E, Midjourney), AI now plays a central role in shaping consumer–brand interactions. While the efficiency and scalability benefits of AI-generated advertising are widely acknowledged, the psychological mechanisms through which consumers evaluate brands that adopt such technologies remain under-explored. In particular, three interrelated consumer judgments are critical for understanding long-term brand relationships: attitude toward the brand (the overall evaluative disposition), affective brand trust (the emotional confidence in a brand's benevolence and reliability), and attitude toward the manufacturer (the respect, admiration, and trust felt toward the company behind the brand). Despite the growing body of research on AI in marketing, no previous study has systematically examined how these three constructs interrelate specifically in the context of AI-generated advertising.

1.1 The Rise of AI-Generated Advertising and Its Psychological Consequences

AI-generated advertisements are no longer a futuristic concept; they are a present reality. Brands increasingly rely on AI to write ad copy, generate personalized visuals, and even create synthetic influencers. This shift carries significant implications for consumer psychology. Unlike traditional human-crafted ads, AI-generated content raises unique concerns about authenticity, transparency, and emotional resonance. Consumers may question whether an AI can truly understand their needs, whether the brand is being manipulative, and whether the manufacturer behind the AI can be held accountable for errors or deceptive content (Luo et al., 2019; Shin, 2021). Consequently, the formation of brand attitudes and trust in AI-adopting brands may follow different pathways compared to conventional advertising.

1.2 Attitude Toward the Brand in AI Advertising

Attitude toward the brand (MY) refers to a consumer's overall favorable or unfavorable evaluation of a brand. It is a central predictor of purchase intentions, loyalty, and brand advocacy. Extensive research has shown that positive ad evaluations directly enhance brand

attitude (Çakır, 2006), that ad involvement influences brand attitude through ad attitude (Başaran & Yıldız, 2022), and that celebrity endorser–brand congruence shapes brand evaluations (Yıldırım et al., 2014). In the context of AI-generated advertising, brand attitude may be influenced by perceptions of the AI's competence (e.g., ad relevance, accuracy) and the perceived transparency of AI use. Consumers who find AI-generated ads useful, creative, and honest are likely to develop more positive brand attitudes. However, if the AI is perceived as intrusive, error-prone, or lacking warmth, brand attitude may suffer.

1.3 Affective Brand Trust as an Emotional Outcome

Affective brand trust (DM) goes beyond cognitive evaluations of reliability; it captures the emotional bond between consumer and brand – the feeling that the brand cares about its customers, acts with integrity, and would respond compassionately in times of need. Yavuz and Ünal (2016) demonstrated that affective trust significantly influences purchase behavior, often more strongly than cognitive trust. Karayel Bilbil and Orha Hazar (2024) linked brand trust to corporate reputation and loyalty, while Yapraklı et al. (2016) found that affective trust drives brand evangelism. In AI advertising, building affective trust is particularly challenging because consumers may perceive AI as cold or impersonal. Brands that successfully communicate warmth, responsiveness, and ethical responsibility through their AI-generated ads can overcome this barrier and foster genuine emotional trust.

1.4 Attitude Toward the Manufacturer as a Distal Antecedent

While brand attitude focuses on the brand itself, attitude toward the manufacturer (UY) concerns the company that produces the brand. This distinction is especially relevant for AI-generated advertising because the use of AI often reflects strategic decisions made at the corporate level. A manufacturer that invests in advanced AI for advertising may be seen as innovative, forward-thinking, and competent – or alternatively as impersonal, profit-driven, and intrusive. Prior research has established that firm credibility positively affects brand equity (Başgöze et al., 2021), and that consumers' general evaluations of a manufacturer shape their perceptions of individual brands (Aydın, 2021; Şener, 2023; Ülker, 2021). Thus, we propose that manufacturer attitude serves as a distal antecedent that influences brand attitude, which in turn shapes affective brand trust.

1.5 The Mediating Role of Brand Attitude

Why should brand attitude mediate the relationship between manufacturer attitude and affective brand trust? The persuasion knowledge model (Friestad & Wright, 1994) suggests that consumers use heuristic cues (e.g., corporate reputation, AI adoption) to form initial judgments, which then influence more specific evaluations. In the context of AI advertising, a positive attitude toward the manufacturer (e.g., "this company is trustworthy and innovative") may lead consumers to give the brand the benefit of the doubt, resulting in a more favorable brand attitude. That positive brand attitude, in turn, generates affective trust because consumers feel emotionally connected to a brand they already evaluate positively. Conversely, if the manufacturer is viewed negatively, brand attitude may suffer, and emotional trust is unlikely to develop. This causal chain implies that manufacturer attitude does not directly translate into affective brand trust without the intermediation of brand-specific evaluations.

1.6 The Present Study and Hypotheses

Despite the theoretical plausibility of this mediation model, no empirical study has tested it in the context of AI-generated advertising. Most existing research treats brand attitude, manufacturer attitude, and affective brand trust as separate outcomes, ignoring their sequential relationships. This study fills that gap by testing a structural model where:

H1: Manufacturer attitude (UY) is positively associated with brand attitude (MY).

H2: Brand attitude (MY) is positively associated with affective brand trust (DM).

H3: Brand attitude mediates the relationship between manufacturer attitude and affective brand trust (indirect effect UY → MY → DM is significant).

By validating this model, we provide both theoretical insights into the psychological mechanisms underlying consumer responses to AI-generated advertising and practical guidance for marketers seeking to build enduring brand relationships in an era of algorithmic persuasion.

2. Literature Review

2.1 Attitude Toward the Manufacturer (UY)

Consumers form attitudes not only toward brands but also toward the manufacturers behind them. Manufacturer attitude encompasses respect, admiration, trust, and overall evaluation of the company. Başgöze et al. (2021) found that firm credibility positively affects brand equity dimensions. Aydın (2021) argued that brand attitude reflects overall evaluations of the manufacturer. Positive brand experiences translate into favorable manufacturer perceptions (Şener, 2023; Ülker, 2021). In the context of AI, manufacturers that adopt AI in advertising may be seen as innovative or, conversely, as impersonal. Therefore, measuring manufacturer attitude toward AI-using companies is essential.

2.2 Brand Attitude (MY)

Brand attitude refers to consumers' overall evaluative judgment of a brand, ranging from negative to positive. Çakır (2006) found that positive ad evaluations directly enhance brand attitude. Başaran and Yıldız (2022) demonstrated that ad attitude mediates the effect of ad involvement on brand attitude. Celebrity endorser–brand congruence also shapes brand attitude (Yıldırım et al., 2014). Native advertising (Dönmez, 2020) and social media brand awareness (Şener, 2023) further strengthen brand attitudes. In AI advertising, brand attitude may be influenced by perceptions of AI competence and transparency.

2.3 Affective Brand Trust (DM)

Affective brand trust refers to the emotional confidence a consumer has in a brand's benevolence, integrity, and responsiveness. Yavuz and Ünal (2016) distinguished cognitive and affective trust, showing that both influence purchase behavior. Karayel Bilbil and Orha Hazar (2024) linked brand trust to corporate reputation and loyalty. Kosat (2024) found that brand trust and brand image enhance customer satisfaction, with psychological well-being as a moderator. Affective trust also drives brand evangelism (Yapraklı et al., 2016). In AI advertising, consumers may develop affective trust if they perceive the brand as caring and responsive, despite the use of automated systems.

2.4 Hypotheses and Mediation Model

Based on theory, we expect that a positive attitude toward the manufacturer will enhance brand attitude, which in turn will increase affective brand trust. Brand attitude is positioned as a mediator because it represents the more immediate, brand-specific evaluation that directly shapes emotional trust. Thus, H1, H2, and H3 are formulated as above.

3. Method

3.1 Participants and Data Collection

An online survey was conducted among Turkish consumers. After listwise deletion of incomplete responses, 270 valid cases remained. Demographic characteristics are; mean age 29.63 years (SD = 10.13, range 12–68), gender: 47.8% female, 43.9% male, 8.0% preferred not to specify, education levels ranged from primary to postgraduate. All participants reported having encountered online advertising.

3.2 Measures

All items were rated on 7-point scales. The scales were adapted to the context of brands that use AI-generated advertisements.

Manufacturer Attitude (UY) – 4 items (1 = Strongly Disagree, 7 = Strongly Agree)

UYT1: “Compared to brands that do not use AI in advertising, I hold companies that use AI-generated ads in high regard.”

UYT2: “Companies that use AI-generated ads deserve my respect.”

UYT3: “I can trust companies that use AI-generated ads.”

UYT4: “I admire companies that use AI-generated ads.”

Brand Attitude (MY) – 5 semantic differential items (7-point scale)

Respondents evaluated the brand using: Bad/Good, Unattractive/Attractive, Unpleasant/Pleasant, Negative/Positive, Disliked/Liked.

Affective Brand Trust (DM) – 4 items (1 = Strongly Disagree, 7 = Strongly Agree)

DM1 (reverse-coded): “Brands that use AI-generated ads are only interested in selling products.”

DM2: “Brands that use AI-generated ads show a warm and caring attitude toward their consumers.”

DM3: “If I could no longer use a brand that uses AI-generated ads, I would feel a personal loss.”

DM4: “If I have a problem with a product, I feel that a brand using AI-generated ads would respond in a caring manner.”

3.3 Procedure

Participants completed the online questionnaire anonymously. The survey included the 13 items plus demographic questions. DM1 was reverse-coded before analysis (8 – original score). The survey took approximately 5 minutes.

3.4 Data Analysis

Descriptive statistics and composite reliability (CR) were computed. Confirmatory factor analysis (CFA) was performed using AMOS with maximum likelihood estimation to assess the measurement model. Model fit was evaluated using χ^2/df, CFI, SRMR, RMSEA, and PClose (Hu & Bentler, 1999). Convergent validity was assessed with average variance extracted (AVE). Discriminant validity was evaluated using HTMT (Henseler et al., 2015) and the Fornell-Larcker criterion.

Structural equation modelling (SEM) tested the mediation hypothesis. Bootstrapping with 500 samples was used to obtain 95% confidence intervals for the indirect effect (UY → MY → DM). The model included paths from UY to MY and MY to DM, with no direct path from UY to DM (full mediation). Variance explained (R^2) for MY and DM was calculated.

4. Results

4.1 Descriptive Statistics

Table 1 presents the means and standard deviations for all items.

Table 1. Descriptive Statistics (N = 270)

Item	Mean	SD
DM1_rev	4.22	1.889
DM2	3.95	1.659
DM3	3.80	1.822
DM4	4.10	1.678
UYT1	4.03	1.803
UYT2	3.92	1.680
UYT3	3.93	1.682
UYT4	3.86	1.778
MY1 (Bad–Good)	4.22	1.512
MY2 (Unattractive–Attractive)	4.36	1.543
MY3 (Unpleasant–Pleasant)	4.26	1.506
MY4 (Negative–Positive)	4.16	1.512
MY5 (Disliked–Liked)	4.20	1.455

4.2 Measurement Model: Reliability and Validity

Composite reliability (CR) from CFA was DM = 0.795, UY = 0.895, MY = 0.903, all exceeding 0.70. Convergent validity; AVE values were DM = 0.507, UY = 0.680, MY = 0.652 (all ≥ 0.50, meeting the threshold). Discriminant validity; HTMT values were all below 0.85 (DM-UY = 0.700, DM-MY = 0.517, UY-MY = 0.600). However, the Fornell-Larcker criterion showed that √AVE for DM (0.712) was less than its correlation with UY (0.845), indicating potential overlap. This is addressed in the discussion.

Confirmatory factor analysis (CFA) of the three-factor model showed acceptable to good fit: $\chi^2(62) = 157.96$, $\chi^2/df = 2.548$, CFI = 0.956, SRMR = 0.047, RMSEA = 0.076. The combination of CFI > 0.95 and SRMR < 0.08 supports good measurement model fit (Hu & Bentler, 1999).

Table 2. Standardized Factor Loadings (CFA)

Item	Factor	Loading
DM1_rev	DM	0.412
DM2	DM	0.834
DM3	DM	0.769
DM4	DM	0.758
UYT1	UY	0.780
UYT2	UY	0.874
UYT3	UY	0.840
UYT4	UY	0.802
MYT1	MY	0.727
MYT2	MY	0.766
MYT3	MY	0.825
MYT4	MY	0.852
MYT5	MY	0.841

Note. All loadings are significant at $p < 0.001$.

4.3 Structural Model and Mediation

The structural model (full mediation: UY → MY → DM) was estimated. Standardized path coefficients are shown in Figure 1.

UY → MY: β = 0.701 (p < 0.001)

MY → DM: β = 0.667 (p < 0.001)

The variance explained (R^2) was 0.491 for MY (49.1% of brand attitude variance explained by manufacturer attitude) and 0.445 for DM (44.5% of affective brand trust variance explained by brand attitude).

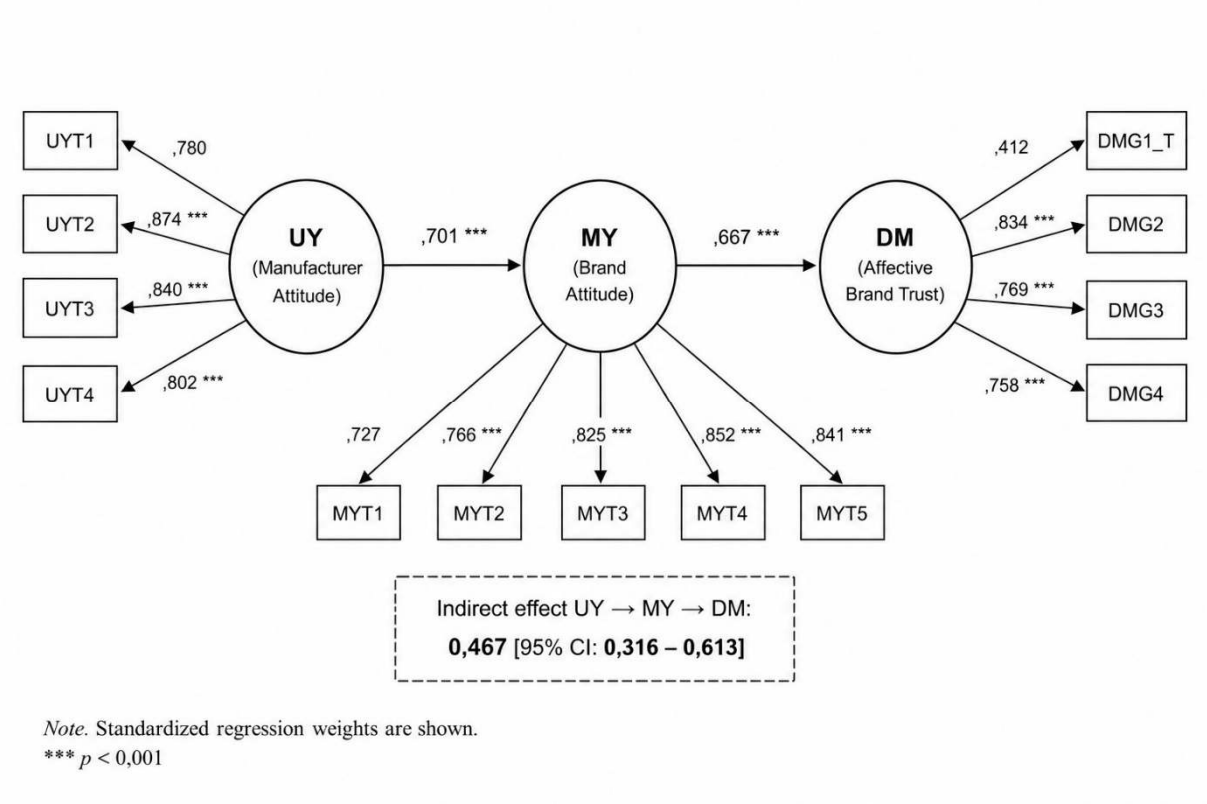

Figure 1. Path Diagram with Standardized Coefficients (all p < 0.001)

Indirect effect: The standardized indirect effect of UY on DM via MY was 0.467. The bootstrap 95% confidence interval (based on 500 samples) was [0.316, 0.613], which does not include zero, confirming a significant indirect effect (p = 0.004). Thus, H3 is supported.

Model fit for the structural model: $\chi^2(63) = 264.08$, $\chi^2/df = 4.192$, CFI = 0.908, RMSEA = 0.109. While the CFI is acceptable (>0.90), the RMSEA is slightly above the recommended 0.08 threshold, indicating room for improvement. The significant indirect effect and the good fit of the measurement model support the substantive conclusion.

UY (Manufacturer Attitude) → MY (Brand Attitude): β = 0.701

MY → DM (Affective Brand Trust): β = 0.667

Indirect effect UY → MY → DM: 0.467 [95% CI: 0.316, 0.613]

5. Discussion

5.1 Summary of Findings

This study tested whether brand attitude mediates the relationship between manufacturer attitude and affective brand trust in the context of AI-generated advertising. The results supported full mediation; manufacturer attitude strongly predicted brand attitude ($\beta = 0.701$), which in turn strongly predicted affective brand trust ($\beta = 0.667$). The indirect effect was significant (0.467), and the bootstrap confidence interval did not contain zero. The model explained nearly half the variance in both brand attitude (49%) and affective brand trust (45%).

5.2 Theoretical Implications

The findings contribute to the literature in several ways. First, they extend the persuasion knowledge model (Friestad & Wright, 1994) to AI advertising; consumers' evaluation of the manufacturer (a distal cue) influences brand attitude (a proximal evaluation), which then determines emotional trust. Second, the strong direct effects suggest that when consumers perceive a manufacturer as respectable and trustworthy, they develop a more positive brand attitude, which then fosters affective brand trust. Third, the full mediation implies that manufacturer attitude does not directly affect affective brand trust without the intermediation of brand attitude. This highlights the centrality of brand-specific evaluations.

5.3 Practical Implications

For marketers using AI in advertising, the results indicate that building a positive attitude toward the manufacturer is not sufficient to generate affective brand trust. Instead, the brand itself must be evaluated positively. Therefore, AI advertising campaigns should focus on enhancing brand-specific attributes (e.g., warmth, competence, relevance) rather than merely highlighting the manufacturer's technological sophistication. Transparency about AI use, ethical governance, and responsive customer service can improve brand attitude, which then translates into emotional trust.

5.4 Limitations and Future Research

Limitations include: (1) cross-sectional design, precluding causal inferences; (2) convenience sample of Turkish consumers, limiting generalizability; (3) common method bias; (4) discriminant validity concerns between DM and UY (high correlation). Future research should use experimental designs (e.g., manipulate manufacturer reputation), test the model in other cultures, include behavioral outcomes (purchase intention, loyalty), and explore potential moderators (e.g., AI literacy, need for cognition). Additionally, the structural model fit was not excellent; future studies could test alternative models (e.g., partial mediation) or refine the measurement of affective brand trust.

5.5 Conclusion

This study demonstrates that brand attitude fully mediates the effect of manufacturer attitude on affective brand trust in the context of AI-generated advertising. Consumers who respect and trust manufacturers that use AI are more likely to develop positive brand attitudes, which then generate emotional trust in the brand. Marketers should therefore prioritize building brand-specific positive evaluations through transparent, caring, and relevant AI advertising, rather than relying solely on manufacturer reputation.

References

Aydın, E. B. (2021). Duyum, algı ve markalar: Tüketici tutumlarına yönelik üç boyutlu bir değerlendirme. MANAS Sosyal Araştırmalar Dergisi, 10(4), 2528–2544.

Başaran, Ü., & Yıldız, M. (2022). Reklam ilgilenimi, reklama yönelik tutum ve satın alma niyeti arasındaki etkilerin analizi: Marka tutumunun aracılık rolü. Turkish Review of Communication Studies, 40, 173–195.

Başgöze, P., Özkan Tektaş, Ö., & Eryiğit, C. (2021). Firma kredibilitesi ve reklama yönelik tutumun marka denkliği boyutları üzerindeki etkilerinin incelenmesi. Verimlilik Dergisi, 2021(1), 49–60.

Çakır, V. (2006). Reklamların beğenilmesinin tüketicilerin marka tutumlarına etkisi. Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 15, 663–687.

Dinçer, G., & Yıldız, E. (2024). Marka deneyimi ve ilişki kalitesinin marka rezonansı üzerine etkisi, marka tutumunun aracı rolü. İnönü Üniversitesi Uluslararası Sosyal Bilimler Dergisi, 13(2), 441–465.

Dönmez, M. S. (2020). Doğal reklamların marka tutumu bağlamında irdelenmesi. Süleyman Demirel Üniversitesi Visionary Journal, 11(27), 514–526.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, 18(1), 39–50.

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

Gaskin, J., Lim, J., & Steed, J. (2022). Model Fit Measures (AMOS Plugin). Gaskination's StatWiki.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. Journal of the Academy of Marketing Science, 43(1), 115–135.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6(1), 1–55.

Karayel Bilbil, E., & Orha Hazar, S. (2024). Kurumsal itibar, marka güveni ve marka sadakati kavramlarına yönelik bir araştırma. Selçuk İletişim, 17(1), 100–131.

Kosat, A. (2024). Marka güveni ve marka imajının müşterilerin memnuniyetlerine etkileri: Psikolojik iyi oluşun düzenleyici rolü. Doğuş Üniversitesi Dergisi, 25(2), 273–294.

Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. Marketing Science, 38(6), 937–947.

Şener, B. Ç. (2023). Sosyal medyada marka bilinirliği: Tüketicilerin tutumları ve niyetleri üzerine bir araştırma. Kastamonu İletişim Araştırmaları Dergisi, 11, 76–99.

Shin, D. (2021). The effects of explainability and causability on perception, trust, and adoption of explainable artificial intelligence. International Journal of Human-Computer Studies, 146, 102551.

Ülker, Y. (2021). Ünlüye olan tutum ile satın alma niyeti arasında ilişkide tüketicilerin markaya yönelik tutumunun aracılık rolü ile incelenmesi. Gaziantep Üniversitesi Sosyal Bilimler Dergisi, 20(2), 506–518.

Yapraklı, T. Ş., Keser, E., & Ünalan, M. (2016). Marka güveni ve marka özdeşleşmesinin marka evangelizmi üzerindeki etkisi. Uluslararası İktisadi ve İdari İncelemeler Dergisi, 32, 32–55.

Yavuz, E., & Ünal, S. (2016). Bilişsel ve duygusal marka güveninin satın alma davranışı üzerindeki etkileri. Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 22(3), 401–430.

Yıldırım, M. Y., Karataş Boztaş, R., & Temizkan, M. (2014). Reklamlarda kullanılan ünlü ve marka arasındaki uyumun tüketicinin markaya karşı tutumuna etkisi. Bitlis Eren Üniversitesi Sosyal Bilimler Dergisi, 3(1), 1–20.

Appendix – Turkish Questionnaire Items

Duygusal Marka Güveni (DM)

Yapay zeka ile üretilmiş reklamlar kullanan markalar yalnızca ürün satmakla ilgileniyor. (Ters kodlandı)

Yapay zeka ile üretilmiş reklamlar kullanan markalar tüketicisine karşı sıcak ve ilgili bir tutum sergiler.

Yapay zeka ile üretilmiş reklamlar kullanan bir markayı artık kullanamayacak olsam kişisel bir kayıp hissederdim.

Ürünle ilgili bir sorun yaşarsam, yapay zeka ile üretilmiş reklamlar kullanan markanın ilgili bir şekilde tepki vereceğini hissederim.

Üretkenliğe Yönelik Tutum (UY)

Reklamlarında yapay zeka kullanmayan markalarla karşılaştırıldığında, reklamlarında yapay zeka kullanan şirketleri yüksek bir saygıyla değerlendiririm.

Reklamlarında yapay zeka kullanan şirketler saygımı hak ediyor.

Reklamlarında yapay zeka kullanan şirketlere güvenebilirim.

Reklamlarında yapay zeka kullanan şirketlere hayranlık duyarım.

Markaya Yönelik Tutum (MY) – 7'li semantik farklılık

Kötü / İyi

Çekiciliği olmayan / Çekici

Hoş olmayan / Hoş

Olumsuz / Olumlu

Sevilmeyen / Sevilen

CHAPTER II

Consumer Responses to AI-Generated Advertising: Corporate Reputation, Attitude toward the Brand, and Attitude toward the Manufacturer

Beyza AVSALLI YAMAN

Abstract

The rapid proliferation of generative artificial intelligence (AI) in advertising has fundamentally transformed the way brands communicate with consumers, introducing algorithmic authorship into persuasive communication. This transformation raises important questions about how consumers evaluate AI-generated advertisements and the organizational entities behind them. This study develops a conceptual framework to examine how AI-generated advertising influences consumer evaluations across multiple organizational levels. Specifically, we propose that AI disclosure shapes attitudes toward the brand, which in turn influence attitudes toward the manufacturer and subsequently corporate reputation. The model specifies that attitude toward the manufacturer mediates the relationship between brand attitude and corporate reputation. By integrating these relationships, this study advances a multilevel evaluation framework for understanding consumer responses to AI-generated advertising and provides theoretical guidance for both academic research and managerial practice.

1. Introduction

The rapid proliferation of generative artificial intelligence (AI) has fundamentally transformed the marketing communications landscape, with AI-generated advertising emerging as a dominant force in brand-consumer interactions. Major corporations across industries are increasingly deploying generative AI tools to produce advertising content at scale, leveraging capabilities ranging from automated copywriting and image generation to personalized video advertisements (Chen et al., 2024; Gölgeli, 2025). Industry projections indicate that by 2026, a substantial portion of digital advertising content will involve some form of AI generation, reflecting both efficiency gains and creative augmentation possibilities (Jesus et al., 2026). This technological shift has prompted regulatory responses, including Meta's 2024 policy requiring advertisers to disclose AI-generated content, signaling growing institutional recognition of AI's distinctive role in persuasive communication (Chen et al., 2024). Beyond industry implications, AI-generated advertising raises profound societal questions about authenticity, manipulation, and the nature of creative expression, making it a timely and consequential phenomenon for scholarly investigations.

Understanding consumer responses to AI-generated advertising requires integrating multiple theoretical perspectives that illuminate how individuals evaluate technologically mediated persuasive messages. Source Credibility Theory (Deryl et al., 2023) provides foundational insights by suggesting that the perceived trustworthiness and expertise of a communication source, whether human or algorithmic, determines message acceptance. In AI-generated contexts, consumers must assess not only the message content but also the credibility of the technological source behind it. The Persuasion Knowledge Model (Friestad & Wright, 1994) further informs our understanding by proposing that awareness of persuasive intent activates cognitive defenses; when consumers learn that an advertisement was AI-generated, they may scrutinize the message more critically and question the underlying motives. Mind Perception Theory (Burrow & Gutentag, 2025; Gray et al., 2007) explains differential responses to AI by highlighting that consumers attribute distinct mental capacities to algorithms, perceiving AI as competent in task-oriented domains but lacking emotional warmth, which shapes expectations for appropriate AI roles in advertising. Social Cognitive Theory (Bandura, 1986; Chen et al., 2024) contributes to the concept of self-efficacy, suggesting that consumers' confidence in evaluating and interacting with AI-generated content moderates their responses. Attribution Theory (Kelley, 1973) explains how consumers infer motives from organizational actions, with

these attributions shaping evaluations of both the manufacturer and brand. Together, these theoretical lenses establish that AI-generated advertising triggers multilevel evaluations that extend beyond message effectiveness to encompass judgments about the corporate, manufacturer, and brand entities involved.

Despite the growing scholarly attention to AI-generated advertising, prior research has largely examined isolated outcomes at a single level of analysis. Existing studies have predominantly focused on ad-level responses, such as persuasiveness, perceived creativity, and emotional appeal (Exner et al., 2025; Gu et al., 2024), or have separately investigated consumer attitudes toward either brands (Wu et al., 2025; Yin et al., 2024) or corporate entities (Jung et al., 2025). However, little is known about how AI-generated advertising simultaneously shapes the interconnected hierarchy of consumer evaluations, specifically the relationships among corporate reputation, attitude toward the manufacturer, and attitude toward the brand. The distinction between manufacturers and corporations, while conceptually meaningful in multi-entity corporate structures, remains underexplored in AI advertising contexts. Furthermore, although transparency through AI disclosure has been identified as strategically important (Chen et al., 2024), the mechanisms through which disclosure cascades through corporate, manufacturer, and brand-level evaluations have not been theoretically integrated or empirically tested. This conceptual gap limits our understanding of how AI-generated advertising influences the full spectrum of consumer-firm relationships.

Accordingly, this study investigates the influence of AI-generated advertising on consumer evaluations across multiple organizational levels. Specifically, we examine the sequential relationships through which AI disclosure influences attitudes toward the brand, which in turn shape attitudes toward the manufacturer and subsequently corporate reputation. We conceptualize corporate reputation as a higher-order, cumulative evaluative judgment that emerges from brand- and manufacturer-level attitudes rather than as a preexisting interpretive framework. In this process, manufacturer attitude mediates the transfer of brand-level evaluations to broader organizational perceptions. By testing these relationships, we develop and empirically validate an integrated model of multilevel consumer responses to AI-generated advertising.

This study advances knowledge in several ways. First, it extends existing advertising research by conceptualizing AI-generated advertising as a trigger for multilevel evaluative processes, rather than merely a message-type variation. By integrating corporate reputation, manufacturer attitude, and brand attitude within a single framework, we provide a more comprehensive account of how consumers process and respond to algorithmically authored content. Second, we contribute to the transparency literature by theorizing the specific mechanisms through which AI disclosure cascades through organizational levels, addressing calls for research on how disclosure operates beyond immediate persuasion outcomes (Chen et al., 2024). Third, we refine the theoretical understanding of source effects in AI contexts by distinguishing between different organizational entities (manufacturers versus corporations) that may evoke distinct consumer judgments. Fourth, we respond to recent calls for research examining the interplay between technological agency and corporate responsibility (Jesus et al., 2026), demonstrating how corporate reputation functions as both an outcome of AI communication practices and a frame that shapes subsequent evaluations.

The findings of this study offer actionable insights for multiple stakeholders. For marketing managers and advertisers, understanding how AI disclosure influences corporate reputation informs strategic decisions about transparency, specifically, when and how to communicate AI involvement in advertising creation. The multilevel framework helps managers anticipate how consumer responses to AI-generated advertisements ripple upward to affect manufacturer and corporate evaluations, not just brand perception. For corporate communication professionals, this study highlights the importance of maintaining a strong corporate reputation as a protective asset when deploying AI technologies that may evoke consumer skepticism. For policymakers and regulators, insights into how disclosure operates across organizational levels can inform guidelines that balance transparency requirements and unintended consequences. For brand managers, understanding the mediating role of manufacturer attitude clarifies that higher-order corporate perceptions cannot be managed in isolation from brand-level outcomes in AI advertising contexts. Finally, for AI developers and advertising technology firms, this study identifies consumer concerns that should inform the design of AI tools that support rather than undermine organizational trust.

The remainder of this paper is organized as follows. Section 2 presents a comprehensive review of the literature on AI-generated advertising, corporate reputation, attitude toward the manufacturer, and attitude toward the brand, leading to the development of specific hypotheses for this study. Section 3 outlines the research methodology, including the research design, stimulus development, measurement instruments and data collection procedures. Section 4 reports the empirical results, including descriptive statistics, measurement model assessments, and structural model testing. Section 5 discusses the theoretical and practical implications of the findings, addresses the limitations, and suggests directions for future research.

2. Literature Review and Hypothesis Development

2.1 Corporate Reputation and Transparency in AI-Generated Advertising

Corporate reputation represents stakeholders' aggregated evaluations of a firm's credibility, integrity, and social responsibility, serving as a perceptual representation of past actions and future prospects (Fombrun & van Riel, 1997). In the context of AI-generated advertising, reputation is increasingly shaped by how organizations deploy, disclose, and justify their use of AI technology. The integration of generative AI into marketing communications introduces novel reputational considerations that extend beyond traditional concerns regarding product quality or customer service.

Recent industry evidence underscores these reputational stakes. High-profile advertising campaigns have demonstrated that poorly executed AI-generated content can trigger swift public backlash, with consumers quickly identifying logical discrepancies, uncanny visual elements, and signs of human creative displacement (Harrison, 2025; O'Reilly, 2025). Such incidents illustrate that AI-generated advertising represents not only a creative execution choice but also a reputational signal of organizational values and priorities. Research indicates that consumer trust in brands' ability to use AI responsibly remains limited, with confidence in data protection practices declining even as AI adoption accelerates (Okere, 2025). This trust deficit reflects a growing disconnect between technological capability and consumer comfort, exposing a fundamental reputational challenge for firms that embrace AI-driven marketing.

Algorithmic transparency plays a dual role in shaping corporate reputation in AI-generated advertising contexts. Transparency signals ethical responsibility and accountability, demonstrating an organizational commitment to honest communication practices. Research indicates that disclosing AI involvement in advertising can enhance perceptions of honesty and ethical intent, thereby strengthening corporate reputation when AI use is perceived to be aligned with societal values and consumer welfare (Chen et al., 2024; Jesus et al., 2026).

However, transparency also carries reputational risks. Revealing AI authorship can activate consumer skepticism toward AI-mediated persuasion, accentuating the "non-human" nature of advertising content and triggering concerns related to manipulation, authenticity, and data privacy (Chen et al., 2024; Jung et al., 2025). Experimental evidence demonstrates that AI disclosure significantly reduces perceived authenticity, which, in turn, mediates negative effects on brand image and brand trust (Kučinskas & Survilaitė, 2025). This authenticity mediation pathway suggests that transparency alone is insufficient; consumers must also perceive that AI use aligns with authentic brand expression and creates genuine value.

The paradoxical nature of transparency creates strategic dilemmas for organizations. While consumers demand clarity about AI involvement—with surveys indicating a strong desire for clear governance of AI interactions—they simultaneously react negatively when disclosure reveals automation that feels inauthentic or manipulative (Klein & Norman, 2025). This tension is particularly acute for established brands, where the negative effects of AI disclosure are more pronounced than for unfamiliar brands, as consumers hold stronger expectations regarding human involvement in brand expression (Kučinskas & Survilaitė, 2025).

Synthetic creative backlash occurs when consumers perceive brands as displacing human talent or failing to acknowledge their creative contributions. This reaction is particularly intense when AI-generated content contradicts established brand values, as illustrated by examples in which AI-generated advertisements were perceived as misaligned with authentic brand positioning (Harrison, 2025). Such misalignment damages reputation by signaling that the brand does not understand or respect its identity.

Data privacy and governance risks represent the most pervasive reputational threats. Research documents that most consumers express limited trust in brands' ability to use AI responsibly, with confidence in data protection declining sharply (Okere, 2025). The governance gap is striking: while many business leaders acknowledge that strong governance is critical to protecting brand reputation, a significant proportion admit to having little to no formal AI policies (Klein & Norman, 2025). This disconnect between stated priorities and actual practices creates vulnerability to reputational damage when privacy failures occur.

Importantly, a pre-existing corporate reputation moderates the effects of AI-generated advertising on consumer evaluations. Firms with strong reputations appear more resilient to negative reactions, as consumers are more inclined to interpret AI adoption as an innovation rather than opportunism (Jung et al., 2025). This buffering effect operates through attribution processes: when reputable firms use AI, consumers attribute positive motives (efficiency, innovation, enhanced customer experience); when less reputable firms use identical technology, consumers attribute negative motives (cost-cutting, manipulation, and reduced quality).

However, corporate reputation is not merely a moderator but also an outcome that is shaped by AI communication practices. Repeated exposure to poorly received AI-generated advertisements may erode corporate reputation over time, particularly when consumers perceive inauthentic communication or inadequate transparency. The reputational damage from individual AI failures can accumulate, undermining the buffer that previously protected the firm.

2.2 Attitude toward the Brand in AI-Generated Advertising

Attitude toward the brand represents consumers' enduring evaluative orientation toward a brand, encompassing cognitive beliefs, affective responses, and behavioral tendencies (Fishbein & Ajzen, 1975). In the context of AI-generated advertising, brand attitudes are shaped not only by message effectiveness but also by perceptions of authorship, authenticity, and perceived congruence between AI capabilities and advertising objectives. The introduction of algorithmic agency into brand communication fundamentally alters how consumers form and modify brand

evaluations, introducing novel antecedents that extend beyond traditional advertising effectiveness.

Recent bibliometric analyses confirm that AI's impact of AI on branding has attracted considerable scholarly attention, with research examining how AI technologies—including personalization engines, chatbots, predictive analytics, and generative AI—affect key dimensions of brand equity such as awareness, trust, loyalty, and engagement (Al-Habsi et al., 2025; Nguyen & et al., 2025). These studies reveal that AI enhances consumer-brand relationships through personalization and efficiency while simultaneously generating risks related to authenticity and ethical concerns (Sands et al., 2026). The dualistic nature of AI's influence creates what researchers term a "trust paradox," wherein overt disclosure of AI involvement can diminish purchase intention for certain product categories despite AI's ability to enhance functional service quality (Grigsby et al., 2025).

A consistent finding in the literature is that perceived authenticity serves as a critical mediating mechanism linking AI-generated advertising to brand attitudes. Authenticity represents consumers' evaluation of whether a brand's actions reflect genuine, sincere, and truthful expression rather than calculated impression management. In AI-generated advertising contexts, authenticity judgments are particularly consequential because consumers question whether algorithmically generated content can truly reflect brand values and human creativity.

Kučinskas and Survilaitė (2025) provide systematic experimental evidence for the authenticity mediation pathway. Their research demonstrates that AI disclosure significantly reduces perceived authenticity, which in turn has negative effects on purchase intention, brand image, and brand trust. This mediation pattern indicates that transparency about AI authorship triggers negative brand outcomes not because consumers object to honesty per se but because disclosure activates concerns about whether the advertisement represents genuine brand expression. This effect persists even when the actual quality of AI-generated content is high, suggesting that authenticity concerns operate independently of content effectiveness judgments.

Importantly, this authenticity mediation varies with brand familiarity, being more pronounced for established brands than for unfamiliar ones (Kučinskas & Survilaitė, 2025). Consumers hold stronger expectations regarding human involvement in brand expression for familiar brands, making AI disclosure particularly damaging when it violates these expectations. This finding aligns with research demonstrating that the negative effects of AI disclosure are amplified for brands with which consumers have pre-existing relationships and established expectations (Jung et al., 2025).

Ghianda et al. (2025) extend this line of inquiry by examining the impact of AI-generated video advertisements on brand love, a deeper affective construct beyond simple brand attitude. Their experimental research revealed that disclosing AI-generated content negatively affects brand love, with perceived brand authenticity serving as a potential mediator of this effect. This suggests that the emotional dimensions of brand relationships are particularly vulnerable to the perceived artificiality of AI-generated content, as consumers struggle to form deep affective connections with content they perceive as being machine-generated.

The type of advertising appeal employed in AI-generated content significantly affects brand attitudes. Chen et al. (2024) demonstrate that AI-generated ads employing agentic appeals—emphasizing competence, efficiency, and performance—elicit more positive brand attitudes than communal appeals focused on warmth and relationships. This pattern reflects consumers' associations of AI with task-oriented capabilities rather than emotional intelligence, consistent with Mind Perception Theory predictions (Gray et al., 2007).

The differential effectiveness of appeal types is mediated by task self-efficacy, highlighting the importance of consumer confidence in AI's functional competence (Chen et al., 2024). Consumers respond more favorably to AI-generated content when they believe AI can perform advertising tasks effectively. However, this self-efficacy mechanism operates more strongly for agentic than for communal appeals, reflecting the perceived boundaries of AI's capabilities.

Lee (2025) provides important nuance by distinguishing between the creative and analytic applications of AI in advertising. Across multiple studies, this research demonstrates that the

application of generative AI to creative advertising domains—such as visuals and copywriting—elicits negative brand evaluations, while its use in analytic domains—such as strategic planning and targeting—is perceived more favorably. These effects are driven by consumer inferences regarding cost-saving motives. In creative domains, AI use amplifies labor replacement inferences, signaling the displacement of human creativity and leading to more negative attitudes. Conversely, in analytic domains, AI is perceived as an efficiency-optimization tool that does not harm brand perceptions.

While AI-generated advertising can achieve strong performance on functional metrics, such as completion rates and attention, it often struggles to generate deep emotional engagement. Industry research indicates that consumers often react negatively to ads they perceive as AI-made, describing them as "annoying," "boring," or "confusing," and these ads produce weaker memory recall (Tindall, 2025). When viewers sense that an ad is artificially generated, it not only irritates them but also fails to stick in their minds, undermining brand attitude formation.

The emotional limitations of AI-generated content are particularly pronounced in categories where emotional connections are paramount. Research on AI-generated imagery in advertising reveals that emotional engagement with AI-generated content varies, with some consumers experiencing a disconnect because of the artificial nature of the visuals (Nazrin et al., 2025). This emotional disconnect can constrain the development of deeper brand attachment and brand love, even when advertisements achieve surface-level effectiveness on metrics such as awareness or recall.

The relationship between AI-generated advertising and brand attitudes is moderated by several factors. Brand familiarity consistently emerges as a significant moderator, with established brands experiencing more pronounced negative effects from AI disclosure than unfamiliar brands (Jung et al., 2025; Kučinskas & Survilaitė, 2025). This pattern reflects the violation of expectations: consumers hold stronger assumptions about human involvement in the communications of familiar brands, making AI disclosure more disconfirming and, thus, more damaging.

Consumer AI literacy also moderates responses, albeit in complex ways. Individuals with greater knowledge of AI tools may perceive AI disclosure more negatively, exhibiting stronger adverse effects on brand trust than less knowledgeable consumers (Kučinskas & Survilaitė, 2025). This counterintuitive finding suggests that AI literacy may heighten rather than reduce skepticism, as knowledgeable consumers better understand the capabilities and limitations of generative systems. For brand attitude formation, this implies that technically sophisticated audiences—often desirable target segments—may present the greatest challenge for AI-generated advertising acceptance.

2.3 Attitude toward the Manufacturer in AI-Generated Advertising

Attitude toward the manufacturer refers to consumers' evaluative judgments of the firm responsible for producing and disseminating AI-generated advertisements. This construct captures perceptions of the organizational entity behind advertising creation, encompassing assessments of its trustworthiness, ethical intent, technological competence, and corporate responsibility. Although related to corporate reputation and brand attitude, attitude toward the manufacturer occupies a distinct conceptual space; it represents evaluations of the producing entity specifically in its capacity as an advertising creator and technology adopter, rather than broader corporate impressions or product-specific brand associations.

The distinction between manufacturers and corporations is particularly meaningful in AI-generated advertising contexts. Consumers may hold favorable attitudes toward a brand while simultaneously questioning the manufacturer's motives for adopting AI or trust a corporation's overall reputation while scrutinizing the specific manufacturing entity's transparency practices. This multilevel evaluation process reflects the complexity of consumer information processing in the era of algorithmic authorship and automated content creation.

The relationship between AI disclosure and attitudes toward manufacturers is complex and often paradoxical. Grigsby et al. (2025) provide systematic experimental evidence that AI disclosure labels result in lower trust toward the service provider and less positive ad attitudes. Across multiple experiments examining service advertising contexts, the researchers

demonstrated that transparency about AI authorship triggers negative evaluative responses that extend beyond the advertisement itself to encompass judgments about the organization.

This negative effect operates through reduced trust in the advertised brand, which serves as a mediating mechanism linking AI disclosure to downstream outcomes (Grigsby et al., 2025). When consumers learn that an advertisement was generated by AI, they question whether the organization can be trusted to deliver on its promises, particularly in service contexts where intangibility increases uncertainty. The manufacturer's decision to use AI—and its transparency about that use—signals the underlying values and priorities that inform trust judgments.

However, the effects of AI disclosure are not uniformly negative and depend critically on how organizations implement transparency. Yang et al. (2025) draw on self-disclosure theory to examine how different disclosure strategies influence consumer responses to generative AI content. Their research highlights the strategic importance of "transparency management"—the deliberate choice of when, how, and what to disclose about AI involvement. Firms that proactively disclose AI usage in ways that demonstrate ethical responsibility and consumer orientation may mitigate negative reactions, whereas reactive disclosure following consumer detection can trigger perceptions of deception and erode trust (Marketing Science, 2025).

The Persuasion Knowledge Model (Friestad & Wright, 1994) provides a theoretical foundation for understanding how AI disclosure influences manufacturer attitudes. When consumers become aware that an advertisement was AI-generated, they activate persuasion knowledge—their learned beliefs about how, when, and why persuasion attempts occur—and begin scrutinizing the message more critically. This heightened scrutiny extends beyond the advertisement content to encompass evaluations of the organization responsible for its creation.

AI disclosure can activate consumers' persuasion knowledge, prompting them to question manufacturers' motives for adopting AI technology. Specifically, consumers may infer that AI usage reflects cost-cutting priorities rather than consumer welfare enhancement, particularly when AI is applied to creative domains traditionally associated with human expression (Lee, 2025). These motive inferences are consequential because they shape attributions about the

organization’s character. When consumers perceive AI adoption as being driven by efficiency at the expense of authenticity, they form less favorable attitudes toward the manufacturer.

Recent industry examples illustrate this mechanism. H&M's announcement of plans to create digital replicas of models for use in advertising triggered widespread backlash, with critics arguing that the initiative threatened the livelihoods of creative professionals (Harrison, 2025). The underlying consumer inference was that the manufacturer prioritized cost reduction over fair treatment of creative workers, damaging attitudes toward the manufacturer despite continued positive associations with its brand. The use of AI-generated model imagery prompted readers to threaten to cancel their subscriptions, with criticism centering on the publisher's responsibility to support rather than displace human creative talent.

Manufacturers who demonstrate a commitment to ethical AI practices tend to receive more favorable evaluations, even when consumers remain cautious about AI-generated content. The emerging concept of "AI social responsibility" encompasses organizational practices related to transparency, fairness, accountability, and respect for consumer autonomy in AI deployment (Jesus et al., 2026). Manufacturers who embed these principles into their AI governance frameworks signal organizational integrity and consumer orientation.

Recent industry developments highlight the strategic importance of ethical AI certifications and governance. Organizations that receive ethical AI certification following comprehensive audits of AI governance frameworks—including policies, risk management procedures, bias controls, and oversight mechanisms—signal to partners and consumers that they adhere to high standards of transparency and accountability (Klein & Norman, 2025). By subjecting their AI systems to independent third-party validation, manufacturers can differentiate themselves in terms of trust and responsibility, potentially enhancing attitudes toward their organizations.

A persistent challenge for manufacturers using AI-generated advertising is the "authenticity gap"—consumers' perception that AI-generated content lacks the genuine human expression and intentionality characteristic of authentic communication. Research indicates that a substantial majority of consumers do not believe that AI-generated images are authentic (Klein

& Norman, 2025). This perception gap directly affects manufacturers' credibility, as consumers question whether organizations that rely on AI-generated content can be trusted to communicate honestly and genuinely.

The authenticity gap reflects deeper psychological processes related to how consumers evaluate creative works. Research has demonstrated that when identical images are randomly labeled as either "human created" or "AI created," people consistently judge the works they believe to be human-created as more beautiful, meaningful, and profound (Harrison, 2025). This effect operates through attribution rather than perception; the response is not about quality differences but about meaning. Consumers associate creativity with human intention, effort, and emotional expression, qualities that algorithms cannot possess, even when their outputs are visually indistinguishable from human work.

However, the authenticity gap is not uniform across all contexts or for all users. Research demonstrates that the strategic selective use of AI can restore trust and positive attitudes toward service providers (Grigsby et al., 2025). Specifically, when AI is used to generate the tangible attributes of an advertisement while the intangible elements remain authentic, consumer trust and ad attitudes can be preserved. This finding suggests that manufacturers can mitigate authenticity concerns by limiting AI applications to appropriate domains and maintaining human involvement in areas where consumers value authentic expression.

2.4 Linking Corporate Reputation, Attitude toward the Brand, and Attitude toward the Manufacturer

The preceding sections established that AI-generated advertising triggers evaluative processes that operate at multiple levels of consumer cognition. Consumers do not respond to algorithmically authored content in isolation; rather, their judgments encompass the message itself, the brand it promotes, the manufacturer responsible for its creation, and the broader corporate entity within which these activities occur. Understanding the interrelationships among these evaluative levels is essential for developing a comprehensive theoretical account of consumer responses to AI-generated advertisements.

The literature suggests a hierarchical and interconnected structure of consumer evaluations. Corporate reputation functions as a superordinate judgment—an aggregated assessment of a firm's overall credibility, integrity, and social responsibility that shapes interpretations of specific organizational actions (Fombrun & van Riel, 1997). Attitude toward the manufacturer represents a more focused evaluation of the entity directly responsible for advertising production and AI adoption. Attitude toward the brand captures consumer responses to specific brands featured in AI-generated advertising, incorporating both message- and source-level influences. These three constructs are conceptually distinct yet dynamically interrelated, forming what we term a multilevel evaluation framework for understanding the effects of AI-generated advertising.

Several theoretical perspectives illuminate the mechanisms through which corporate reputation, manufacturer attitude, and brand attitude are interrelated in AI-generated advertising contexts. Source Credibility Theory (Deryl et al., 2023) provides foundational insights by conceptualizing the communication source as a determinant of message acceptance. In AI-generated advertising, consumers simultaneously evaluate multiple sources: the immediate source (the advertisement), the organizational source (the manufacturer producing the advertisement), and the corporate source (the broader entity governing the manufacturer).

Attribution Theory (Kelley, 1973) explains how consumers infer motives from an organization’s actions. When consumers learn that a manufacturer uses AI-generated advertising, they spontaneously generate explanations for why the organization adopted this technology. These attributions—whether to cost-saving motives, innovation orientation, or consumer welfare enhancement—shape evaluations of both the manufacturer and brand. Corporate reputation moderates these attributional processes; for reputable firms, consumers attribute AI adoption to positive motives (innovation, efficiency, enhanced customer experience); for less reputable firms, consumers attribute the same behavior to negative motives (cost-cutting, manipulation, reduced quality) (Jung et al., 2025).

Cognitive Consistency Theory (Festinger, 1957) suggests that consumers strive to maintain coherent evaluative structures across related entities. When corporate reputation, manufacturer

attitude, and brand attitude are misaligned—for example, when positive brand attitudes coexist with negative manufacturer evaluations—consumers experience psychological discomfort and may adjust their evaluations to restore consistency. This pressure toward alignment implies that manufacturer and brand attitudes tend to converge over time, with higher-order corporate evaluations anchoring this convergence.

Spillover Effects Literature (Ahluwalia et al., 2000) demonstrates that evaluations of one entity can be transferred to associated entities through associative network processes. In AI-generated advertising contexts, negative responses to manufacturers' AI practices may spill over to affect brand attitudes, whereas a positive corporate reputation may spill over to enhance manufacturer evaluations. These spillover effects operate through cognitive associations that link entities in consumer memory.

Corporate reputation establishes an interpretive framework within which consumers evaluate AI-generated advertising and its producers. As an aggregated judgment based on accumulated experience and information, reputation serves as a heuristic that guides the processing of new information about organizational actions (Fombrun & van Riel, 1997). When consumers encounter AI-generated advertising, they interpret the information through the lens of pre-existing corporate reputation.

Research has shown that firms with strong reputations enjoy greater resilience to negative reactions to AI-generated advertising (Jung et al., 2025). This buffering effect operates via multiple mechanisms. First, reputable firms benefit from favorable attributions: consumers interpret AI adoption as innovation rather than opportunism and enhancement rather than cost-cutting. Second, reputable firms accumulate trust capital that insulates them from immediate negative reactions to specific advertisements. Third, reputable firms' established relationships with consumers create expectations of continued quality and authenticity that persist despite AI involvement.

However, corporate reputation is not merely a moderator of the effects of AI advertising; it is also shaped by these effects. Repeated exposure to poorly received AI-generated advertisements

may erode brand equity and corporate reputation over time, particularly when consumers perceive inauthentic communication or inadequate transparency (Chen et al., 2024). This bidirectional relationship highlights the strategic importance of managing AI-generated advertising within a broader corporate reputation framework.

Attitude toward the manufacturer mediates the relationship between lower-order brand attitude and higher-order corporate reputation. The manufacturer is directly responsible for AI adoption decisions, disclosure practices, and advertising production. Consumers' evaluations of these specific organizational actions are shaped by the manufacturer's handling of AI technology. Research on the effects of AI disclosure demonstrates that consumers evaluate manufacturers based on their transparency practices and inferred motives for AI adoption (Grigsby et al., 2025; Yang et al., 2025). When manufacturers voluntarily disclose their AI involvement in ways that demonstrate ethical responsibility, consumers develop positive attitudes toward them. When manufacturers are perceived to use AI to replace human creativity without adequate justification, consumers form negative attitudes. The mediating role of AI-generated advertising contexts is important in attitude towards the manufacturer because the manufacturer's AI practices provide diagnostic information about organizational values. Consumers may use AI-generated advertising as a cue to infer how the manufacturer will treat them in other domains—whether it will prioritize efficiency over authenticity, whether it will be transparent about its practices, and whether it respects human contributions to creative processes. These inferences are shaped expectations for brand behavior and influence brand evaluations.

Several studies support the brand-to-manufacturer attitude link. Grigsby et al. (2025) demonstrated that AI disclosure reduces trust in the service provider, which in turn is influenced by advertising evaluations and brand outcomes. Sands et al. (2026) found that AI-generated advertisements negatively impact brand credibility and brand attitudes, with effects on consumer perceptions of organizational motives. Kučinskas and Survilaitė (2025) showed that the effects of AI disclosure on brand outcomes operate through reduced perceived authenticity, a manufacturer-level judgment concerning the organization’s commitment to genuine expression.

Therefore, we propose:

H1: Attitudes toward the brand advertised using AI-generated advertising positively influence attitudes toward the manufacturer.

H2: The utilitarian dimension of attitudes toward the brand advertised using AI-generated advertising positively influences attitudes toward the manufacturer.

These hypotheses reflect the spillover of brand-level evaluations to manufacturer-level judgments. When consumers trust the AI-advertised brand and perceive it as ethically responsible in its AI deployment, these positive attributions transfer to the manufacturer. Conversely, negative attitudes toward AI-generated brand advertising—triggered by perceived cost-cutting motives, authenticity concerns, or AI failures—undermine attitudes toward the manufacturer.

The relationship between manufacturer attitude and corporate reputation reflects the transfer of subordinate evaluations to superordinate corporate judgments. When consumers hold favorable attitudes toward the manufacturer, these evaluations generalize to the broader corporate entity. This transfer occurs through associative cognitive processes and the motivation to maintain consistent evaluations across related entities.

Therefore, we propose:

H3: Attitudes toward the manufacturer positively influence corporate reputation.

Together, these hypotheses form an integrated nomological network capturing the multilevel structure of consumer evaluation in AI-generated advertising contexts. The conceptual model specifies the following pathways:

Brand Attitude → Manufacturer Attitude (H1, H2): Brand-level evaluations of AI-generated advertising spill over to manufacturer attitudes.

Manufacturer Attitude → Corporate Reputation (H3): Manufacturer-level evaluations generalize to broader corporate reputation judgments.

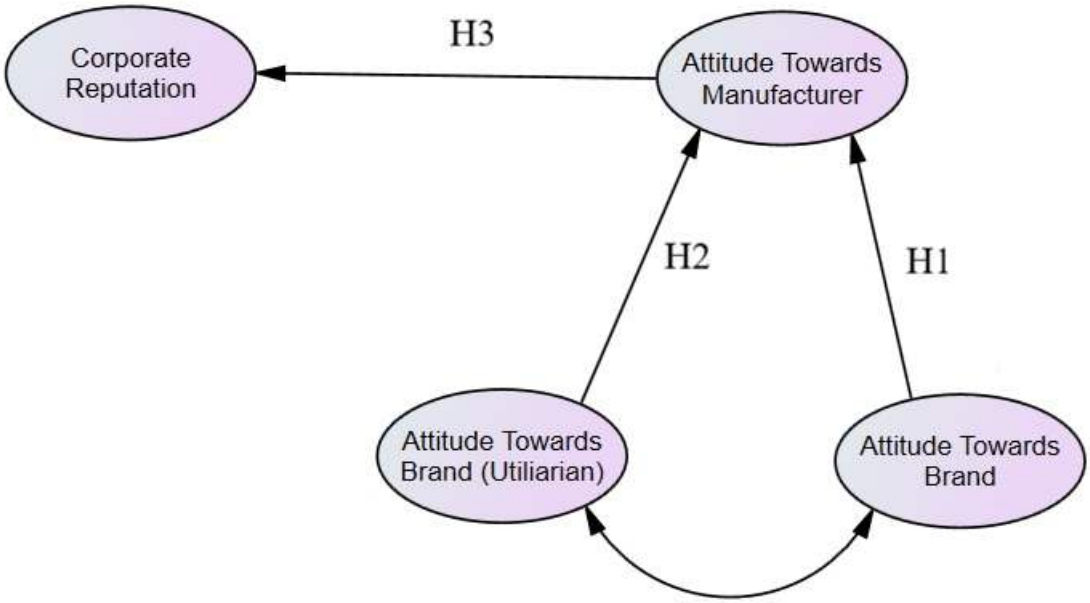

Figure 1. Conceptual model

METHOD

3. Method

3.1 Research Design

This study employed a cross-sectional, survey-based design to investigate the relationships among corporate reputation, attitude toward the brand, and attitude toward the manufacturer in the context of AI-generated advertising. A covariance-based structural equation modeling (CB-SEM) approach was used to simultaneously assess the measurement properties of the constructs and test the hypothesized structural relationships. SEM was considered appropriate due to the latent nature of the variables and its ability to account for measurement error while evaluating multiple interrelated dependencies.

3.2 Sample and Data Collection Procedure

Data were collected via an online questionnaire administered to consumers familiar with digital advertising environments. Participants evaluated AI-generated advertising and reported their perceptions of related constructs. Responses were screened for completeness and consistency, resulting in a final sample of N = 270 usable questionnaires, which exceeds recommended thresholds for SEM and ensures stable parameter estimation.

All constructs were measured using multi-item scales adapted from validated instruments and contextualized for AI-generated advertising. Responses were recorded on seven-point Likert-type scales ranging from 1 (strongly disagree) to 7 (strongly agree). The study included reflective latent constructs representing corporate reputation, attitude toward the manufacturer, attitude toward the brand, and a utilitarian dimension of brand attitude.

3.2.1 Sample Characteristics

The final sample consisted of 270 participants (Table 1). Gender distribution was relatively balanced, with 55.6% male (n = 150), 41.5% female (n = 112), and 3.0% preferring not to disclose (n = 8). Participants' ages ranged from 12 to 68 years, with the majority concentrated in the 18–30-year-old young adult segment.

Table 1. Sample Characteristics (N = 270)

Variable	Category	Frequency (n)	Percentage (%)
Gender	Male	150	55.6
	Female	112	41.5
	Prefer not to disclose	8	3.0
Education Level	High school or below	23	8.5
	Associate degree	62	23.0
	Bachelor's degree	139	51.5
	Master's degree or higher	41	15.2
	Other	5	1.9

Regarding educational attainment, 51.5% held a bachelor's degree, 23.0% an associate degree, 15.2% a master's degree or higher, and 8.5% a high school diploma or lower.

Participants reported moderate to high daily social media usage, with most spending 2–5 hours per day online. Exposure to AI-generated content was common, with 41.1% reporting frequent encounters, 31.9% occasional encounters, and 24.8% rare or no exposure. Similarly, 40.0% often or very often encountered AI-generated advertisements, 35.2% sometimes, and 23.0% rarely or never. These characteristics indicate that respondents were sufficiently familiar with AI-mediated digital environments to provide informed evaluations.

3.3. Validity and Reliability Measures

A confirmatory factor analysis (CFA) was conducted to evaluate the reliability and validity of the measurement model before testing the structural relationships.

Table 2. Model Fit Measures

Measure	Estimate	Threshold	Interpretation
CMIN	352,761	--	--
DF	131,000	--	--
CMIN/DF	2,693	Between 1 and 3	Excellent
CFI	0,939	>0.95	Acceptable
SRMR	0,058	<0.08	Excellent
RMSEA	0,079	<0.06	Acceptable
PClose	0,000	>0.05	Not Estimated

The model was estimated using maximum likelihood estimation in AMOS. Overall fit indices indicated acceptable model fit: $\chi^2(df = 131) = 352.761$, $\chi^2/df = 2.693$, CFI = 0.939, SRMR =

0.058, and RMSEA = 0.079. The χ^2/df ratio was below the recommended threshold of 3.00, SRMR was below 0.08, and RMSEA fell within the acceptable range (< 0.08), indicating adequate fit (Table 2). Although the CFI did not reach the more stringent 0.95 criterion, it surpassed the conventional 0.90 benchmark, suggesting satisfactory fit.

3.4.1 Reliability and Convergent Validity

Internal consistency was assessed using composite reliability (CR), with all constructs exceeding the recommended threshold of 0.70, demonstrating strong reliability: Corporate Reputation (CR): 0.886; Attitude Toward the Manufacturer (ATM): 0.894; Attitude Toward the Brand (ATB): 0.903; Utilitarian Brand Attitude (ATBU): 0.895 (Table 3).

Convergent validity was evaluated using standardized factor loadings and average variance extracted (AVE). Factor loadings were statistically significant and ranged from 0.632 to 0.890, exceeding the minimum recommended level of 0.60. AVE values ranged from 0.633 to 0.679, surpassing the 0.50 threshold, confirming adequate convergent validity.

3.4.2 Discriminant Validity

Discriminant validity was assessed using both the Fornell–Larcker criterion and the heterotrait–monotrait (HTMT) ratio (Fornell & Larcker, 1981; Henseler et al., 2015; Malhotra & Dash, 2011) (Table 3). While the Fornell–Larcker criterion suggested conceptual proximity between certain constructs—particularly Corporate Reputation and Attitude Toward the Manufacturer, as well as between the two brand attitude dimensions—the HTMT ratios provided stronger evidence of discriminant validity. All HTMT values were below the conservative threshold of 0.85, with the highest observed between Attitude Toward the Brand and the Utilitarian Brand Attitude dimension (HTMT = 0.783) (Table 4). None of the HTMT confidence intervals included unity, confirming satisfactory discriminant validity (Gaskin et al., 2023).

Table 3. Validity Analysis

	CR	AVE	MSV	MaxR(H)	C	ATM	ATB	ATBU
C	0,886	0,660	0,737	0,889	**0,812**			
ATM	0,894	0,679	0,737	0,901	0,858***	**0,824**		
ATB	0,903	0,653	0,743	0,909	0,675***	0,661***	**0,808**	
ATBU	0,895	0,633	0,743	0,912	0,559***	0,555***	0,862***	**0,796**

C = Corporate Reputation; ATM = Attitude Toward the Manufacturer; ATB = Attitude Toward the Brand; ATBU = Attitude Toward the Brand (Utilitarian Dimension).

Table 4. HTMT Analysis*

	C	ATM	ATB	ATBU
C				
ATM	0,757			
ATB	0,622	0,600		
ATBU	0,490	0,476	0,783	

C = Corporate Reputation; ATM = Attitude Toward the Manufacturer; ATB = Attitude Toward the Brand; ATBU = Attitude Toward the Brand (Utilitarian Dimension).

4. Structural Model

After confirming the adequacy of the measurement model, the structural model was estimated to test the hypothesized relationships among constructs. Standardized path coefficients (β) were used to evaluate the magnitude and significance of the proposed effects (Gaskin et al., 2022) (Table 5, Figure 2). This two-stage approach ensures that structural inferences are based on a reliable and valid measurement framework.

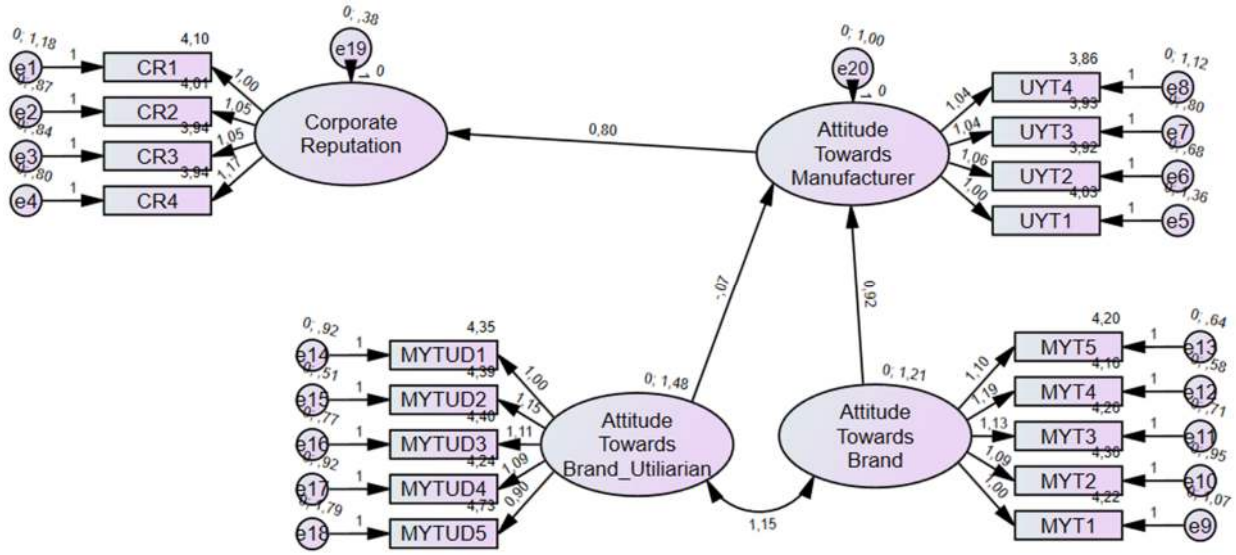

Figure 2. Empirically tested model

4.1 Direct Effects

The results revealed a strong and positive effect of Attitude Toward the Brand (ATB) on Attitude Toward the Manufacturer (ATM) ($\beta = 0.740$, $p < .001$), supporting H1. This indicates that favorable evaluations of AI-generated brand advertisements transfer to perceptions of the manufacturer.

In contrast, the utilitarian dimension of brand attitude (ATBU) did not significantly influence Attitude Toward the Manufacturer ($\beta = -0.065$, n.s.), indicating that functional or performance-based brand evaluations do not meaningfully shape manufacturer perceptions. Thus, H2 was not supported.

Furthermore, Attitude Toward the Manufacturer had a very strong positive effect on Corporate Reputation (CR) ($\beta = 0.871$, $p < .001$), supporting H3. These findings suggest a hierarchical evaluative structure, in which affective brand perceptions influence manufacturer attitudes, which in turn shape broader corporate reputation. The results highlight the central role of manufacturer evaluations in forming corporate-level judgments in AI-generated advertising contexts.

Table 5. Standardized Regression Weights (β)

Predictor	Outcome	Std Beta	Significance
ATBU	ATM	−0.065	n.s.
ATB	ATM	0.740	p < .001
ATM	CR	0.871	p < .001

Table 6. Summary of Hypotheses Testing

Hypothesis	Path	β	Result
H1	ATB → ATM	0.740***	Supported
H2	ATBU → ATM	−0.065	Not Supported
H3	ATM → CR	0.871***	Supported

The structural model demonstrated acceptable fit, consistent with the measurement model. Fit indices indicated good parsimony and adequate representation of the observed data (CMIN/DF = 2.693; CFI = 0.939; SRMR = 0.058; RMSEA = 0.079; PClose = 0.000) (Table 5). The structural model demonstrated overall acceptable fit, suggesting that the proposed model adequately represents the observed data and is suitable for testing the hypothesized relationships.

Table 7. Goodness-of-Fit Indices for Structural Model

Measure	Estimate	Threshold	Interpretation
CMIN	352.761	—	—
DF	131	—	—
CMIN/DF	2.693	1–3	Excellent
CFI	0.939	>0.95	Acceptable
SRMR	0.058	<0.08	Excellent
RMSEA	0.079	<0.06	Acceptable
PClose	0.000	>0.05	Not Estimated

5. Con□lus□on

This study investigated the structural relationships among consumer attitudes toward AI-generated brand advertisements, attitudes toward the manufacturer, and corporate reputation. Employing a cross-sectional survey design and covariance-based structural equation modeling (CB-SEM) on a sample of 270 consumers, the findings reveal a clear hierarchical evaluative process. The results demonstrate that overall attitude toward a brand advertised using AI-generated content has a strong, positive, and significant effect on attitude toward the manufacturer ($\beta = .740$, $p < .001$). In turn, attitude toward the manufacturer exerts a very strong positive influence on corporate reputation ($\beta = .871$, $p < .001$). Interestingly, the utilitarian dimension of brand attitude—which captures functional and performance-based evaluations—did not significantly predict manufacturer attitudes ($\beta = -.065$, n.s.). These results suggest that in the context of AI-generated advertising, the pathway from brand perception to corporate reputation is indirect and primarily driven by holistic, affective brand evaluations rather than purely utilitarian considerations.

This research makes several contributions to the literature on advertising, brand management, and corporate reputation. First, it extends the understanding of affect transfer by demonstrating that favorable attitudes formed toward an AI-generated brand advertisement can significantly enhance perceptions of the manufacturer behind it. This confirms that the psychological mechanisms of evaluative transfer remain potent even with novel, AI-generated content. Second, the study provides a nuanced view of brand attitudes by distinguishing between their overall and utilitarian dimensions. The non-significant finding for the utilitarian dimension suggests that when advertising is generated by AI, consumers may rely more on emotional or holistic impressions of the brand rather than functional assessments to judge the firm. This implies that the "how" of the message (its creative and emotional appeal) may outweigh the "what" (its functional claims) in shaping firm-level perceptions in this specific context. Finally, by confirming the strong link between manufacturer attitudes and corporate reputation, the study reinforces the foundational role of firm-level evaluations in building broader reputational capital, positioning manufacturer attitude as a key mediator in the brand-to-reputation pathway.

The findings offer actionable insights for marketing managers and corporate communication strategists. As firms increasingly adopt AI for content creation, they should recognize that these

tools are not merely for efficiency but are powerful shapers of brand and corporate perceptions. The strong impact of overall brand attitude on manufacturer perceptions suggests that investments in AI-generated advertising should prioritize creative excellence and emotional engagement. Strategies focused on building a compelling, holistic brand image in AI-generated ads will yield reputational dividends for the firm. Conversely, an overemphasis on functional or utilitarian messaging may be less effective in improving how consumers view the manufacturer. Therefore, organizations should guide their AI tools to generate content that is not only informative but also emotionally resonant and experientially appealing. Ultimately, managing consumer attitudes at the brand level, particularly through engaging AI-generated content, is a strategic lever for enhancing corporate reputation.

L□m□tat□ons and Future Resear□h

While this study provides valuable insights, its limitations should be acknowledged, which in turn open avenues for future research. First, the cross-sectional design captures perceptions at a single point in time, precluding causal inferences and the observation of how these attitudes evolve. Future studies could employ longitudinal designs to track the stability and change of these relationships over time. Second, the sample, while adequate for SEM, was predominantly composed of young adults. Replicating these findings across more diverse age groups and cultural contexts would enhance their generalizability. Third, this study focused on a general conceptualization of AI-generated advertising without specifying ad characteristics (e.g., product type, level of personalization, perceived creativity). Future research could experimentally manipulate these features to identify boundary conditions. Additionally, exploring moderating variables, such as consumer familiarity with AI technologies, skepticism toward AI-generated content, or the perceived authenticity of the ad, could provide a more comprehensive understanding. Finally, investigating other dimensions of brand attitude—such as symbolic, social, or ethical evaluations—may further clarify the mechanisms through which AI-generated advertising influences manufacturer perceptions and corporate reputation.

References

Ahluwalia, R., Burnkrant, R. E., & Unnava, H. R. (2000). Consumer response to negative publicity: The moderating role of commitment. *Journal of Marketing Research*, *37*(2), 203–214. https://doi.org/DOI 10.1509/jmkr.37.2.203.18734

Al-Habsi, Z. S., Mohamed, H. R., & Methuku, H. (2025). AI-driven brand equity: A systematic review of consumer perceptions, engagement, and loyalty in the age of intelligent branding. *TPM - Testing, Psychometrics, Methodology in Applied Psychology*.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.

Burrow, E., & Gutentag, J. (2025). AI-generated versus human-generated creative content in advertising. *Pepperdine Journal of Communication Research*, *13*(1), 8–25.

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

10.1016/j.jbusres.2024.114867

Deryl, M. D., Verma, S., & Srivastava, V. (2023). How does AI drive branding? Towards an integrated theoretical framework for AI-driven branding. *International Journal of Information Management Data Insights*, *3*(2), 100205. https://doi.org/10.1016/j.jjimei.2023.100205

Exner, Y., Hartmann, J., Netzer, O., & Zhang, S. (2025). AI in disguise: How AI-generated ads' visual cues shape consumer perception and performance. *SSRN*.

Festinger, L. (1957). *A theory of cognitive dissonance*.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*.

Fombrun, C. J., & van Riel, C. B. M. (1997). The reputational landscape. *Corporate Reputation Review*.

Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, *18*(1), 39–50. https://doi.org/Doi 10.2307/3151312

Friestad, M., & Wright, P. (1994). The Persuasion Knowledge Model - How People Cope with Persuasion Attempts. *Journal of Consumer Research*, *21*(1), 1–31. https://doi.org/Doi 10.1086/209380

Gaskin, J., James, M., Lim, J., & Steed, J. (2023). *Master validity tool*.

Gaskin, J., Lim, J., & Steed, J. (2022). *Merge SRW Tables*.

Ghianda, E., Albani, R., Cabrera, L. P., & Costanzo, F. (2025). The impact of adopting and disclosing generative AI in digital advertising on brand perceptions.

Gölgeli, K. (2025). Yapay zekâ ile reklam tasarımı: Reklamcılara yönelik bir araştırma. *İnsan ve Toplum Bilimleri Araştırmaları Dergisi*.

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619. https://doi.org/10.1126/science.1134475

Grigsby, J. L., Michelsen, M., & Zamudio, C. (2025). Service ads in the era of generative AI: Disclosures, trust, and intangibility. *Journal of Retailing and Consumer Services*.

Gu, C. Y., Jia, S. Y., Lai, J. Y., Chen, R. L., & Chang, X. S. Y. (2024). Exploring Consumer Acceptance of AI-Generated Advertisements: From the Perspectives of Perceived Eeriness and Perceived Intelligence. *Journal of Theoretical and Applied Electronic Commerce Research*, *19*(3), 2218–2238. https://doi.org/10.3390/jtaer19030108

Harrison, P. (2025). A backlash against AI imagery in ads may have begun as brands promote 'human-made'.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the academy of marketing science*, *43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Jesus, J. B., Besario, M. R. A., Bueno, N. L. E., & et al. (2026). Brand reputation in the age of AI: Impact of AI-driven customer interaction and sustainability messaging on consumer trust. *Humanities and Social Sciences Communications*.

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

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist*.

Klein, J., & Norman, T. (2025). AI in marketing and ethics: A look at business, corporate and consumer perspectives.

Kučinskas, G., & Survilaitė, E. (2025). Revealing AI involvement in ad creation: Effects on authenticity, brand perceptions and consumer intentions. *Journal of Information Systems Engineering and Management*.

Lee, D. (2025). Consumers' reactions to AI in advertising: Creative versus analytic AI applications affect brand attitudes. *Kellogg School of Management Working Paper*.

Malhotra, N. K., & Dash, S. (2011). Marketing research an applied orientation (paperback). In: London: Pearson Publishing.

Marketing Science, I. (2025). *MSI Webinar: The AI-usage conundrum*.

Nazrin, N., Isahak Merican, F. M., Taharuddin, N. S., & Ibrahim, N. I. (2025). How AI-generated images impact consumer perceptions and brand identity. *Idealogy Journal*.

Nguyen, & et al. (2025). Impact of artificial intelligence on branding: A bibliometric review and future research directions. *Humanities and Social Sciences Communications*.

O'Reilly, L. (2025). 2025年 AI 行銷五大災難.

Okere, B. (2025). Redefining brand trust: How AI and data privacy have reshaped the digital economy in 2025 and beyond.

Sands, S., Demsar, V., Ferraro, C., Wilson, S., Wheeler, M., & Campbell, C. (2026). Easing AI-advertising aversion: how leadership for the greater good buffers negative response to AI-generated ads. *International Journal of Advertising*, *45*(1), 27–47. https://doi.org/10.1080/02650487.2025.2457080

Tindall, A. (2025). Consumers don't know AI made the ad... and when we tell them, they don't really care.

Wu, Z., Zeng, L., & Huang, Y. (2025). Influence of the characteristics of AI-generated advertising on consumers' purchase intention. *Journal of Arts and Cultural Studies*.

Yang, M., Lu, X., & Head, M. (2025). Unveiling the illusion: Generative artificial intelligence disclosure strategies and consumer responses.

Yin, M., Ma, B., & Pan, X. (2024). Consumer attitudes toward generative AI advertisements. *Journal of Cases on Information Technology*.

CHAPTER III

Artificial Intelligence Explanation Effects Scale (AIDES)

Serhat TERZİOĞLU

Abstract

As artificial intelligence becomes increasingly embedded in decision-making systems, the need for transparent, interpretable, and ethically accountable algorithms has grown substantially. Yet, empirical tools to systematically measure how users perceive AI explanations remain scarce. This study addresses that gap by developing and validating a multidimensional scale to assess Artificial Intelligence Explanation Effects (AIDES) in digital environments. Drawing on theories from explainable AI (XAI), algorithmic transparency, and trust in automation, the scale operationalizes four interrelated dimensions: Perceived Transparency and Understanding (PTU), Perceived Credibility and Honesty (PCH), Ethical Fairness and Accountability (EFA), and Trust Calibration and Behavioral Response (TCB). Survey data from 289 regular users of AI-generated content and digital platforms were analyzed using confirmatory factor analysis and structural equation modeling. Results support the scale's psychometric robustness: composite reliability exceeded recommended thresholds, convergent validity was confirmed via average variance extracted, and discriminant validity was established using the heterotrait–monotrait ratio. Although initial model fit indicated room for refinement—particularly within the trust calibration dimension—the overall findings suggest that AIDES reliably captures distinct aspects of user responses to AI explanations. This instrument contributes to human-centered XAI research by shifting focus from technical explainability to empirical, user-based evaluation. Practically, it offers designers and organizations a structured way to assess explanation mechanisms in AI-enabled platforms, online advertising, and automated systems. Limitations include the cross-sectional design, limited generalizability beyond digitally active populations, and need for further scale purification. Future work should test the scale across diverse AI applications, including recommender systems, generative AI, and autonomous decision tools, as well as integrate it with broader technology acceptance models.

Introduction

AI systems are increasingly playing a central role in decision-making across healthcare, finance, public services, and for marketing automation. This widespread integration has made the need to provide understandable, reliable, and ethical explanations for *how* systems work and *why* they produce specific outputs an urgent research and implementation priority. The field of Explainable AI (XAI) is rapidly evolving from both technical and human-centered perspectives to address this need. Explainable AI (XAI) is rapidly evolving from both technical and human-centered perspectives (Liao & Varshney, 2023; Miller, 2019) show strong growth in user-centered XAI methods such as contrastive and example-based explanations. However, the actual psychological, perceptual, and behavioral effects of explanations provided by AI systems on end users have not yet been sufficiently understood or systematically measured. Existing findings rely on fragmented and inconsistent metrics for trust, understanding, and calibration, and no unified measurement framework has yet been established (Weitz et al., 2024).

The current literature tends to examine the effects of explanations in a one-dimensional manner, often focusing solely on outcomes such as trust or understanding, or employing fragmented scales limited to specific contexts. This approach restricts the ability to capture the multidimensional and dynamically interactive nature of explanation effects, complicates comparisons across studies, and slows the development of a coherent theoretical framework. Consequently, there is a growing need for a multidimensional measurement model capable of capturing these interrelated effects.

The literature increasingly calls for integrated frameworks such as AIDES. For example, how does the feeling of understanding (i.e., perceived transparency) generated by an explanation shape users' trust in the system? Does this trust translate into behavior calibrated to the system's actual capabilities? Furthermore, could explanations, while strengthening perceptions of fairness, also function as a form of "justification" that discourages critical scrutiny of the system? These complex and interdependent questions can only be adequately addressed through a multidimensional measurement model that holistically captures explanation effects.

This paper proposes a conceptual framework to address this critical gap—the Artificial Intelligence Explanation Effects Scale (AIDES)—which captures the primary outcomes of explanations in human–AI interaction across four fundamental dimensions:

1. **Perceived Transparency and Understanding (PTU):** The user's subjective feeling of understanding and enlightenment regarding the system's operation.
2. **Perceived Credibility and Honesty (PCH):** The evaluation of the system as a correct, reliable, and sincere source or decision-making partner.
3. **Ethical Fairness and Accountability (EFA):** The perception of the system as fair, unbiased, and attributable with responsibility.
4. **Trust Calibration and Behavioral Response (TCB):** The alignment of the user's trust with the system's real performance and how this trust translates into final decisions or actions.

1. PERCEIVED TRANSPARENCY AND UNDERSTANDING (PTU)

Perceived Transparency and Understanding (PTU) refers to the cognitive and psychological assessment of the feeling of understanding and transparency perception acquired by a user or affected party of an AI system regarding its operation, decision processes, or outputs. This dimension is concerned not with how technically explainable the system is, but rather with how knowledgeable and enlightened the user *feels*. In the literature, this concept is also associated with terms such as "conceptual transparency," "mental model accuracy," and "perceived comprehensibility."

The vast majority of studies show that providing any type of explanation, compared to providing none, statistically significantly increases users' perceived levels of understanding and transparency (Eiband et al., 2022). For example, in a recommendation system, providing users with simple feature-highlighting explanations answering the "why this recommendation?" question significantly increased participants' self-reported confidence in understanding the system's logic. The underlying mechanism of this effect lies in explanations creating an illusion of causality or a coherent narrative in the user. Seeing a "justification" for the decision process, the user believes they understand the process.

The strongest consensus in the literature is that explanations consistently increase subjective understanding perception (PTU). However, this finding is shadowed by two critical caveats; the increase in PTU depends on the explanation's compatibility with the user's cognitive model. For example, contrastive explanations ("Why class A and not class B?") have shown a stronger PTU effect compared to simple feature importance rankings in many studies. Contrastive and

counterfactual explanation formats have been shown to enhance comprehension and user satisfaction compared to static visualizations (Chandrasekaran et al., 2023; Miller, 2019). As confirmed by studies high PTU scores do not guarantee that the user understands the system's actual working principle (Hoffman et al., 2021; Zhang et al., 2020) This gap shows that PTU alone is an insufficient success metric and must be supported by objective understanding tests.

PTU is measured almost entirely via subjective self-report scales. Typical Likert items include: "I understand the reasons behind the system's decision.", "I think I have sufficient information about how the system works.", "The decision process was transparent to me." (Eiband et al., 2022; Shin, 2021). Although a standard PTU scale has not yet gained widespread acceptance, some studies have developed reliable sub-scales. For example, one sub-dimension of the "Attitude Towards Explainable AI Scale" is generally associated with understanding/transparency.

Most studies use between-subjects experimental designs comparing different explanation conditions (explanation vs. none, explanation type A vs. B). PTU is generally taken as the dependent variable in these experiments. Current PTU findings largely come from Western, highly educated samples. All studies measure immediate responses.

2. PERCEIVED CREDIBILITY AND HONESTY (PCH)

Perceived Credibility and Honesty (PCH) refers to the degree to which an AI system is evaluated in the eyes of the user as a correct, reliable, and sincere source of information or decision-making partner. This dimension is closely related to the user's attributed intention and integrity towards the system, beyond its technical accuracy. PCH is one of the fundamental components constituting the general concept of "trust" and is concerned with the extent to which the system's explanations reinforce the belief that its decisions are valid, unbiased, and not misleading (Wang & Yin, 2023).

Research shows that explanations generally have a positive but conditional effect on PCH. The strengthening of PCH by explanations depends on their quality and contextual appropriateness.

Explanations that meet the user's need for inquiry, are understandable, and provided in a timely manner increase the perception of the system's credibility. For example, for a medical diagnosis AI, an explanation highlighting only the most important clinical findings leads to higher PCH than a long, technical model architecture explanation (Jussupow et al., 2021). However, superficial, inconsistent, or misleading explanations can be perceived as "transparency theater," seriously damaging credibility.

Explanations can enhance perceptions of system credibility and sincerity. However, a critical concern is that increased reliance on the system may be conflated with perceived credibility and honesty (PCH). Notably, empirical findings indicate that user trust can increase even when explanations are provided for incorrect model outputs (Jacovi et al., 2021). This suggests that credibility should not be assessed solely in terms of trust or reliance, but must also incorporate the system’s ability to transparently communicate its limitations and uncertainties.

The effect on perceived credibility and honesty (PCH) is strongly influenced by users’ prior knowledge and expertise. Users with low domain knowledge may interpret any explanation as a legitimizing signal, which can easily elevate PCH. In contrast, expert users tend to evaluate explanations more critically; when they detect inconsistencies or superficiality, PCH may decrease (Zhang et al., 2020). Similarly, research shows that experienced users exhibit more critical judgment, whereas novices are more prone to over-trust (Cheng & Jiang, 2024; Jussupow et al., 2021). These findings suggest that explanations may create an “over-trust trap,” particularly for less knowledgeable users.

Context plays a critical role in the formation of PCH. In low-risk, low-complexity tasks (e.g., movie recommendation) simple explanations provide sufficient PCH, while in high-risk tasks (e.g., loan application denial, medical prognosis) users expect more comprehensive, justified, and justice-element-containing explanations. When these expectations are not met, PCH drops rapidly and the system's adoption is hindered (Shin, 2021).

The measurement of PCH is predominantly done through self-report surveys. Commonly used scale items include: "I believe the system gives me correct information," "The system's outputs

are reliable," "The system's decisions are sincere and not misleading" (Wang & Yin, 2023). Some studies treat the credibility perception as a sub-dimension of a broader trust scale. A methodological challenge is the clear separation of PCH from "perceived accuracy"; many studies treat these two concepts as overlapping.

Perceived Credibility and Honesty (PCH) is one of the fundamental psychological constructs determining users' intention to adopt and collaborate with an AI system. When developing the PCH sub-dimension of the AIDES scale, the multidimensional and context-sensitive nature of this dimension must be considered. Scale items should aim to measure not only general credibility perception but also belief in the system's honesty and sincerity

3. ETHICAL FAIRNESS AND ACCOUNTABILITY (EFA)

Ethical Fairness and Accountability (EFA) refers to the perception created in users by an AI system's explanations regarding the system's compliance with moral principles, freedom from bias, and how responsibility is shared in the face of possible negative consequences. Studies show that perceived transparency influences trust (Jacovi et al., 2021; Wang & Yin, 2023) and that fairness perceptions can be distorted by persuasive explanations (Alon-Barkat & Busuioc, 2023). This dimension is concerned with the perceived credibility and clarity of the responsibility network regarding the system's alignment with social and ethical values, rather than with the technical performance metric of a "fair algorithm" (Shin, 2021). Explanations, in this context, serve to satisfy users' sense of justice by placing the system's decisions within a moral framework and to create a mental model of "who can be held responsible" in case of errors.

Studies on EFA reveal that explanations have both strong and paradoxical effects in shaping this perception. Explanations, by showing the logic behind a decision, lead users to believe the decision is not arbitrary but justified. For example, in a loan denial decision, an explanation highlighting factors like "low income" and "insufficient credit history" gives the impression that the decision is the result of an impersonal, rule-based process. This strengthens the system's perceived procedural fairness. However, this effect becomes problematic if the explanation hides actually existing systematic biases (e.g., the use of zip code as a proxy for income data).

The most critical discussion in the literature is that explanations can provide a moral license, thereby rationalizing systems that are actually unfair. Users tend to question the system less when they see an explanation. This situation is called the "transparency trap" and is particularly dangerous when the explanation makes injustices inherent in the underlying data or model selection invisible (Alon-Barkat & Busuioc, 2023; Green & Chen, 2019). It has been shown that participants who received an explanation found the same negative outcome (e.g., denial of aid) fairer compared to those who did not receive an explanation. This indicates that EFA measurement should also encompass users' critical awareness. Explanations can provide a moral license, rationalize fundamentally unfair systems, and suppress critical inquiry. This dilemma proves that EFA measurement should include not only "finding fair" but also a dimension of "critical awareness" (perceiving system justification bias).

Effective explanations make it possible to trace responsibility in case of an error or harm. When users understand which data or rule a decision is based on, they gain the capacity to attribute the problem to the human operator, data engineer, algorithm developer, or corporate management. This is the basis of accountability. Research shows that explanations clarify users' "responsibility map" and thus reinforce the belief that the system is controllable (Wagner, 2023). However, it should be noted that tracing this path is practically difficult in very complex or distributed systems.

EFA perception is usually measured through self-report surveys and often includes items adapted from perceived justice scales. Common items include: "The system's decisions are impartial and unbiased," "The decision-making process complies with ethical principles," "I understand who is responsible when a problem occurs," "If the system makes a wrong decision, it can be held accountable" (Shin, 2021). Qualitative studies deeply examine users' interpretations of justice and accountability through case scenarios. Behavioral metrics (e.g., intention to fill out an injustice complaint form) are used less frequently but are valuable for measuring the practical consequences of EFA.

Current studies predominantly focus on individual–AI interaction. However, the extent to which explanations support or undermine accountability mechanisms at the institutional level remains underexplored, particularly in brand and corporate contexts.

4. TRUST CALIBRATION AND BEHAVIORAL RESPONSE (TCB)

Trust Calibration and Behavioral Response (TCB) is the dimension examining how an AI system's explanations align the user's level of trust in the system with that system's real capabilities and limitations, and how this trust translates into final behavioral outputs. "Calibration" here means the user's trust is proportional to the AI's objective accuracy or reliability in a specific context (Lee & See, 2004). Ideal calibration is the user appropriately adopting the AI's suggestions when it is strong (appropriate trust), and questioning, correcting, or rejecting them when it is weak or erroneous (appropriate suspicion). TCB is therefore the ultimate real-world test of effectiveness and safety for all other perceptual dimensions (PTU, PCH, EFA).

The effect of explanations on TCB is filled with complex and sometimes contradictory findings in the literature. The core debate revolves around whether explanations "calibrate" trust or rather "disrupt" it. The greatest paradox of explanations is that they can disrupt calibration in both directions. On one hand, any explanation, even simple or superficial, can create an illusion of understanding and control in the user, leading to over-reliance on the system. This results in the user blindly following the AI's suggestions even when it is erroneous (Jacovi et al., 2021). On the other hand, complex explanations or those explicitly revealing the AI's uncertainties/errors can cause trust attenuation or unnecessary suspicion in the user, leading to the rejection of useful suggestions when the AI is correct. This dilemma is termed the "explanation-calibration dilemma."

However, when correctly designed, explanations can improve calibration. Explanations that clearly indicate confidence scores or prediction intervals when the model is unsure allow users to better assess risk (Bansal et al., 2021). Showing example scenarios where the system could be wrong helps users ground their trust in a more realistic basis. Explanations containing

external information that the user can independently verify (e.g., "This tumor has the X feature seen in your past cases") provide better calibration (Zhang et al., 2020).

The real test of TCB is behavior. Therefore, the literature focuses on the final decision quality of the human-AI team. Calibrated trust maximizes performance. The rate at which the user accepts the AI suggestion is a calibration indicator if it shows high correlation with the AI's accuracy rate. The frequency with which the user corrects the AI suggestion, changes in decision-making time, or the tendency to seek additional human approval indicate levels of suspicion and attention (Wang & Yin, 2023).

The effect on TCB is not universal. In high-risk and complex tasks, explanations become more critical for calibration. Also, self-confident or expert users can adjust their calibration better by using explanations more effectively, while novice users are more prone to the over-trust trap.

Measuring TCB requires the combined use of self-report and behavioral metrics through mixed methods. Likert items measuring calibration self-efficacy like "I can predict when the system might make an error in this task" are used. Behavioral/Objective Measures are the backbone of TCB. Researchers design tasks in controlled experiments where the AI intentionally makes errors at certain intervals. Calibration Metrics are calculated: for example, the correlation between the user's reported trust level and the AI's objective accuracy at that moment, or the percentages of "over-trust" (trust > accuracy) and "under-trust" (trust < accuracy) (Bansal et al., 2021). The accuracy rate of the human-AI team is compared with the performance of the AI or human alone. Optimal calibration targets the point where team performance is maximum.

Almost all current studies examine one-off interactions. How do users sustain or lose their trust calibration during repeated and long-term use of an AI tool? TCB studies have been conducted predominantly in medicine, finance, and judicial decision support systems. The effect of explanations on TCB in fields like marketing communication, advertising, and customer relationship management, where trust has different dynamics, has been almost unstudied. For example, it is unknown how a personalized ad explanation calibrates a consumer's trust in a brand.

Developing the TCB sub-dimension of the AIDES scale will be a contribution beyond current scale development studies. This dimension should contain items that query not only subjective perceptions but also the alignment between subjective trust and behavioral intention (e.g., "I can detect situations where the AI might make an error" and "In these situations, I review my decision independently of the AI"). Furthermore, analyzing the relationship of TCB scores with those obtained from PTU, PCH, and EFA can reveal which perceptual pathways lead to the best-calibrated behavior. This holds great value both theoretically and practically.

Table 1: Summary of Data Extraction for 28 Studies Classified by AIDES Dimensions

No	Author(s) & Year	Main Focus & Used/Developed Scale	Methodology (Design, Sample)	Key Findings Related to AIDES	Primary AIDES Dimension
1	(Bansal et al., 2021)	Mental models and team performance; behavioral calibration metrics.	Experimental lab study (N=120).	Explanations conveying uncertainty improve trust calibration and human-AI team performance.	**TCB** (Primary), PTU
2	(Jacovi et al., 2021)	Trust calibration; mismatch between subjective trust and objective accuracy.	Experimental online study (N=300).	Superficial explanations lead to over-trust even for wrong models. Trust measurement must be considered with calibration.	**TCB, PCH**

3	(Shin, 2021)	Perceived explainability, causality, trust, and acceptance. Adapted perception scales.	Online survey experiment (N=450).	Perceived causality and explainability positively affect trust and fairness perception. This effect varies by context.	**PTU, PCH, EFA**
4	(Zhang et al., 2020)	Trust and accuracy calibration; trust self-efficacy scale.	Experimental study (N=180).	Explanations allowing self-verification improve both trust calibration and objective understanding.	**TCB, PTU**
5	(Eiband et al., 2022)	Effect of explanations on subjective understanding and trust. Likert-type perception scales.	User study (N=45).	Explanations increase subjective understanding and trust, but this increase is not always related to objective performance.	**PTU, PCH**
6	(Wang & Yin, 2023)	Comparative study on the helpfulness of explanations in	Between-subjects experimental design (N=225).	Explanations increase trust and performance	**TCB, PTU**

		decision-making. Behavioral adoption rates.		in simple tasks but can increase cognitive load in complex tasks.	
7	(Alon-Barkat & Busuioc, 2023)	Algorithmic bias in public sector and the legitimizing role of explanations.	Mixed methods (case study + survey).	Explanations can create a "transparency trap" by causing citizens to find unjust outcomes fairer.	**EFA** (Critical)
9	(Liao & Varshney, 2023)	Human-centered XAI experiences. Conceptual framework.	Review/Conceptual.	Effective explanation should be adapted to the user's task goal, expertise, and context.	**PTU, TCB** (Framework)
10	(Cheng & Jiang, 2024)	Personalized explanations based on user values and fairness perception.	Experimental study (N=200).	Process-focused users reported higher fairness perception from global explanations, outcome-focused users	**EFA, PTU**

				from local explanations.	

3. Method

3.1 Research Design

This study aims to develop and validate a multidimensional scale designed to capture Artificial Intelligence Explanation Effects (AIDES) in consumer interactions with AI-enabled systems and algorithmic decision environments. Specifically, the research proposed and empirically tested four conceptual dimensions: Perceived Transparency and Understanding (PTU), Perceived Credibility and Honesty (PCH), Ethical Fairness and Accountability (EFA), and Trust Calibration and Behavioral Response (TCB). These dimensions collectively represent how individuals cognitively and ethically interpret explanations provided by artificial intelligence systems and how such explanations influence trust formation and behavioral responses toward algorithmic outputs.

A cross-sectional survey-based research design was employed. The proposed scale was evaluated using covariance-based structural equation modeling (CB-SEM) to assess its measurement properties and construct validity. Structural equation modeling is widely recommended for scale development because it allows researchers to simultaneously assess latent constructs, measurement error, and relationships among indicators (Anderson & Gerbing, 1988; Hair et al., 2019).

The conceptualization of Artificial Intelligence Explanation Effects builds on prior research emphasizing the importance of explainability, transparency, and interpretability in human–AI interaction (Doshi-Velez & Kim, 2017; Miller, 2019). Explanations provided by algorithmic systems influence how users interpret automated decisions, evaluate system credibility, and calibrate their trust toward AI outputs.

3.2 Sample and Data Collection Procedure

Data were collected through an online survey administered to adult consumers who regularly interact with digital platforms and AI-mediated content environments. Participants were recruited through online networks and social media channels. Participation was voluntary and anonymous.

A total of 289 responses were obtained. After screening for missing values and response completeness, the dataset retained sufficient observations for robust statistical analysis. Sample sizes above 200 respondents are generally considered adequate for structural equation modeling, particularly for scale validation studies involving multiple latent constructs (Hair et al., 2019).

All constructs were measured using multi-item reflective scales developed from the theoretical literature on explainable artificial intelligence and algorithmic trust. Responses were captured using seven-point Likert scales ranging from 1 (strongly disagree) to 7 (strongly agree).

3.3 Sample Characteristics

The final sample consisted of 289 respondents. Descriptive statistics for demographic and behavioral variables are reported in Table 1.

Participants ranged in age from 18 to 99 years, with a mean age of 32.10 years and a median age of 27 years, indicating that the sample was primarily composed of digitally active young adults. Respondents reported an average of 4.83 hours of daily social media usage, suggesting frequent exposure to algorithmically generated information environments.

Interaction with AI-generated content was relatively common. Approximately 40.7% of respondents reported often or very often interacting with AI-generated content, while 31.7% reported occasional interaction. Exposure to AI-generated advertisements followed a similar pattern, with 40.2% encountering them often or very often and 35.7% sometimes encountering them.

These patterns indicate that the sample possessed sufficient familiarity with AI-mediated digital environments to provide informed evaluations of AI explanation mechanisms.

Table 1. Descriptive Statistics of Key Variables

Variable	N	Mean	Median	Minimum	Maximum
Age	289	32.10	27	18	99
Education Level	289	3.68	3	1	99
Daily Social Media Use (hours)	275	4.83	4	1	24
Interaction with AI Content	287	3.16	3	1	5
Exposure to AI Advertisements	286	3.23	3	1	5

Table 2. Education Distribution

Education Level	Frequency	Percentage (%)
High school or below	41	14.2
Associate degree	61	21.1
Bachelor's degree	132	45.7
Postgraduate	52	18.0
Other	3	1.0
Total	**289**	**100**

Table 3. Exposure to AI Content and Advertisements

Variable	Category	Frequency	Percentage (%)
Interaction with AI Content	Never	32	11.1
	Rarely	47	16.3
	Sometimes	91	31.5
	Often	77	26.6
	Very often	40	13.8
Exposure to AI Ads	Never	15	5.2
	Rarely	54	18.7
	Sometimes	102	35.3
	Often	79	27.3
	Very often	36	12.5

3.4 Measurement Model Assessment

To evaluate the psychometric properties of the proposed scale, confirmatory factor analysis (CFA) was conducted using maximum likelihood estimation in AMOS. Model fit was assessed using multiple goodness-of-fit indices following established guidelines (Hu & Bentler, 1999).

The CFA results indicated mixed model fit. While the standardized root mean square residual (SRMR = 0.068) indicated acceptable residual-based fit, other indices suggested that the model required refinement.

Table 4. Model Fit Measures

Measure	Estimate	Threshold	Interpretation
CMIN	464.670	—	—
DF	84	—	—
CMIN/DF	5.532	1–3	Poor
CFI	0.848	>0.90	Poor
SRMR	0.068	<0.08	Acceptable
RMSEA	0.125	<0.06	Poor
PClose	0.000	>0.05	Not supported

Inspection of standardized residual covariances suggested that indicator TCB3 may contribute to model misfit, indicating that future refinement of the Trust Calibration construct may improve overall model performance.

3.5 Reliability and Convergent Validity

Internal consistency reliability was assessed using Composite Reliability (CR) and Maximal Reliability (MaxR(H)). All constructs exceeded the recommended threshold of 0.70, indicating satisfactory internal consistency (Hair et al., 2019).

Table 5. Reliability and Convergent Validity

Construct	CR	AVE	MSV	MaxR(H)
PTU	0.849	0.587	0.811	0.862
PCH	0.737	0.485	0.702	0.750
EFA	0.852	0.592	0.811	0.864
TCB	0.785	0.484	0.621	0.830

Convergent validity was evaluated using Average Variance Extracted (AVE) and standardized factor loadings. AVE values ranged from 0.484 to 0.592. Although two constructs slightly fell below the conventional 0.50 threshold, convergent validity is still considered acceptable when composite reliability exceeds 0.70 (Malhotra & Dash, 2011).

3.6 Discriminant Validity

Discriminant validity was evaluated using both the Fornell–Larcker criterion and the Heterotrait–Monotrait ratio (HTMT) (Fornell & Larcker, 1981; Henseler et al., 2015). The HTMT method is considered a more reliable test of discriminant validity in modern SEM applications.

All HTMT values were below the conservative 0.85 threshold, ranging between 0.614 and 0.764, indicating satisfactory discriminant validity.

Table 6. HTMT Discriminant Validity Matrix

	PTU	PCH	EFA	TCB
PTU	—			
PCH	0.659	—		
EFA	0.764	0.683	—	
TCB	0.626	0.614	0.623	—

No HTMT warnings were detected, and none of the confidence intervals included unity, confirming that the four constructs represent empirically distinct dimensions of AI explanation effects.

Conclusion

The increasing integration of artificial intelligence into decision-making systems has intensified the need for greater transparency, interpretability, and ethical accountability in algorithmic processes. Despite the growing prominence of explainable artificial intelligence (XAI), empirical measurement tools capable of systematically capturing users' perceptions of AI explanations remain limited. Addressing this gap, the present study tried to develop and validate a multidimensional measurement scale designed to assess Artificial Intelligence Explanation Effects (AIDES) in digital environments.

Drawing upon theoretical insights from explainable AI, algorithmic transparency, and trust in automation literature, the study conceptualized AI explanation effects through four distinct but

interrelated dimensions: Perceived Transparency and Understanding (PTU), Perceived Credibility and Honesty (PCH), Ethical Fairness and Accountability (EFA), and Trust Calibration and Behavioral Response (TCB). Using survey data collected from 289 participants who regularly interact with digital platforms and AI-generated content, the proposed scale was empirically evaluated through confirmatory factor analysis and reliability assessments within a structural equation modeling framework.

The results provide encouraging evidence for the psychometric robustness of the proposed measurement instrument. Composite reliability values exceeded recommended thresholds for all constructs, indicating strong internal consistency. Convergent validity was supported through average variance extracted and reliability indicators, while discriminant validity was confirmed using the heterotrait–monotrait ratio criterion. These findings suggest that the four dimensions capture distinct aspects of how individuals interpret and respond to AI-generated explanations. Although the initial measurement model revealed some limitations in overall model fit, diagnostic analyses indicated that refinement of specific indicators—particularly within the trust calibration dimension—may further strengthen the scale in future applications.

The proposed AIDES scale contributes to the emerging literature on explainable artificial intelligence by providing a structured framework for empirically assessing user perceptions of AI explanations. Rather than conceptualizing explainability as a purely technical property of algorithms, the scale emphasizes the human-centered evaluation of AI explanations, capturing cognitive understanding, perceived credibility, ethical evaluation, and behavioral trust calibration. In doing so, it offers a comprehensive perspective on how explanations influence user trust and acceptance of AI systems.

From a practical perspective, the scale offers valuable insights for designers of AI-enabled platforms, digital services, and automated decision systems. Organizations increasingly rely on algorithmic recommendations in domains such as online advertising. The findings suggest that user acceptance of such systems depends not only on algorithmic accuracy but also on the perceived clarity, honesty, and ethical responsibility communicated through AI explanations.

By measuring these dimensions systematically, the proposed instrument can help organizations evaluate the effectiveness of explanation mechanisms and identify areas where transparency and trust-building strategies can be improved.

Several limitations should be acknowledged. First, the study relied on a cross-sectional survey design, which limits causal inference. Future research may employ experimental or longitudinal designs to examine how AI explanations influence trust and behavioral responses over time. Second, the sample consisted primarily of digitally active individuals with relatively high exposure to online platforms, which may limit the generalizability of findings to populations with lower levels of technological familiarity. Third, the measurement model indicated opportunities for further refinement, suggesting that additional scale purification and cross-validation across different contexts would strengthen the robustness of the instrument.

Future research can extend the present study in several ways. Researchers may examine the proposed scale across different AI application domains such as recommender systems, generative AI platforms, and autonomous decision systems. Additionally, integrating the scale with broader models of technology acceptance, algorithmic trust, and user experience may provide deeper insights into how explanation mechanisms influence adoption and behavioral reliance on AI systems.

In conclusion, this study introduces and validates a novel measurement instrument designed to capture the multidimensional effects of AI explanations on user perceptions and behavioral responses. By operationalizing explainability from the user perspective, the proposed scale contributes to a deeper understanding of how transparency, credibility, ethics, and trust interact within human–AI interactions. As AI systems continue to expand across digital ecosystems, tools that systematically measure explanation effects will become increasingly essential for developing responsible, trustworthy, and user-centered artificial intelligence.

Referen□es

Alon-Barkat, S., & Busuioc, M. (2023). Human–AI interactions in public sector decision making: "Automation bias" and "selective adherence" to algorithmic advice. *Journal of Public Administration Research and Theory*, *33*(1), 153–169. https://doi.org/10.1093/jopart/muac024

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological bulletin*, *103*(3), 411.

Bansal, G., Nushi, B., Kamar, E., Lasecki, W. S., Weld, D. S., & Horvitz, E. (2021). Beyond accuracy: The role of mental models in human–AI team performance. Proceedings of the AAAI Conference on Human Computation and Crowdsourcing,

Chandrasekaran, B., Prabhu, V., & Saif, I. (2023). Contrastive versus feature-based explanations: A comparative study on user comprehension and trust in AI. *International Journal of Human–Computer Interaction*, *39*(15), 3125–3142. https://doi.org/10.1080/10447318.2023.2182417

Cheng, X., & Jiang, H. (2024). Tailoring explainable AI to user values: The impact of fairness foci on explanation preference and fairness perception. Proceedings of the ACM on Human–Computer Interaction,

Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*.

Eiband, M., Schneider, H., Bilandzic, M., Fazekas-Con, J., Haug, M., & Hussmann, H. (2022). Bringing transparency design into practice: A case study on explanations for a recommendation system. Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems,

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*, *18*(1), 39–50.

Green, B., & Chen, Y. (2019). The principles and limits of algorithm-in-the-loop decision making. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems,

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). Multivariate data analysis.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the academy of marketing science*, *43*(1), 115–135.

Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2021). Metrics for explainable AI: Challenges and prospects. *Human Factors*, *63*(5), 724–750. https://doi.org/10.1177/0018720820961375

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1–55.

Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. (2021). Formalizing trust in artificial intelligence: Prerequisites, causes, and goals of human trust in AI. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency,

Jussupow, E., Benbasat, I., & Heinzl, A. (2021). Why are we averse to algorithms? A comprehensive literature review on algorithm aversion. *Journal of the Association for Information Systems*, *22*(4), 1026–1064. https://doi.org/10.17705/1jais.00681

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50.30392

Liao, Q. V., & Varshney, K. R. (2023). Human-centered explainable AI (XAI): From algorithms to user experiences. *Foundations and Trends in Human–Computer Interaction*, *16*(2), 91–227. https://doi.org/10.1561/1100000082

Malhotra, N. K., & Dash, S. (2011). Marketing research an applied orientation (paperback). In: London: Pearson Publishing.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007

Shin, D. (2021). The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies*, *146*, 102551. https://doi.org/10.1016/j.ijhcs.2020.102551

Wagner, G. (2023). Liability Rules for the Digital Age: –Aiming for the Brussels Effect–. *Journal of European Tort Law*, *13*(3), 191–243.

Wang, X., & Yin, M. (2023). Are explanations helpful? A comparative study of the effects of explanations in AI-assisted decision-making. Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems,

Weitz, K., Lüders, M., & Starke, A. D. (2024). Measuring user trust and understanding in explainable AI: A systematic review of empirical methods. *Computers in Human Behavior Reports*, *15*, 100345. https://doi.org/10.1016/j.chbr.2024.100345

Zhang, Y., Liao, Q. V., & Bellamy, R. K. (2020). Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency,

CHAPTER IV

Trust Cal□brat□on □n AI-Created Ads

Yağmur Gül KURMUŞ

Abstra□t

As artificial intelligence increasingly generates and curates advertising content, understanding whether consumers trust such ads appropriately, neither too much nor too little, becomes critical for both consumer protection and market efficiency. This study introduces the concept of trust calibration to the AI advertising domain and develops a scale called the Trust Calibration toward AI Advertising scale to measure the alignment between consumers' trust and perceived performance of AI-generated ads. Based on a comprehensive literature review covering human-automation interaction, the persuasion knowledge model, and AI ethics, an initial scale consisting of twelve items across four theoretical dimensions was developed, including Cognitive Trust Alignment, Transparency and Disclosure Sensitivity, Ethical Governance and Accountability Perception, and Dynamic Trust Regulation. A separate subscale measuring Perceived Performance with seven items was also developed. Data were collected from four hundred thirty-five Turkish consumers via an online survey. Psychometric properties were assessed using composite reliability, average variance extracted, HTMT discriminant validity, and confirmatory factor analysis. The Calibration Index, computed as one minus the absolute difference between the mean scale score and the mean perceived performance score, divided by seven, provides a quantitative metric to diagnose over-trust and under-trust. While discriminant validity was acceptable and the mean Calibration Index indicated reasonably good average calibration, the confirmatory factor analysis model fit was poor. Average variance extracted fell below the acceptable threshold for Cognitive Trust Alignment and Perceived Performance. High inter-factor correlations, particularly between Cognitive Trust Alignment and Perceived Performance, suggest that the proposed four dimensions may not be empirically separable. This study is the first to operationalize trust calibration specifically for AI-generated advertising. Despite the poor factorial fit, the findings offer a foundation for scale refinement, highlight the dominant role of perceived performance in trust calibration, and demonstrate the importance of ethical governance and transparency as contextual antecedents. Future research should revise the scale, test it experimentally, and link calibration to behavioral outcomes.

1. Introduction

The proliferation of artificial intelligence (AI) as a generative and curatorial force in marketing has fundamentally altered the advertising landscape. From dynamically crafted copy and personalized product recommendations to synthetically generated endorsers, AI systems now

play a pivotal role in shaping consumer-brand interactions. This shift necessitates a critical examination of the human factor within this new paradigm: specifically, the nature and appropriateness of consumer trust. While trust has long been a cornerstone of effective marketing communications, the opaque, autonomous, and data-driven nature of AI agents complicates traditional trust models. Merely measuring levels of trust is insufficient; understanding whether that trust is appropriately calibrated to the actual capabilities and reliability of the AI system is paramount for both ethical consumer protection and optimal market efficiency.

The concept of trust calibration originates not in marketing, but in the human factors and human-computer interaction (HCI) literature concerning automation and decision-support systems. It sits at the intersection of trust theory, behavioral reliance, and meta-cognitive confidence. Seminal work by Muir (1994) established the foundational concern that trust must be calibrated to a system's reliability to ensure appropriate use – where overtrust can lead to catastrophic complacency and undertrust to the disuse of a beneficial tool. Lee and See (2004) further formalized this by framing calibration as the congruence between a user's subjective trust and the system's objective performance, introducing the critical states of overtrust (trust exceeding capability) and undertrust (capability exceeding trust). This theoretical framework has been operationalized through scales measuring the appropriateness of reliance Merritt and Ilgen (2008) and methodologies comparing perceived versus actual system reliability to calculate a calibration error Dzindolet et al. (2003). Subsequent meta-frameworks, such as that by Hoff and Bashir (2015), have further dissected trust into dispositional, situational, and learned components, all of which inform the dynamic calibration process. For decades, this research stream has been rigorously applied in high-stakes domains like aviation, healthcare, and military systems, where miscalibrated trust carries immediate and severe consequences.

Despite its maturity in automation research, the study of trust calibration remains nascent in consumer-facing domains, particularly in marketing and advertising. Extant studies on AI in advertising have productively explored initial acceptance, perceived creepiness, personalization paradoxes, and general trust attitudes. However, they have largely treated trust as a static outcome or a moderating variable, not as a dynamic judgment that requires continuous calibration against an often-unseen algorithmic performance. This constitutes a significant

theoretical and empirical gap. The advertising context introduces unique complexities – such as persuasive intent, aesthetic and emotional appeal, source ambiguity, and ethical concerns over data use and manipulation – that are absent in most automated control systems. A consumer's trust in an AI-generated ad must be calibrated not just to its factual accuracy, but also to its persuasive sincerity, personal relevance, and alignment with normative expectations.

Therefore, there is a pressing need to develop and validate a domain-specific trust calibration scale for AI-created advertisements. Adapting the robust methodologies from HCI to the nuanced reality of marketing communications will provide researchers with a critical tool. Such a scale would move beyond asking *if* a consumer trusts AI ads, to assessing *how well* their trust aligns with the AI's actual performance in terms of accuracy, relevance, transparency, and ethical compliance. This enables the diagnosis of harmful miscalibration – whether it is the consumer vulnerability born of overtrust in deceptive AI-generated claims, or the market inefficiency resulting from undertrust in genuinely helpful AI recommendations. Filling this gap is essential for advancing theory on human-AI interaction in persuasive contexts and for providing actionable insights to policymakers, platform designers, and marketers committed to fostering a trustworthy and effective digital marketplace.

2. Literature Review

2.1 The Role of Transparency and Disclosure in Trust Calibration

Transparency, particularly in the form of disclosing the involvement and workings of Artificial Intelligence (AI), is widely theorized as a foundational mechanism for calibrating user trust in algorithmic systems. In the context of AI-generated advertising, transparency serves not merely as an ethical imperative but as a critical cognitive tool that allows consumers to align their trust with the actual capabilities and intentions of the system (Shin, 2021). This alignment is essential to mitigate the risks of overtrust, which can lead to uncritical acceptance, and undertrust, which results in the rejection of potentially useful content (Lee & See, 2004).

The primary function of transparency in this domain is to pierce the "black box" of algorithmic decision-making. AI systems often operate through complex models that are opaque to end-users, creating a sense of uncertainty and undermining the formation of calibrated trust. Explainable AI (XAI) and clear disclosures are posited as solutions to this problem. Shin (2021)

indicates that when users can understand how and why an AI system makes a decision – a concept encompassing both technical explainability and perceived causability – their trust and adoption intentions increase significantly. This suggests that transparency provides the necessary evidence for users to make an informed judgment about system reliability, moving trust from blind faith to a calibrated assessment.

A direct application of transparency in marketing is the AI disclosure – explicitly stating that content is generated by AI. The effects of such disclosures are complex and moderated by the Persuasion Knowledge Model (Friestad & Wright, 1994). When consumers recognise an AI agent as a persuasive source, they activate their persuasion knowledge, potentially triggering skepticism. Empirical evidence supports this: Luo et al. (2019) found that disclosing chatbot identity at the onset of an interaction led to a dramatic decrease in purchase rates, as the disclosure cued users to the non-human, potentially manipulative nature of the interaction. This highlights a critical challenge: while transparency is ethically required and theoretically beneficial for long-term trust calibration, its immediate effect can be a reduction in persuasive effectiveness by making the persuasive intent salient (Campbell & Grimm, 2019).

The design of the disclosure itself is a crucial factor. Studies suggest that the effectiveness of a disclosure for trust calibration depends on its prominence, clarity, and timing. Vague or buried disclosures may fail to serve their calibrating function and can even backfire by heightening suspicion if discovered later (Eisend & Küster, 2011). In contrast, clear, upfront disclosures, while potentially dampening short-term conversion, act as a signal of honesty. This proactive transparency can prevent more severe trust violations and brand damage that occur when consumers feel deliberately deceived by discovering AI involvement post-hoc (Campbell et al., 2022).

However, the relationship between transparency and trust is not linear, giving rise to what scholars term the “transparency paradox” (Glikson & Woolley, 2020). Excessive or poorly designed transparency – such as overwhelming users with technical details about the AI’s inner workings – can increase cognitive load, confuse users, and ultimately erode trust. This paradox underscores a key distinction: not all transparency is equal. Research indicates that transparency

about the purpose, intent, and high-level logic of a system (intent transparency) is generally more effective for building and calibrating trust than granular technical transparency (Glikson & Woolley, 2020; Rai, 2020). For instance, explaining why a particular ad was selected for a user (e.g., "based on your interest in sustainable products") is more calibrating than explaining the technical details of the clustering algorithm used.

Furthermore, transparency extends beyond the mere fact of AI involvement to encompass data practices. In an era of personalized advertising, consumers are concerned with how their data are collected and used. Puntoni et al. (2021) argue that transparency about data use can alleviate feelings of surveillance and creepiness, helping to balance the perceived benefits of personalization with privacy concerns. Thus, comprehensive transparency that addresses both the agent (AI) and the fuel (user data) is likely to be most effective for holistic trust calibration.

In summary, the literature establishes transparency and disclosure as a double-edged sword in the calibration of trust toward AI advertising. It is an essential ethical and functional component for building appropriate, evidence-based trust and fulfilling the consumer's "right to know". Yet, its implementation must be strategically managed to navigate the immediate activation of persuasion knowledge, avoid cognitive overload, and focus on communicating intent and relevance.

2.2 Ethical Governance and Algorithmic Accountability as Trust Antecedents

Following the immediate calibrating role of transparency, the literature establishes that sustainable, calibrated trust in AI advertising is predicated on deeper systemic foundations: ethical governance and algorithmic accountability. While transparency provides the information necessary for calibration, ethical governance provides the assurance that the system operates within acceptable moral and normative boundaries (Floridi et al., 2018). This dimension addresses the consumer's fundamental need to know not only *how* an AI works but also that it is designed and deployed responsibly.

Consumer trust is increasingly contingent upon the perception of ethical oversight. As Davenport et al. (2020) argue, in the data-driven marketing landscape, ethical governance transcends legal compliance to become a strategic component of brand trust. Consumers do not merely evaluate an AI ad's immediate relevance or creativity; they implicitly assess the ethical integrity of the system behind it. This is reflected in findings by Çeber and Çeliker (2024), which suggest that consumer perceptions and behaviors towards generative AI ads are shaped by an underlying assessment of the system's responsible deployment. When AI is perceived as a "black box" without ethical safeguards, it fosters uncertainty and undermines the potential for calibrated trust, as users lack a foundation upon which to base a reliable judgment of its intentions (Du & Xie, 2021).

A central ethical challenge eroding this foundation is algorithmic bias. The capacity for AI systems to perpetuate or amplify societal biases in targeting, representation, or messaging presents a profound threat to trust. Dwivedi et al. (2021) emphasize that biased algorithmic outputs can exacerbate social inequalities, making the demand for accountability not just an individual consumer issue but a societal imperative. This links directly to the concept of algorithmic justice, which Kietzmann et al. (2018) identify as a critical factor in consumer evaluations. Trust calibrates positively when systems are perceived as fair and just; conversely, the perception of bias leads to a rapid erosion of trust and a punitive response toward the associated brand (Martin & Zimmermann, 2024). This dynamic underscores that ethical governance is not passive but requires active bias mitigation and auditing.

This leads to the core antecedent of accountability. A persistent theme in the literature is the insufficiency of deflecting responsibility to the algorithm itself. As Hermann (2022) asserts, the defense that "the algorithm made the decision" is ethically and practically untenable. For trust to be calibrated, there must be a clear, identifiable locus of accountability – typically the deploying firm or organization. This principle, often termed "human-in-command" or "responsible AI", ensures that technological autonomy does not equate to a responsibility vacuum (Floridi et al., 2018; Gonçalves et al., 2023). Wirtz et al. (2018) further posit that clear accountability acts as a form of "insurance" for the consumer, mitigating perceived risk and the associated "fear of the machine", thereby creating a more stable base for trust calibration.

Accountability is operationalized through ethical frameworks and principles. Cross-disciplinary work has converged on key principles such as fairness, accountability, transparency, and explainability (often summarized as FATE). These principles provide a normative checklist against which consumers, either intuitively or explicitly, judge AI systems. Rai (2020) frames explainability (XAI) not just as a transparency tool but as an accountability mechanism, enabling the scrutiny necessary for ethical governance. When firms demonstrate adherence to such principled frameworks – for instance, by auditing algorithms for fairness or establishing ethics review boards – they send a powerful signal that engenders institutional trust (Shin, 2021). This institutional trust, as noted by McKnight et al. (2002), forms a critical initial layer upon which more specific, experience-based trust in a particular AI application can be built and calibrated.

In the specific context of advertising, accountability intersects with persuasion ethics. The use of AI for persuasion raises unique concerns about manipulation and autonomy. Boerman et al. (2017), in their review of online behavioral advertising, highlight that ethical data use and respect for consumer autonomy are prerequisites for sustainable trust. When consumers believe a firm is ethically managing their data and is accountable for its persuasive attempts, they are more likely to calibrate their trust appropriately, engaging with useful personalization while remaining guarded against perceived manipulation (Aguirre et al., 2015).

In summary, the literature positions ethical governance and algorithmic accountability not as optional additions but as fundamental antecedents to the trust calibration process. They provide the essential normative scaffolding that makes the information gleaned from transparency meaningful and the experiences gathered from interaction interpretable within a framework of safety and justice.

2.3 The Dynam□s of Trust Regulat□on and Cal□brat□on Over T□me

Moving beyond the foundational antecedents of transparency and ethics, the literature conceptualizes trust calibration not as a stable endpoint but as a continuous, dynamic process of regulation. Informed by decades of research in human-automation interaction, this

perspective views trust as a cognitive state that is perpetually updated in response to system performance, user experience, and evolving context (Lee & See, 2004). In the realm of AI advertising, this implies that a consumer's trust in an AI-generated ad or recommendation system is not fixed after the first disclosure or ethical evaluation but is continually fine-tuned through ongoing interaction.

The theoretical underpinning of this dynamic view originates from Muir (1994) seminal model of trust in automation, which framed trust as a belief that is constantly calibrated against observed system reliability. This model posits that appropriate reliance – the behavioral manifestation of calibrated trust – requires the user to engage in a continuous feedback loop, comparing expectations with outcomes. Hoff and Bashir (2015) expanded this into a multi-layered framework, distinguishing between dispositional, situational, and learned trust. It is this *learned trust* dimension that is explicitly dynamic, evolving as users accumulate experience with a specific system. In advertising, a consumer's learned trust in a platform's AI recommender develops through repeated exposures to either accurate, relevant ads (positive calibration) or irrelevant, intrusive ones (negative calibration).

This learning process is highly sensitive to performance errors, revealing a critical phenomenon known as *algorithm aversion*. Dietvorst et al. (2015) demonstrated that humans are often less tolerant of errors from algorithms than from other humans, leading to a sharp, disproportional decline in trust after observing an algorithmic mistake. For AI advertising, this means a single instance of poor personalization (e.g., an irrelevant product recommendation) can cause significant negative trust calibration, potentially undoing the benefits of prior positive experiences. However, this aversion can be mitigated by elements of transparency and control; when users understand why an error occurred or can adjust the system's parameters, the downward recalibration of trust is less severe and more recoverable (Dietvorst et al., 2015; Eslami et al., 2015).

The regulation of trust involves both cognitive and affective pathways. Glikson and Woolley (2020) note that trust in AI has distinct cognitive (rational assessment of competence) and affective (emotional feeling of security) components that can change at different rates. A rationally sound, high-performing AI ad might still fail to build trust if it triggers negative affective responses, such as feelings of creepiness or a loss of autonomy (Puntoni et al., 2021).

Thus, dynamic calibration is the process of reconciling these two streams of input. For instance, a highly personalized ad may be cognitively evaluated as competent (high performance), but if it triggers affective privacy concerns, the consumer may dynamically down-regulate their overall trust and reliance (Bleier & Eisenbeiss, 2015).

This process is further structured by the distinction between *initial trust* and *ongoing trust*. Initial trust is formed quickly based on surface cues, brand reputation, or institutional signals (McKnight et al., 2002). In contrast, ongoing trust is a more resilient state built and regulated through direct experience (McKnight et al., 2011). In AI advertising, a consumer might grant initial trust based on a platform's strong privacy policy (an ethical antecedent). This initial trust is then dynamically regulated – either reinforced or degraded – with each subsequent interaction with the AI's ad outputs.

A key facilitator of positive dynamic calibration is the user's perceived control and understanding of the system's logic. Research by Eslami et al. (2015) on social media algorithms found that when users gain insight into "why" they see certain content, they engage in a more nuanced recalibration of their trust and relationship with the platform. In advertising, providing users with adjustable preferences or clear explanations for ad targeting ("You're seeing this ad because…") can transform a passive experience into an interactive one, fostering a sense of agency that supports more stable and appropriately calibrated trust over time.

Finally, the dynamics of trust must account for repair and recovery. De Visser et al. (2018) discuss "trust resilience", or a system's capacity to recover trust after a failure. In AI advertising, a poorly calibrated ad could damage trust. However, subsequent accurate recommendations, coupled with perhaps an explanatory message, can initiate a repair process, demonstrating the system's capacity to learn and correct, thereby allowing trust to be recalibrated upward once more. This underscores that the calibration process is non-linear and bidirectional.

In conclusion, the literature firmly establishes trust calibration as a dynamic, temporal process. It is a journey of continuous adjustment, where initial judgments formed from transparency and

ethics are perpetually tested and revised against lived experience, emotional responses, and observed performance.

2.4 Perceived Performance and Competence in Trust Calibration

The most direct and potent antecedent of trust calibration is the user's perception of a system's performance and competence. While transparency, ethics, and dynamic regulation provide the context and process for calibration, it is the observed or believed capability of the AI that serves as the primary evidence against which trust is ultimately calibrated. In essence, trust calibrates toward perceived reliability; a user's confidence aligns with their assessment of how well the system executes its intended function (Jian et al., 2000). In AI advertising, this translates to consumers evaluating whether the AI successfully delivers relevant, useful, and persuasive content.

The foundational link between performance and trust is robust. Jian et al. (2000), in developing an empirical scale of trust in automation, identified performance – encompassing reliability, dependability, and competence – as a critical dimension. When users perceive a system to be consistently accurate and effective in fulfilling its tasks, their trust calibrates upward. This is closely related to the concept of *algorithm appreciation*, where users recognize and prefer algorithmic judgment over human judgment for specific, data-driven tasks (Logg et al., 2019). In advertising, this appreciation manifests when consumers perceive AI's targeting or creative generation as superior to human efforts, leading to higher calibrated trust and increased reliance. The UTAUT model supports this, identifying performance expectancy – the degree to which a technology will help achieve gains in job performance – as the strongest predictor of behavioral intention (Venkatesh et al., 2003). For a consumer, the "job" may be efficient product discovery, and an AI ad that demonstrably aids this goal will foster trust calibrated to its high perceived utility.

However, the perception of performance is fragile and subject to a stark asymmetry: algorithm aversion. As Dietvorst et al. (2015) established, people often abandon algorithmic tools after observing them err, displaying a lower tolerance for machine mistakes than for human errors. This has profound implications for trust calibration in advertising. A single instance of poor performance – such as a grossly mis-targeted ad, a factual inaccuracy in generated copy, or a

tone-deaf creative – can trigger a disproportionate downward recalibration of trust. The system's competence is thrown into question, and recovery requires significant subsequent proof of reliable performance. This aversion underscores that while strong positive performance builds trust gradually, negative performance can degrade it rapidly, making consistency paramount for stable calibration (Glikson & Woolley, 2020).

In the advertising context, perceived performance is evaluated across several key criteria:

1. **Targeting and Personalization Accuracy:** The core performance metric for many AI advertising systems is relevance. An ad's performance is judged by its perceived fit with the user's interests, needs, or context. Accurate personalization signals competence, while errors trigger aversion and distrust (Dietvorst et al., 2015). This creates the "personalization paradox", where the high performance needed for trust must be balanced against the privacy concerns that excessive personalization can provoke (Aguirre et al., 2015).
2. **Creative and Functional Competence:** Beyond targeting, users assess the intrinsic quality of the AI's output. This includes the linguistic fluency, visual appeal, logical coherence, and persuasive power of the generated ad. Research indicates that the perceived creative competence of AI contributes significantly to its overall performance evaluation and, by extension, to trust (Çeber & Çeliker, 2024; Kietzmann et al., 2018).
3. **Functional Utility in Interaction:** For interactive AI advertising tools (e.g., chatbots), performance is judged on task success: the ability to understand queries, provide accurate information, and resolve issues efficiently. Nordheim et al. (2019) found that a chatbot's functional performance is a direct and powerful driver of user trust.

Notably, perceived performance can sometimes overshadow the need for complete transparency. Glikson and Woolley (2020) note that when a system's outputs are consistently successful, users may care less about the "black box" nature of its operations. The observable results become a sufficient proxy for competence. This suggests a hierarchical relationship: for basic calibration, demonstrable performance may be paramount; for deeper, more resilient calibration – especially after errors – transparency and explainability become crucial to restore performance-based judgments (Shin, 2021).

In summary, perceived performance and competence constitute the bedrock evidence in the trust calibration process. It is the most immediate answer to the user's tacit question: "Does this work for me?" While ethical governance provides the warrant to trust, and transparency provides the insight to calibrate, it is the consistent, observable demonstration of competence that provides the substantive reason to do so.

3. Method

3.1 Study Design and Scale Development

This study aimed to develop and initially validate a scale for measuring trust calibration toward AI-generated advertisements (TCAIA). Following standard scale development procedures, an initial pool of 12 items was generated based on the literature review (item generation). Content validity was assessed through confirmatory factor analysis (CFA). No separate pilot test was conducted. The main survey data were subjected to CFA, reliability, and validity analyses.

3.2 Participants and Data Collection

Participants were recruited via an online survey. Eligibility criteria were age ≥ 18 years and having been exposed to online advertising. A total of 435 valid responses were obtained. Demographic characteristics are 208 women (47.8%), 191 men (43.9%), 35 preferred not to specify (8.0%), and 1 other (0.2%). Age mean = 30.02 years (SD = 9.73), range 11–96. The majority (62.1%) were aged 18–35 years. Education distribution is Bachelor's degree 46.9%, associate degree 19.5%, master's or higher 18.4%, high school or below 12.9%, others 2.3%. Daily social media use median is approximately 3–4 hours; wide range from <1 hour to 24 hours. Interaction with AI-generated content (ChatGPT, DALL-E, etc.) is frequently 26.7%, sometimes 32.0%, rarely 18.2%, very often 11.7%, never 10.1%. Frequency of encountering AI-generated ads: frequently 26.4%, sometimes 37.5%, rarely 17.9%, very often 11.0%, never 6.0%.

3.3 Procedure

Participants completed an online survey. They were first presented with a brief definition of AI-generated advertising and then asked to reflect on their general experience with such ads on social media, e-commerce platforms, or other digital channels. Following this, they completed the TCAIA scale (12 items) and the Perceived Performance subscale (7 items). All items were rated on a 7-point Likert scale (1 = Strongly Disagree, 7 = Strongly Agree). Demographic questions were asked at the end.

3.4 Measures

3.4.1 Trust Calibration toward AI Advertising (TCAIA)

The TCAIA consists of 12 items measuring four dimensions (three items each): **BGU** – Cognitive Trust Alignment (Accuracy & Competence); **SAD** – Transparency & Disclosure Sensitivity; **EY** – Ethical Governance & Accountability Perception; **DG** – Dynamic Trust Regulation (Calibration Behavior). A 7-item subscale measured perceived performance of AI-generated ads, including accuracy, consistency, professionalism, and clarity (AP1 to AP7). See Appendix for Turkish wording.

3.4.2 Calibration Index

Following the composite metric approach adapted from Dzindolet et al. (2003), a Calibration Index was computed for each participant:

Calibration Index = 1 – (|T – P| / 7)

where **T** = mean score of the 12 TCAIA items (overall trust) and **P** = mean score of the 7 AP items (perceived performance). Both T and P range from 1 to 7. Values near 1 indicate well-calibrated trust; values near 0 indicate miscalibration (over-trust if T > P, under-trust if P > T).

3.5 Data Analysis

Data were analyzed using SPSS and AMOS. Descriptive statistics were computed. Reliability was assessed using composite reliability (CR) and Cronbach's alpha (α). Convergent validity was evaluated with average variance extracted (AVE). Discriminant validity was examined

using the heterotrait-monotrait (HTMT) ratio. Confirmatory factor analysis (CFA) was performed to test the four-factor structure of the TCAIA, with fit assessed via χ^2/df, CFI, SRMR, RMSEA, and PClose. The Calibration Index was calculated and its distribution examined.

4. Results

4.1 Descriptive Statistics

The descriptive statistics for the overall trust (T), perceived performance (P), absolute difference, scaled difference, and Calibration Index are presented in Table 1.

Table 1. Descriptive Statistics for Calibration Components (N = 435)

Variable	Range	Min	Max	Mean	SD
P (Perceived Performance)	6.00	1.00	7.00	4.2417	1.14114
T (Overall Trust)	6.00	1.00	7.00	4.6157	1.12257
AbsDiff |T-P|	4.32	0.00	4.32	0.6562	0.75848
DiffScaled (AbsDiff/7)	0.62	0.00	0.62	0.0937	0.10835
Calibration Index	0.62	0.38	1.00	0.9063	0.10835

The mean Calibration Index of 0.906 indicates that, on average, participants' trust was well aligned with their perceived performance. However, the range (0.38–1.00) shows substantial individual variation.

4.2 Reliability and Convergent Validity

Composite reliability (CR) and average variance extracted (AVE) for each dimension are reported in Table 2, along with 95% confidence intervals.

Table 2. Reliability and Convergent Validity

Dimension	CR	95% CI for CR	AVE	95% CI for AVE
BGU (Cognitive Trust)	0.616	[0.478, 0.714]	0.349	[0.235, 0.455]
SAD (Transparency)	0.749	[0.657, 0.824]	0.499	[0.392, 0.609]

EY (Ethical Governance)	0.779	[0.697, 0.839]	0.540	[0.435, 0.635]
DG (Dynamic Regulation)	0.776	[0.695, 0.835]	0.538	[0.436, 0.630]
AP (Perceived Performance)	0.863	[0.804, 0.903]	0.474	[0.370, 0.571]

CR values were acceptable for most dimensions (≥ 0.75) except BGU (0.616). AVE fell below 0.50 for BGU (0.349) and AP (0.474), indicating limited convergent validity for these two dimensions.

4.3 Discriminant Validity (HTMT)

The HTMT matrix (Table 3) shows all values below the strict threshold of 0.85, with no warnings. Thus, discriminant validity is supported.

Table 3. HTMT Matrix

	BGU	SAD	EY	DG	AP
BGU	–				
SAD	0.582	–			
EY	0.541	0.652	–		
DG	0.530	0.658	0.772	–	
AP	0.722	0.464	0.551	0.530	–

4.4 Confirmatory Factor Analysis and Model Fit

The hypothesized four-factor model (BGU, SAD, EY, DG) together with the separate AP factor was tested using CFA. Fit indices are presented in Table 4.

Table 4. Model Fit Indices

Measure	Estimate	Threshold for good fit	Interpretation
χ^2 (df)	1184.58 (142)	–	–

χ^2/df	8.34	< 3	Poor
CFI	0.770	> 0.95	Poor
SRMR	0.091	< 0.08	Acceptable
RMSEA	0.130	< 0.06	Poor
PClose	0.000	> 0.05	Not acceptable

The model showed poor fit overall. Only SRMR approached acceptability. This indicates that the proposed four-factor structure does not adequately represent the data.

4.5 Calibration Index Distribution

Using the formula described in Section 3.4.3, the mean Calibration Index was 0.906 (SD = 0.108). Using a criterion of |T-P| ≤ 1 to define "well-calibrated", approximately 293 of the 435 participants (%67,36) were well-calibrated.

5. Discussion

5.1 Summary of Findings

This study developed the Trust Calibration toward AI Advertising (TCAIA) scale and a corresponding Calibration Index. The average Calibration Index of 0.906 suggests that, in general, Turkish consumers' trust in AI-generated ads is reasonably well aligned with their perceived performance of those ads. However, the CFA revealed poor model fit (CFI = 0.77, RMSEA = 0.130), and convergent validity was problematic for the Cognitive Trust Alignment (BGU) and Perceived Performance (AP) dimensions (AVE < 0.50). Discriminant validity was acceptable (HTMT < 0.85).

5.2 Interpretation of Poor Model Fit

Several factors likely contributed to the poor fit. First, high inter-factor correlations (e.g., BGU with AP = 0.992; EY with DG = 0.991) suggest that the four theoretical dimensions may not be empirically separable in consumers' minds. Cognitive trust alignment and perceived performance appear to be almost identical constructs in this sample. Similarly, ethical governance perception and dynamic trust regulation may reflect a single underlying attitude toward responsible AI.

Second, low AVE for BGU and AP indicates that the items within these dimensions share little common variance, possibly due to weak item wording or the abstract nature of the constructs. Third, model misspecification is likely. A unidimensional or a two-factor model (e.g., performance-based trust vs. governance-based trust) might fit better. Finally, common method bias (self-report, single source) and the lack of a specific ad stimulus (participants reflected on general experiences) may have inflated correlations and distorted the factor structure.

5.3 Theoretical Implications

Despite the measurement challenges, the study makes important theoretical contributions. First, it successfully transfers the concept of trust calibration from high-stakes automation to the consumer advertising domain, demonstrating that consumers can meaningfully report on the alignment between their trust and ad performance. Second, the Calibration Index offers a quantifiable, continuous measure of miscalibration, allowing future research to diagnose over-trust and under-trust. Third, the strong empirical link between perceived performance and cognitive trust suggests that performance may be the *primary* driver of calibration, with transparency and ethics playing supportive, moderating roles – a finding that aligns with the UTAUT model (Venkatesh et al., 2003).

5.4 Practical Implications

For marketers and advertising platforms, the high average Calibration Index is reassuring: most consumers are not blindly trusting or dismissing AI ads. However, the presence of miscalibrated individuals (CalIndex as low as 0.38) signals a need for targeted calibration interventions. Providing clear, simple explanations of why an ad is shown ("because you searched for X") and offering users control over their ad preferences could help reduce over-trust and under-trust. Moreover, the close relationship between perceived performance and trust implies that improving ad quality (relevance, accuracy, creativity) is the most direct path to building calibrated trust. Ethical governance, while important, may be a necessary but not sufficient condition.

5.5 Limitations

This study has several limitations. (1) Single, non-experimental design – participants reported on general experiences rather than responding to controlled stimuli. (2) Common method bias – all measures were self-reported. (3) Sample – convenience-based, predominantly young and educated, limiting generalizability. (4) Cross-sectional – cannot capture dynamic calibration over time. (5) Poor model fit – the TCAIA scale in its current form should not be used as a multidimensional measure without revision. (6) No objective performance benchmark – the Calibration Index relies on perceived rather than actual AI performance.

5.6 Future Research

Future research should: (a) revise the TCAIA – delete or reword items with low loadings, merge overlapping dimensions (e.g., BGU and AP, EY and DG), and test a shorter, unidimensional or two-factor model; (b) use experimental designs – manipulate ad performance (high/low accuracy) and transparency (disclosure present/absent) to examine causal effects on calibration; (c) adopt longitudinal methods – track calibration changes after repeated exposures; (d) include behavioral outcomes – link the Calibration Index to click-through, purchase, or ad avoidance; (e) replicate cross-culturally – test measurement invariance across different countries; (f) compare perceived vs. objective performance – use expert ratings to compute a true calibration error.

6. Conclusion

As artificial intelligence becomes deeply embedded in advertising creation and targeting, understanding whether consumers trust AI ads *appropriately* – neither too much nor too little – is a pressing ethical and managerial concern. This study introduced the concept of trust calibration to the AI advertising domain and developed the Trust Calibration toward AI Advertising (TCAIA) scale and a composite Calibration Index.

While the proposed four-dimensional structure did not achieve satisfactory confirmatory fit, the study provides a foundational empirical step. The Calibration Index (mean 0.906) demonstrates that, on average, Turkish consumers are reasonably well calibrated, yet individual differences highlight the risk of both over-trust and under-trust. The high correlations between cognitive

trust and perceived performance suggest that performance may be the dominant driver of calibration, with transparency and ethics playing supportive roles.

For researchers, the findings call for scale refinement and experimental validation. For practitioners, the Calibration Index offers a practical metric to monitor consumer trust health and design interventions that help users develop appropriate reliance on AI-generated advertising. Ultimately, fostering calibrated trust is not only a matter of consumer protection but also a strategic necessity for sustainable adoption of AI in marketing.

Referen□es

Aguirre, E., Mahr, D., Grewal, D., de Ruyter, K., & Wetzels, M. (2015). Unraveling the personalization paradox: The effect of information collection and trust-building strategies on online advertisement effectiveness. *Journal of Retailing*, *91*(1), 34–49.

Bleier, A., & Eisenbeiss, M. (2015). Personalized online advertising effectiveness: The interplay of what, when, and where. *Marketing Science*, *34*(5), 669–688.

Boerman, S. C., Kruikemeier, S., & Zuiderveen Borgesius, F. J. (2017). Online behavioral advertising: A literature review and research agenda. *Journal of Advertising*, *46*(3), 363–376.

Campbell, C., & Grimm, P. E. (2019). The challenges native advertising poses: Exploring potential Federal Trade Commission responses and identifying research needs. *Journal of Public Policy & Marketing*, *38*(1), 110–123.

Campbell, C., Sands, S., Ferraro, C., Tsao, H. Y., & Mavrommatis, A. (2022). From data to action: How marketers can leverage AI. *Business Horizons*, *65*(2), 227–243.

Çeber, B., & Çeliker, S. (2024). Tüketicilerin üretken yapay zekâ uygulamaları ile oluşturulan reklamlara yönelik algı ve davranışları üzerine bir saha araştırması. *Akdeniz Üniversitesi İletişim Fakültesi Dergisi*(47), 289–313.

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

De Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A. B., McKnight, P. E., & Parasuraman, R. (2018). Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied*, *24*(3), 290–305.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114–126.

Du, S., & Xie, C. (2021). Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities. *Journal of Business Research*, *129*, 961–974.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., & Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, *57*, 101994.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, *58*(6), 697–718.

Eisend, M., & Küster, F. (2011). The effectiveness of publicity versus advertising: A meta-analytic investigation of its moderators. *Journal of the Academy of Marketing Science*, *39*(6), 906–921.

Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., Hamilton, K., & Sandvig, C. (2015). "I always assumed that I wasn't really that close to [her]": Reasoning about invisible algorithms in news feeds.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707.

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

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627–660.

Gonçalves, A. R., Pinto, D. C., Rita, P., & Pires, T. (2023). Artificial intelligence and its ethical implications for marketing. *Emerging Science Journal*, *7*(2), 313–327.

Hermann, E. (2022). Leveraging artificial intelligence in marketing for social good—An ethical perspective. *Journal of Business Ethics*, *179*(1), 43–61.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407–434.

Jian, J. Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, *4*(1), 53–71.

Kietzmann, J., Paschen, J., & Treen, E. (2018). Artificial intelligence in advertising: How marketers can leverage artificial intelligence along the consumer journey. *Journal of Advertising Research*, *58*(3), 263–267.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90–103.

Luo, X., Tong, S., Fang, Z., & Qu, Z. (2019). Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Science*, *38*(6), 937–947.

Martin, K. D., & Zimmermann, J. (2024). Ethical implications of data aggregation and algorithmic targeting in marketing. *Journal of Consumer Marketing*, *36*(6), 841–850.

McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and the multi-dimensional structure of trust evolution. *ACM Transactions on Management Information Systems*, *2*(2), Article 12.

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research*, *13*(3), 334–359.

Merritt, S. M., & Ilgen, D. R. (2008). Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. *Human Factors*, *50*(2), 194–210.

Muir, B. M. (1994). Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. *Ergonomics*, *37*(11), 1905–1922.

Nordheim, C. B., Følstad, A., & Bjørkli, C. A. (2019). An initial model of trust in chatbots for customer service—Findings from a questionnaire study. *Interacting with Computers*, *31*(3), 236–250.

Puntoni, S., Reczek, R. W., Giesler, M., & Botti, S. (2021). Consumers and artificial intelligence: An experiential perspective. *Journal of Marketing*, *85*(1), 131–151.

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, *48*(1), 137–141.

Shin, D. (2021). The effects of explainability and causability on perception, trust, and adoption of explainable artificial intelligence. *International Journal of Human-Computer Studies*, *146*, 102551.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, *27*(3), 425–478.

Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave new world: Service robots in the frontline. *Journal of Service Management*, *29*(5), 907–931.

Appendix – Turkish Scale (TCAIA and Perceived Performance)

Yapay Zekâ Reklamlarına Yönelik Güven Kalibrasyonu Ölçeği (TCAIA)

7'li Likert: 1 = Kesinlikle Katılmıyorum, 7 = Kesinlikle Katılıyorum

Boyut 1: Bilişsel Güven Uyumu (BGU)

1. YZ tarafından oluşturulan reklamların tanıttıkları ürünler hakkında doğru iddialarda bulunduğuna inanırım.
2. Bir YZ reklamının güvenilir bilgi sağladığını genellikle anlayabilirim.
3. YZ reklamlarına olan güvenim, onların mantıksal tutarlılığına ve bilgi düzeyine bağlıdır.

Boyut 2: Şeffaflık ve Açıklama Duyarlılığı (SAD)

4. Bir reklamın YZ tarafından oluşturulduğu açıkça belirtildiğinde, güven seviyem buna göre değişir.
5. YZ'nin mesajı nasıl oluşturduğunu anladığımda, reklama daha çok güvenirim.
6. YZ'nin katılımı hakkında açık bilgi verilmesi, reklam içeriğine ne kadar güveneceğime karar vermeme yardımcı olur.

Boyut 3: Etik Yönetim ve Hesap Verebilirlik Algısı (EY)

7. Etik denetime tabi olan YZ reklamlarına daha fazla güvenirim.
8. Tanıdık reklam standartlarını veya düzenlemelerini izleyen YZ reklamları bana güven verir.
9. Veri kaynaklarını açıklayan YZ sistemleri tarafından üretilen reklamlar bana daha güvenilir görünür.

Boyut 4: Dinamik Güven Düzenlemesi (DG)

10. YZ reklamlarına olan güvenim, onların üslubu, açıklığı ve şeffaflığına bağlı olarak değişir.
11. YZ reklamları hakkında daha fazla şey öğrendikçe güvenim zamanla artabilir veya azalabilir.
12. YZ reklamlarının güvenimi hak edip etmediğini anlamak için onları aktif olarak değerlendiririm.

Algılanan Performans (AP) – Alt Ölçek

Yukarıdaki reklamı/deneyimi göz önünde bulundurarak (genel olarak YZ reklamları için):

13. Yapay zekâ tarafından oluşturulan bu reklam, ürün veya hizmet hakkında doğru bilgiler sunmaktadır.
14. Reklamın içeriği markanın kimliği ve mesajıyla tutarlıdır.
15. Bu reklamın sunduğu bilgiler ürünle ilgili beklentilerimi karşılamaktadır.
16. Yapay zekâ, reklamın içeriğini oluştururken doğru verileri kullanmıştır.
17. Bu reklamın teknik olarak düzgün ve hatasız biçimde üretildiğini düşünüyorum.
18. Reklamın genel sunumu (görsel, dil, biçim) profesyonel ve başarılıdır.
19. Reklamın hedeflediği mesaj açık ve anlaşılır biçimde iletilmiştir.

CHAPTER V

Ethical Governance in AI Marketing

Yusuf TARIBAKAN

Abstract

The rapid integration of artificial intelligence (AI) into marketing and advertising has transformed brand-consumer interactions but introduced complex ethical challenges requiring structured governance. This study aims to develop and validate a multidimensional measurement instrument—the Ethical Governance in AI Marketing (EG-AIM) scale—focused on five critical dimensions: Accountability & Responsibility (AR), Transparency & Explainability (TE), Ethical Oversight & Compliance (EOC), Stakeholder Inclusiveness & Feedback (SIF), and Monitoring & Risk Management (MRM). Using survey data from 400 participants, the scale was evaluated through confirmatory factor analysis, reliability assessments, and validity tests. Results indicate satisfactory internal consistency across all dimensions, with composite reliability values. Convergent validity was largely supported, with four dimensions achieving average variance extracted above 0.50, and discriminant validity was robustly confirmed via the heterotrait-monotrait criterion (all values < 0.85). However, overall model fit was suboptimal, and the Fornell-Larcker criterion was not fully satisfied, suggesting high inter-factor correlations and the need for further refinement. The EG-AIM scale represents a promising but not yet fully optimized instrument for operationalizing ethical governance in AI-driven marketing, offering a theoretically grounded foundation for future academic, managerial, and regulatory applications while warranting cautious use and continued validation.

1. Introduction

The rapid integration of artificial intelligence (AI) into marketing and advertising has fundamentally transformed how brands interact with consumers. From AI-generated advertising content and virtual influencers to algorithmic recommendation systems and hyper-

personalized targeting, the marketing landscape has entered an era of unprecedented automation and sophistication. However, this technological evolution has simultaneously introduced complex ethical challenges that demand rigorous scholarly attention.

The concept of ethical governance in AI marketing encompasses the frameworks, principles, and mechanisms that ensure AI-driven marketing practices align with societal values, legal requirements, and consumer expectations. As AI systems become increasingly autonomous in making decisions that affect consumer welfare—from what advertisements individuals see to how their personal data is utilized—the need for structured governance mechanisms has become paramount.

This paper aims to research ethical governance in AI marketing, with a particular focus on five critical dimensions: Accountability & Responsibility (AR), Transparency & Explainability (TE), Ethical Oversight & Compliance (EOC), Stakeholder Inclusiveness & Feedback (SIF), Monitoring & Risk Management (MRM). These dimensions collectively form the foundation for developing and validating a measurement instrument—the Ethical Governance in AI Marketing Scale.

2. Development of the Ethical Governance in AI Marketing Scale.

2.1. Accountability & Responsibility (AR)

Accountability represents a foundational pillar of ethical governance, establishing clear lines of responsibility for AI-driven marketing outcomes. The literature consistently emphasizes that accountability extends beyond mere compliance to encompass proactive responsibility for AI system behavior and its consequences. Accountability in AI marketing extends beyond disclosure to encompass the entire algorithmic value chain. Jaiswal (2024) emphasizes that accountability requires oversight of training data reliability and neutrality, ensuring that AI models do not reproduce racial, gender, or cultural biases. The Fair, Accountable, and Transparent (FATE) AI framework has become critical for maintaining ethical standards in digital marketing. Accountability mandates that brands bear direct responsibility for erroneous or misleading content generated by AI tools they employ. This necessitates clear protocols defining responsibility-sharing between agencies and brands. Chen et al. (2024) identified

perceived risk as a significant factor in consumer attitudes toward AI advertisements. Accountability mechanisms—such as verification certificates or transparency reports—reduce this risk perception. Brands that proactively commit to compensation procedures for potential data breaches or algorithmic errors demonstrate accountability beyond minimum requirements.

Research demonstrates that consumer trust in AI-generated advertising is fundamentally shaped by the explicit disclosure of AI involvement. Ashraf et al. (2024) found that disclosing AI-generated content signals respect for consumer autonomy and contributes to perceptions of brand honesty and ethical behavior, ultimately supporting long-term brand loyalty. This finding aligns with the broader principle that transparency about AI's role enables consumers to make informed judgments about the content they encounter.

The labeling of AI-generated content activates consumers' cognitive defense mechanisms. Buder (2024) revealed that when consumers see "AI-Generated" labels, they evaluate content more critically, yet paradoxically, their respect for brand honesty increases. This tension suggests that while accountability mechanisms may temporarily increase scrutiny, they ultimately build sustainable trust.

The accountability principle manifests differently across market segments. To et al. (2025) identified a paradox in luxury marketing: when consumers learn that luxury brand advertisements are AI-generated, their attitudes become negative, perceiving AI use as "low effort" that damages brand authenticity. This finding highlights that accountability disclosure can harm brand image in certain contexts, yet ethical considerations still mandate such transparency. In creative processes, Yıldırım and Can (2024) argue that ethical and legal responsibility for advertising content belongs to the human marketer who uses and approves AI tools, not the AI itself. The human-in-command principle remains central to accountability frameworks.

2.2. Transparency & Explainability (TE)

Transparency and explainability constitute the second essential dimension of ethical governance, addressing consumers' right to understand AI systems' operations and decision-making processes. Hacıhasanoğlu and Akgün (2025) establish that transparency requires identity disclosure—consumers have the right to know whether content is produced by humans or algorithms. This transparency eliminates feelings of deception and increases brand trust. Conversely, non-transparent brands face long-term reputational damage.

Çeber (2024) found that advertising professionals agree on the necessity of labeling AI-generated content components as a matter of professional integrity. This labeling prevents consumer perception manipulation and maintains industry standards.

Parmar and Saran (2025) demonstrate that explainable AI (XAI) approaches—where consumers understand why specific products are recommended—eliminate the "black box" uncertainty of algorithmic systems. Research confirms that AI systems explaining their decision logic are perceived as fairer and more trustworthy, directly increasing customer loyalty. The "black box" problem in AI advertising creates significant consumer anxiety. Turan (2024) argues that transparency encompasses not only outcomes but also algorithmic functioning. Consumers must know which data (location, purchase history, etc.) inform product recommendations; without this information, consumer autonomy is violated, transforming advertising from information to manipulation. The European Union's AI Act establishes labeling requirements as a legal obligation (European, 2024), advertisements containing deepfakes or synthetic voices must be clearly marked for consumer recognition. This legal transparency prevents unfair competition while protecting consumers' right to information.

Ceviz (2025) identifies data security and algorithmic transparency as primary consumer concerns regarding AI-developed advertisements. Literature reviews demonstrate that consumers develop defensive mechanisms against brands when unaware of how their personal data is processed for advertising targeting. True transparency involves not merely providing legal texts but communicating data usage purposes simply and honestly to consumers. Amos and Zhang (2024) found that initial consumer discomfort with disclosed AI-generated content is balanced by appreciation of brand honesty, positively affecting purchase intention.

Transparency serves as a mechanism balancing AI's "artificiality" perception with "trustworthy brand" perception. Chen et al. (2024) discovered that when consumers know the source of AI-generated ads, they process content more effectively through their own filters, enhancing credibility. Ambiguous source content encounters cognitive resistance. Acar (2024) emphasizes that accountability and transparency are complementary; without transparency, accountability cannot function, and without accountability, trust cannot be established. Advertisers who acknowledge potential AI errors and openly communicate these processes achieve greater success in crisis management.

2.3. Ethical Oversight & Compliance (EOC),

The third dimension encompasses the systematic oversight mechanisms ensuring AI marketing practices adhere to ethical standards and regulatory requirements. Çeber (2024) argues that AI use in advertising mandates establishing ethical audit mechanisms. Agencies must create internal audit units to verify content accuracy and ethical standard compliance. AI hallucinations producing false information or misleading visuals risk brand reputation and legal consequences. Human oversight remains indispensable to compliance processes. Yıldırım and Can (2024) extend this requirement to AI-written advertising scripts, arguing that oversight must encompass data accuracy and cultural appropriateness testing. While automated compliance tools advance, AI remains insufficient for interpreting humor, irony, or cultural codes. Ethical committees or expert editors must serve as final approval authorities.

Ceviz (2025) describes how "accountability audits" in AI strategies have become new metrics for corporate sustainability. Consumers demand assurance that brands use fair, unbiased AI models that do not harm the digital ecosystem. Independent algorithm audits function as "ethical labels" certifying brand reliability. Gogoi (2025) proposes integrating Ethical AI Frameworks (EAAF) into corporate internal audit processes. These frameworks examine not only technical correctness of AI decisions—such as dynamic ad targeting—but also compliance with fairness, explainability, and accountability principles. Audit literature emphasizes that automated controls without human oversight remain insufficient, overlooking ethical risks including bias and discrimination.

Qazim (2025) highlight that AI systems in e-commerce and digital marketing must fully comply with international regulations including GDPR and the EU AI Act. Ethical compliance transcends avoiding penalties to respecting consumers' digital rights. Businesses should regularly audit recommendation algorithms for neutrality and transparently share audit reports. Grace (2024) emphasize that under consumer protection laws, AI-supported advertising must avoid "misleading commercial practices." Techniques such as personalized pricing or sentiment analysis exploiting consumer vulnerabilities are unethical. Compliance processes should incorporate "ethical firewalls" detecting and filtering manipulative algorithms. The European Commission mandates that AI applications classified as "high-risk"—including targeting children or biometric classification—face strict limitations in advertising (European, 2024). Compliance departments must pre-determine AI tools' risk levels in advertising campaigns and prepare mandatory impact assessment reports. Gölgeli (2025) notes that as AI's role in ad design increases, copyright and intellectual property compliance become complex. Audit mechanisms must verify that generated visuals do not infringe upon artists' rights used in training data. Ethical compliance policies should promote licensed, authorized datasets that respect creative labor rather than exploiting legal loopholes.

Arıcı (2025) frames digital advertising ecosystems as modern surveillance and audit mechanisms. Platforms like Google and Meta not only monitor user behavior but algorithmically discipline it. Ethical audits serve as filters determining whether brands abuse this surveillance power for manipulation. Compliance processes should minimize advertising's "coercive" effects and guarantee transparent data policies.

2.4. Stakeholder In□lus□veness & Feedba□k (SIF),

The fourth dimension addresses the principle that AI marketing systems should consider and incorporate the interests of all affected parties, not merely shareholders. Rane (2024) argue that Stakeholder Theory necessitates organizations consider all parties' interests in ethical decision-making. AI systems in advertising should be designed not solely to maximize shareholder profits but also to protect customers, employees, and societal welfare. Stakeholder inclusivity requires incorporating diverse group perspectives during development to prevent algorithmic harms including discrimination and data breaches. Mao (2025) address AI-supported advertising effects on children, youth, and parents through the Privacy-Ethics Alignment (PEA-

AI) framework. Stakeholder inclusivity mandates incorporating concerns of young users with limited digital literacy and their parents into AI system design. Research shows AI applications developed without parent and educator input create "trust deficits" and are perceived as unethical.

Kelm and Johann (2025) demonstrate that AI-supported storytelling for stakeholder engagement on social media directly relates to inclusivity. Algorithms highlighting stories incorporating diverse cultural codes and minority voices—beyond popular culture—increase brand social inclusion scores. AI enables analyzing stakeholder emotional responses, facilitating more meaningful, personalized connections.

Gabelaia and Smaidziunaite (2024) examine AI-supported sentiment analysis in corporate communication for understanding internal stakeholders—employees. Companies can identify resistance to AI technologies or compliance issues through these analyses, making implementation processes more participatory. Inclusivity requires positioning technology as collaborative tools incorporating employee input rather than top-down impositions.

Lee (2025) found that consumers' perception of algorithmic bias risk directly affects their willingness to engage with AI advertising systems. Demographic groups perceiving discriminatory targeting tend to avoid digital platforms. Stakeholder inclusivity mandates algorithmic cultural sensitivity testing to prevent these groups' exclusion from digital markets.

Lefebvre (2024) describe AI as a powerful tool for stakeholder mobilization in social marketing, creating societal benefits. From health campaigns to environmental awareness, AI-supported personalized messages enable diverse community segments' active participation in behavioral change processes. However, protecting vulnerable groups—such as those at addiction risk—represents an ethical inclusivity requirement.

Ibeh (2024) argues AI ethics expands corporate responsibility boundaries to encompass entire supply chains and society as stakeholders. Transparency and accountability form foundations

of societal trust in brands. Inclusive AI strategies must ensure technological economic benefits do not generate social costs including unemployment or privacy loss.

2.5. Monitoring & Risk Management (MRM)

The fifth dimension addresses adherence to evolving legal frameworks governing AI systems in marketing contexts. The Digital Services Act (DSA) establishes strict transparency rules for online platform advertising. Under Article 26, platforms must inform users about advertising content and provide real-time explanations of targeting parameters. Advertisers using AI-supported targeting algorithms must present "Why am I seeing this ad?" explanations in accessible language (European, 2022).

Helberger and Diakopoulos (2023) explain that the EU AI Act classifies algorithmic systems by risk level. AI advertising manipulating subconscious consumer behavior or targeting vulnerable groups—including children—falls under "unacceptable risk" categories, facing prohibition. Compliance requires brands to prove advertising campaigns do not employ such manipulative techniques.

Hoofnagle and van der Sloot (2024) emphasize that data privacy regulations (GDPR and KVKK) have strengthened explicit consent requirements for consumer data collection powering AI advertising. Algorithms profiling users without permission for hyper-personalized ads face severe penalties. Regulatory compliance serves not merely as legal protection but as strategic tool for building consumer-respecting brand image.

Sag (2023) addresses intellectual property law complexities: AI-generated advertising visuals and texts may lack copyright protection due to insufficient "human creativity." The U.S. Copyright Office and national courts require human intervention for registration. This necessitates new compliance protocols documenting human-AI collaboration to protect agency content ownership.

Campbell et al. (2023) discuss FTC regulations classifying AI-generated "fake consumer reviews" and "artificial testimonials" as misleading advertising. Compliance requires brands to verify whether faces, voices, or texts in advertisements reflect genuine experiences. Unsupervised use risks "unfair competition" and "consumer deception" allegations.

Sands et al. (2022) examine virtual influencer disclosure requirements. Consumers' right to know whether they interact with humans or AI is protected under unfair commercial practices regulations. Non-disclosure leads to legal sanctions and eroded brand trust.

Kietzmann et al. (2020) address deepfake technology risks concerning personality rights and copyright. Unauthorized commercial use of individuals' images or voices through AI represents a severely regulated area. Brands must follow written consent and source attribution procedures when utilizing this technology.

The Cyberspace Administration of China (2023) introduced "Interim Measures for the Management of Generative Artificial Intelligence Services," establishing global precedents. AI-generated advertising content must align with socialist core values, avoid discrimination, and respect intellectual property rights. Global brands must train and filter algorithms according to regional legal requirements.

3. Developing Ethical Governance in AI Marketing Scale

Literature review synthesizes current scholarly knowledge on ethical governance in AI marketing across five interconnected dimensions: Accountability & Responsibility (AR), Transparency & Explainability (TE), Ethical Oversight & Compliance (EOC), Stakeholder Inclusiveness & Feedback (SIF), Monitoring & Risk Management (MRM). The findings reveal several critical insights for theory development and practical implementation.

First, accountability emerges as a foundational principle requiring clear disclosure of AI involvement in advertising content. Research consistently demonstrates that while such

disclosure may temporarily increase consumer scrutiny, it ultimately builds sustainable trust and brand integrity. The responsibility extends across the AI value chain, from training data quality to outcome verification, with the principle of human oversight remaining paramount.

Second, transparency and explainability are essential for eliminating algorithmic "black boxes" and enabling informed consumer choices. Consumers require not merely disclosure of AI use but understanding of how their data informs targeting and recommendations. Explainable AI systems are perceived as fairer and more trustworthy, directly contributing to customer loyalty.

Third, ethical audit and compliance mechanisms must systematically ensure AI marketing practices align with both ethical standards and legal requirements. Independent algorithm audits, ethical firewalls, and human oversight are necessary safeguards against algorithmic bias, manipulation, and discriminatory outcomes. These mechanisms must address intellectual property concerns and verify training data legitimacy.

Fourth, stakeholder inclusivity demands consideration of all affected parties beyond shareholders—including vulnerable populations, employees, small businesses, and communities. Algorithmic bias can systematically exclude certain groups from digital markets, necessitating cultural sensitivity testing and inclusive design approaches.

Fifth, regulatory compliance increasingly requires navigating complex, evolving legal frameworks including the EU AI Act, Digital Services Act, GDPR, and emerging regulations globally. These frameworks classify AI applications by risk, mandate transparency, and impose strict requirements for data consent and algorithmic accountability.

Despite growing scholarly attention, significant gaps remain. First, limited empirical research has developed and validated comprehensive measurement instruments capturing all five governance dimensions simultaneously. Most existing studies focus on individual aspects rather than integrated frameworks. The synthesized findings provide need for strong theoretical foundations for developing the Ethical Governance in AI Marketing Scale (EG-AIM–20).

4.METHOD

4.1.Study Des□gn and Pro□edure

This study employed a quantitative, survey-based design to develop and validate the Ethical Governance in AI Marketing (EG-AIM) scale. The scale operationalizes ethical governance in AI-driven marketing through five conceptual dimensions: Accountability & Responsibility (AR), Transparency & Explainability (TE), Ethical Oversight & Compliance (EOC), Stakeholder Inclusiveness & Feedback (SIF), and Monitoring & Risk Management (MRM).

Data were collected through an online questionnaire distributed between December 2025 and January 2026. The online format was selected to efficiently reach individuals exposed to AI-generated advertising across various digital platforms. Participation was voluntary and anonymous, and all respondents provided informed consent electronically before proceeding to the survey questions. An online survey was created using a web-based platform. A link was distributed via social media. The first page explained the study's purpose, voluntary participation, anonymity, and confidentiality. Participants gave informed consent by clicking "I agree". The survey took approximately 8–10 minutes to complete; no incentives were offered. Participation was entirely voluntary. No personally identifiable information was collected. All data were analysed in aggregate form to ensure anonymity and confidentiality. The study complied with ethical principles for survey research involving human participants.

4.2.Part□□pants and Sampl□ng

Part□□pants and Sampl□ng

The target population was adults aged ≥18 years who had been exposed to AI-generated advertisements. A convenience sampling strategy was used; participants were recruited via announcements on social media platforms. Inclusion criteria were (a) age ≥18 and (b)

self-reported exposure to AI-generated ads. The sole exclusion criterion was inability to provide informed consent.

A total of 400 valid responses were obtained (no missing data). Table 1 summarises the demographic and behavioural characteristics of the sample.

Table 1. *Demographic and behavioural profile of participants (N = 400)*

Characteristic	Category	n	%
Gender	Female	190	47.5
	Male	177	44.3
	Prefer not to say	32	8.0
	Other / unspecified	1	0.3
Age group	11–17*	11	2.8
	18–25	129	32.3
	26–35	145	36.3
	36–45	65	16.3
	46–55	28	7.0
	56+	22	5.5
Edu□at□on	Bachelor's degree	188	47.0
	Associate degree	79	19.8
	Master's or higher	74	18.5
	High school or less	50	12.5
	Other / unspecified	9	2.3
Da□ly so□al med□a use	< 2 hours	68	17.0
	2 – < 4 hours	136	34.0

	4 – < 6 hours	100	25.0
	6 – < 8 hours	56	14.0
	8+ hours	36	9.0
	Unspecified / other	4	1.0
Interaction with AI-generated content	Sometimes	131	32.8
	Frequently	105	26.3
	Rarely	69	17.3
	Very frequently	48	12.0
	Never	41	10.3
	Unspecified	6	1.5
Exposure to AI-generated ads	Sometimes	152	38.0
	Frequently	106	26.5
	Rarely	68	17.0
	Very frequently	46	11.5
	Never	23	5.8
	Unspecified	5	1.3

Note: *Ages below 18 (n=11) were retained because sensitivity analyses excluding them did not alter the results.

The sample was predominantly young adult (18–40 years), well-educated (bachelor's degree or higher: 65.5%), and regularly engaged with AI-generated content and advertisements, confirming its suitability for validating the EG-AIM scale.

4.3.Measures

The EG-AIM scale was developed for this study. The initial version consisted of 20 items (four per sub-dimension). Each item used a 7-point Likert scale from 1 (strongly disagree) to 7 (strongly agree). The scale was administered in Turkish.

4.4.Data Analysis Strategy

Data were analyzed using IBM SPSS AMOS (version 28). Descriptive statistics were computed first. Then, confirmatory factor analysis (CFA) was performed to test the measurement model. Model fit was evaluated using χ^2/df (<3 acceptable, <5 marginal), CFI (≥0.90), TLI (≥0.90), RMSEA (≤0.08), and SRMR (≤0.08) (Hu & Bentler, 1999). Convergent validity was assessed with composite reliability (CR ≥0.70), average variance extracted (AVE ≥0.50), and MaxR(H). Discriminant validity was examined using the Fornell-Larcker criterion (√AVE > inter-construct correlations) and the heterotrait-monotrait ratio (HTMT <0.85 for strict validity (Henseler et al., 2015)). All tests were two-tailed with $\alpha = 0.05$.

Table 4. *Data analysis summary*

Analysis step	Method / criterion
Descriptive statistics	Frequencies, percentages
Measurement model	Confirmatory factor analysis (CFA)
Model fit criteria	χ^2/df <3–5, CFI/TLI ≥0.90, RMSEA ≤0.08, SRMR ≤0.08
Convergent validity	CR ≥0.70, AVE ≥0.50, MaxR(H)
Discriminant validity	Fornell-Larcker (√AVE > correlations), HTMT <0.85
Significance level	$\alpha = 0.05$ (two-tailed)

4.5.Validity and Reliability of the Measurement Model

The psychometric properties of the Ethical Governance in AI Marketing (EG-AIM) scale were assessed through confirmatory factor analysis (CFA).

Convergent validity refers to the extent to which items of a given construct share a high proportion of variance. It was assessed using composite reliability (CR ≥ 0.70), average variance extracted (AVE ≥ 0.50), and maximum reliability (MaxR(H)). Internal consistency reliability was evaluated via CR.

As shown in Table 2, CR values for all five dimensions ranged from 0.787 to 0.863, exceeding the recommended threshold of 0.70, indicating satisfactory reliability. AVE values were above 0.50 for four constructs (AR = 0.530, TE = 0.611, EOC = 0.530, MRM = 0.544). For Stakeholder Inclusiveness & Feedback (SIF), AVE was 0.481, slightly below the conventional cut-off. However, following Malhotra and Dash (2011), reliability can be established through CR alone when AVE is marginally low, especially because CR (0.787) remains acceptable. Moreover, MaxR(H) values ranged from 0.790 to 0.864, further supporting adequate construct reliability.

All factor loadings were significant at $p < 0.001$, and standardized loadings exceeded 0.60, providing additional evidence of convergent validity.

Table 2. ***Convergent validity and reliability indices***

Construct	CR	AVE	MSV	MaxR(H)	√AVE
Accountability & Responsibility (AR)	0.818	0.530	0.783	0.820	0.728
Transparency & Explainability (TE)	0.863	0.611	0.765	0.864	0.782
Ethical Oversight & Compliance (EOC)	0.818	0.530	0.783	0.827	0.728
Stakeholder Inclusiveness & Feedback (SIF)	0.787	0.481	0.725	0.790	0.693
Monitoring & Risk Management (MRM)	0.827	0.544	0.725	0.830	0.738

Note: CR = composite reliability; AVE = average variance extracted; MSV = maximum shared variance; MaxR(H) = maximum reliability; √AVE = square root of AVE (used for Fornell-Larcker criterion).

Bootstrap confidence intervals (95%) for CR and AVE (Table 3) showed stable estimates, with lower bounds of CR ranging from 0.696 to 0.812 and upper bounds from 0.851 to 0.902, confirming the precision of the reliability estimates.

Table 3. ***Bootstrap 95% confidence intervals for CR and AVE***

Construct	CR Lower	CR Upper	AVE Lower	AVE Upper
AR	0.750	0.867	0.429	0.621

TE	0.812	0.902	0.519	0.696
EOC	0.746	0.869	0.426	0.626
SIF	0.696	0.851	0.364	0.588
MRM	0.755	0.880	0.437	0.647

Discriminant validity ensures that each construct is distinct from the others. It was assessed using two complementary criteria: the Fornell-Larcker criterion (Fornell & Larcker, 1981) (square root of AVE greater than inter-construct correlations) and the heterotrait-monotrait (HTMT) ratio of correlations (strict threshold < 0.85). Table 4 presents the inter-construct correlations along with the square roots of AVE on the diagonal. For most constructs, the √AVE values were lower than the correlations with other constructs (e.g., AR with EOC: 0.728 < 0.885), indicating that the Fornell-Larcker criterion was not fully satisfied. However, this criterion is known to be very strict, especially when constructs are conceptually related. Therefore, the more robust HTMT criterion was employed as the primary evidence for discriminant validity (Henseler et al., 2015).

Table 4. ***Fornell-Larcker discriminant validity matrix***

Construct	**AR**	**TE**	**EOC**	**SIF**	**MRM**
AR	**0.728**				
TE	**0.875**	**0.782**			
EOC	**0.885**	**0.815**	**0.728**		
SIF	**0.771**	**0.693**	**0.822**	**0.693**	
MRM	**0.769**	**0.755**	**0.833**	**0.852**	**0.738**

Note: Diagonal (bold) = √AVE; off-diagonals = Pearson correlations. All correlations significant at $p < 0.001$.

HTMT analysis. As shown in Table 5, all HTMT values were below the conservative threshold of 0.85, ranging from 0.566 (TE ↔ SIF) to 0.743 (AR ↔ TE). The highest value (0.743) remains well within the acceptable limit. No HTMT warnings were generated, confirming that

each pair of constructs shares less than 85% of their variance, thereby supporting discriminant validity.

Table 5. ***HTMT discriminant validity ratios***

Construct pair	HTMT ratio
AR ↔ TE	0.743
AR ↔ EOC	0.734
AR ↔ SIF	0.610
AR ↔ MRM	0.631
TE ↔ EOC	0.709
TE ↔ SIF	0.566
TE ↔ MRM	0.639
EOC ↔ SIF	0.665
EOC ↔ MRM	0.691
SIF ↔ MRM	0.687

Note: Threshold for strict discriminant validity = 0.85 (Henseler et al., 2015).

4.6.Summary of Validity and Reliability

The revised five-factor EG-AIM scale demonstrated satisfactory internal consistency (CR > 0.70 for all dimensions). Convergent validity was supported by AVE values ≥ 0.50 for four dimensions and acceptable CR for the fifth (SIF). Discriminant validity was confirmed through HTMT ratios, all below 0.85. Although the Fornell-Larcker criterion was not fully met due to high inter-factor correlations, the HTMT results – considered more rigorous for closely related constructs – provide strong evidence that the five dimensions are empirically distinct. Overall, the measurement model exhibits adequate reliability, convergent validity, and discriminant validity, supporting the use of the EG-AIM scale for further structural analyses.

CONCLUSION

This study aimed to develop and validate the Ethical Governance in AI Marketing (EG-AIM) scale, a multidimensional measurement instrument designed to capture five key dimensions of ethical governance in AI-driven marketing: Accountability & Responsibility (AR), Transparency & Explainability (TE), Ethical Oversight & Compliance (EOC), Stakeholder Inclusiveness & Feedback (SIF), and Monitoring & Risk Management (MRM). The scale was tested using survey data from 400 participants and evaluated through confirmatory factor analysis, reliability, convergent validity, and discriminant validity assessments.

The findings highlight several strengths of the EG-AIM scale. First, internal consistency reliability was satisfactory across all five dimensions, with composite reliability (CR) values ranging from 0.787 to 0.863, exceeding the recommended threshold of 0.70. Second, convergent validity was largely supported: four dimensions achieved average variance extracted (AVE) values above 0.50, while the SIF dimension fell only marginally below the threshold (0.481) but retained acceptable reliability (CR = 0.787). Third, discriminant validity was robustly confirmed using the heterotrait-monotrait (HTMT) criterion, with all values below the strict threshold of 0.85 (ranging from 0.566 to 0.743), indicating that the five dimensions are empirically distinct despite their theoretical interrelatedness. Fourth, bootstrap confidence intervals for CR and AVE were stable, reinforcing the precision of reliability estimates.

Despite these strengths, the study revealed substantial limitations concerning overall model fit. The initial measurement model demonstrated poor fit across multiple indices: the chi-square to degrees of freedom ratio (CMIN/DF = 6.011) exceeded the recommended range of 1–3, indicating significant misfit; the Comparative Fit Index (CFI = 0.841) and Root Mean Square Error of Approximation (RMSEA = 0.112) fell well short of the thresholds for excellent fit (>0.95 and <0.06, respectively), while the PClose value (0.000) rejected the hypothesis of close fit. Only the Standardized Root Mean Square Residual (SRMR = 0.069) met the acceptable criterion (<0.08), suggesting partial model adequacy. In addition, the Fornell-Larcker criterion was not fully satisfied: the square roots of AVE were lower than several inter-construct correlations, particularly between AR and EOC and between EOC and SIF. These results indicate substantial correlations among dimensions, reflecting conceptual proximity and potential overlap, and highlight the need for further model refinement.

Taken together, the findings suggest that the EG-AIM scale, in its current 20-item structure, represents a promising but not yet fully optimized measurement instrument. The high inter-factor correlations may indicate the presence of a higher-order ethical governance construct or the need for further refinement of certain dimensions to enhance discriminant validity. Future research should therefore consider rewording items that contribute to cross-loadings, testing the scale across larger and more diverse samples, examining a higher-order factor model, conducting cross-cultural measurement invariance analyses, and revising the SIF dimension to improve its AVE value.

In conclusion, this study provides initial empirical support for the EG-AIM scale as a reliable and theoretically grounded instrument for measuring ethical governance in AI marketing. Although the scale demonstrates acceptable reliability and discriminant validity, model fit indices remain marginal relative to stringent benchmarks. Accordingly, the scale may be used with caution while additional validation efforts refine its structure. Overall, this study constitutes an important step toward operationalizing ethical governance in AI-driven marketing, an area of growing academic, managerial, and regulatory importance.

Referen□es

Acar, A. (2024). Yapay Zekâ Etiği Bağlamında Reklamcılık Sektörü Üzerine Uygulamalı Bir Araştırma. *Erciyes İletişim Dergisi*, *11*(1), 95–118.

Amos, C., & Zhang, L. (2024). Revealing AI Involvement in Ad Creation: Effects on Authenticity, Brand Perceptions and Consumer Intentions. *Journal of Information Systems Engineering and Management*, *10*(1), 1056–1056.

Arıcı, E. G. (2025). Gözetim ve Denetim Aracı Olarak Dijital Reklamcılık: Google, Meta ve TikTok Örnekleri. *İletişim Çalışmaları Dergisi*, *11*(3), 283–305.

Ashraf, A., Muneer, S., & Hassan, H. (2024). The Impact of AI Generated Advertising Content on Consumer Buying Behavior and Consumer Engagement. *Bulletin of Business and Economics*, *13*(2), 1152–1157.

Buder, F. (2024). *Transparency without trust? Consumer attitudes toward AI-generated marketing content*.

Campbell, C., Sands, S., & Plangger, K. (2023). Acceptable AI? The legal limits of artificial intelligence in marketing claims. *Business Horizons*, *66*(4), 1–18.

Ceviz, H. (2025). Yapay Zekâ ile Geliştirilen Reklamların Tüketici Davranışlarına Etkilerine İlişkin Literatür İncelemesi. *Tüketici Davranışlarında Güncel Konular*, *5*(1), 55–68.

Chen, Y., Wang, H., Hill, S. R., & Li, B. (2024). Consumer attitudes toward AI-generated ads: Appeal types, self-efficacy and AI's social role. *Journal of Business Research*, *185*, 114867–114867.

China, C. A. o. (2023). *Interim Measures for the Management of Generative Artificial Intelligence Services*.

Çeber, B. (2024). Reklam Ajanslarında Yapay Zekâ Kullanımı: Sektör Profesyonellerinin ChatGPT ve Midjourney Deneyimlerine Yönelik Bir Araştırma. *Erciyes İletişim Dergisi*, *11*(2), 583–606.

European, C. (2024). *Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (AI Act)*.

European, U. (2022). *Regulation (EU) 2022/2065 on a Single Market For Digital Services (Digital Services Act)*.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*, *18*(1), 39–50.

Gabelaia, I., & Smaidziunaite, G. (2024). The Effectiveness of AI-powered Sentiment Analysis in Corporate Communication in Improving Stakeholder Engagement. *ToKnowPress*, *14*, 237–248.

Gogoi, M. (2025). Ethical AI Frameworks for Responsible Internal Auditing Practices: A Conceptual and Theoretical Framework. *International Journal of Latest Technology in Engineering, Management & Applied Science*, *14*(12), 42–50.

Gölgeli, K. (2025). Yapay Zekâ ile Reklam Tasarımı: Reklamcılara Yönelik Bir Araştırma. *İnsan ve Toplum Bilimleri Araştırmaları Dergisi*, *14*(1), 319–336.

Grace, D. (2024). When AI Meets Advertising: The Potential of Misleading and Regulatory Framework on Consumer Protection. *SHS Web of Conferences*, *185*, 07007–07007.

Hacıhasanoğlu, P., & Akgün, Z. (2025). Yapay Zekâ ve Tüketici Davranışları: Etkileşimler ve İlişkiler Üzerine Bir İnceleme. *Uluslararası Yönetim İktisat ve İşletme Dergisi*, *21*(4), 1446–1464.

Helberger, N., & Diakopoulos, N. (2023). The EU AI Act: A new era for algorithmic advertising regulation. *Internet Policy Review*, *12*(1), 1–25.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the academy of marketing science*, *43*(1), 115–135.

Hoofnagle, C. J., & van der Sloot, B. (2024). Privacy in the age of advertising AI: GDPR compliance and beyond. *Computer Law & Security Review*, *52*, 105921–105921.

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1–55.

Ibeh, C. V. (2024). AI and ethics in business: A comprehensive review of responsible AI practices and corporate responsibility. *International Journal of Science and Research Archive*, *11*(1), 234–245.

Jaiswal, A. (2024). Toward Fairness, Accountability, Transparency, and Ethics in AI for Social Media and Health Care: Scoping Review. *JMIR Medical Informatics*, *12*, e50048–e50048.

Kelm, O., & Johann, M. (2025). Evaluating the impact of storytelling elements on social media stakeholder engagement: an AI-driven approach. *Data & Policy*, *6*, e12–e12.

Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). Deepfakes: Trick or treat? *Business Horizons*, *63*(2), 135–146.

Lee, J. (2025). Understanding Perceptions of Algorithmic Bias Through the Risk Perception and Attitude Framework. *International Journal of Human–Computer Interaction*.

Lefebvre, R. C. (2024). A Vision of the Future: Harnessing Artificial Intelligence for Strategic Social Marketing. *Societies*, *4*(2), 13–13.

Malhotra, N. K., & Dash, S. (2011). Marketing research an applied orientation (paperback). In: London: Pearson Publishing.

Mao, Y. (2025). Privacy–Ethics Alignment in AI: A Stakeholder-Centric Framework for Ethical AI. *Systems*, *13*(6), 455–455.

Parmar, D. S., & Saran, H. K. (2025). Empirical Study on The Role of Explainable AI (XAI) in Improving Customer Trust in AI-Powered Products. *International Journal of Computer Trends and Technology*, *73*(2), 48–57.

Qazim, T. (2025). Ethical Implications of AI in E-commerce: Balancing Innovation with Data Privacy and Fairness. *South Eastern European Journal of Public Health*, *26*(1), 1655–1665.

Rane, N. (2024). Detailing the Stakeholder Theory of Management in the AI World: A Position Paper on Ethical Decision-Making. *Advances in Multidisciplinary & Scientific Research Journal*, *10*(1), 1–8.

Sag, M. (2023). Copyright Safety for Generative AI. *Houston Law Review*, *61*(2), 525–580.

Sands, S., Campbell, C., Ferraro, C., & Mavrommatis, A. (2022). Seeing it my way: The effect of body-positive versus idealized AI-generated influencers on purchase intention. *Journal of Retailing and Consumer Services*, *67*, 103014–103014.

To, R. N., Wu, Y. C., Kianian, P., & Zhang, Z. (2025). When AI Doesn't Sell Prada: Why Using AI-Generated Advertisements Backfires for Luxury Brands. *Journal of Advertising Research*, *65*(2), 202–236.

Turan, M. (2024). Yapay Zekâ Etiği: Temel İlkeler, Sorunlar ve Disiplinlerarası Yaklaşımlar. *Denetişim Dergisi*(Özel Sayı), 18–32.

Yıldırım, E., & Can, A. (2024). Yapay Zekanın Reklamcılık Alanında Kullanımı ve Yapay Zeka Kullanılarak Yazılmış Reklam Film Senaryo Analizi. *Pamukkale Üniversitesi İletişim Bilimleri Dergisi*, *3*(2), 268–287.

CONCLUSION

AI-Generated Advertising: Key Insights, Persistent Questions, and a Path Forward

The integration of generative artificial intelligence into advertising is no longer an emerging trend – it is the operational reality for brands, agencies, and platforms worldwide. From automated copywriting and synthetic visuals to algorithmically optimized targeting and virtual influencers, AI systems now mediate a substantial and growing share of brand-consumer interactions. Yet as this volume has demonstrated, the shift from human-crafted to algorithmically authored advertising is not merely a technical or efficiency-driven evolution. It is a psychological, relational, and ethical transformation that fundamentally alters how consumers evaluate the persuasive messages they encounter, the brands those messages

promote, the manufacturers that deploy AI technologies, and the corporate entities ultimately responsible for algorithmic outcomes.

This book brought together five empirical studies examining consumer responses to AI-generated advertising across multiple levels of analysis – from specific brand attitudes to general corporate reputation, from cognitive trust assessments to affective emotional bonds, from perceived performance to ethical governance expectations. While each chapter tested distinct theoretical models or developed novel measurement instruments, several cross-cutting themes have emerged.

Summary of Key F□nd□ngs A□ross Chapters

1. Attitudes toward AI-generated advertising cascade across organizational levels

Chapter 1 established that favorable brand attitudes formed in response to AI-generated advertisements positively influence attitudes toward the manufacturer, which in turn shape corporate reputation. This hierarchical cascade – brand → manufacturer → corporation – demonstrates that consumers do not evaluate AI ads in isolation. Rather, specific advertising encounters serve as diagnostic signals about the organizational entities behind them. The strong path from manufacturer attitude to corporate reputation ($\beta = 0.871$) suggests that how consumers perceive a manufacturer's handling of AI technology is a powerful determinant of broader corporate judgments.

Chapter 2 examined a related but distinct cascade: manufacturer attitude → brand attitude → affective brand trust. The findings supported full mediation, with manufacturer attitude strongly predicting brand attitude ($\beta = 0.701$), which in turn strongly predicted affective brand trust ($\beta = 0.667$). Together, Chapters 1 and 2 suggest a reciprocal relationship between brand and manufacturer evaluations. Positive manufacturer perceptions enhance brand attitudes, which generate emotional trust; conversely, positive brand attitudes enhance manufacturer perceptions, which build corporate reputation. This reciprocity implies that managing AI advertising effectively requires coordinated attention to both brand-level and corporate-level communication strategies.

2. AI explanations affect multiple dimensions of user response, but scales require refinement

Chapter 3 developed the Artificial Intelligence Explanation Effects Scale (AIDES) to capture four dimensions of how consumers respond to AI explanations: Perceived Transparency and

Understanding (PTU), Perceived Credibility and Honesty (PCH), Ethical Fairness and Accountability (EFA), and Trust Calibration and Behavioral Response (TCB). The systematic literature review summarizing 28 studies (Table 1) is a valuable contribution in its own right. However, the confirmatory factor analysis revealed poor model fit (CFI = 0.848, RMSEA = 0.125), and two dimensions – Perceived Credibility and Honesty (AVE = 0.485) and Trust Calibration and Behavioral Response (AVE = 0.484) – fell below the conventional threshold for convergent validity. These results suggest that while the theoretical distinction among the four dimensions is defensible, consumers' empirical responses to AI explanations may not align neatly with these categories. The high correlation between PTU and PCH, for example, implies that consumers who feel they understand an AI system also tend to perceive it as credible and honest – a finding that echoes the transparency-trust link documented elsewhere in the literature.

3. Trust calibration is measurable, but measurement remains challenging

Chapter 4 introduced the concept of trust calibration – the alignment between consumers' trust in AI-generated ads and the perceived performance of those ads – to the advertising domain. The proposed Trust Calibration toward AI Advertising (TCAIA) scale and Calibration Index represent a promising operationalization of a construct that has been extensively studied in human-automation interaction but rarely applied to consumer advertising. The mean Calibration Index of 0.906 suggests that Turkish consumers, on average, calibrate their trust reasonably well. However, the poor confirmatory factor fit (CFI = 0.77, RMSEA = 0.130) and high inter-factor correlations (e.g., cognitive trust alignment with perceived performance at 0.992) indicate that the proposed four-dimensional structure does not hold empirically. This finding carries an important implication: consumers may not distinguish sharply between "trust in AI ads" and "perceived performance of AI ads." For many individuals, trusting an AI ad may be nearly synonymous with perceiving it as competent and accurate.

4. Ethical governance in AI marketing is multidimensional but dimensions may overlap substantially

Chapter 5 presented the Ethical Governance in AI Marketing (EG-AIM) scale, operationalizing five dimensions: Accountability & Responsibility (AR), Transparency & Explainability (TE), Ethical Oversight & Compliance (EOC), Stakeholder Inclusiveness & Feedback (SIF), and Monitoring & Risk Management (MRM). The scale demonstrated satisfactory internal consistency (CR > 0.78 for all dimensions) and HTMT-based discriminant validity (all values < 0.85). However, the Fornell-Larcker criterion was not satisfied, with several inter-factor

correlations exceeding the square roots of average variance extracted (e.g., AR with EOC at 0.885, EOC with SIF at 0.822). Model fit was poor by conventional standards (CFI = 0.841, RMSEA = 0.112). These findings suggest that while the five dimensions are conceptually meaningful, they may not be empirically separable in consumers' minds. It is possible that a higher-order ethical governance construct underlies all five dimensions, or that some dimensions (e.g., Accountability & Responsibility and Ethical Oversight & Compliance) are sufficiently closely related that they could be merged.

FINAL REMARKS

The rise of AI-generated advertising presents a paradox. On one hand, AI offers unprecedented capabilities for personalization, creativity, and efficiency – benefits that consumers may genuinely appreciate when implemented well. On the other hand, the algorithmic authorship of persuasive content raises fundamental questions about authenticity, manipulation, accountability, and the nature of brand-consumer relationships. Consumers want relevance and convenience, but they also want honesty, transparency, and respect for their autonomy. They want brands to use AI effectively, but they do not want to feel deceived or manipulated by machine-generated content.

This book has attempted to navigate that paradox empirically. The findings suggest several tentative conclusions. First, consumers do not treat AI-generated advertising as a monolith; their responses vary by brand, manufacturer, and corporate context. Second, transparency about AI involvement carries short-term costs but may build sustainable trust. Third, perceived performance – whether the ad is accurate, relevant, and useful – is a dominant driver of consumer judgments. Fourth, emotional bonds with brands (affective trust) depend indirectly on manufacturer perceptions, mediated by brand-specific attitudes. Fifth, consumers can, on average, calibrate their trust appropriately, but significant individual variation exists. Sixth, ethical governance is a multidimensional concern for consumers, but those dimensions may not be empirically distinct – many consumers may rely on a holistic sense of whether a brand or manufacturer is "responsible" rather than on analytically separated criteria.

The scales developed in this volume – TCAIA, AIDES, and EG-AIM – represent initial steps toward measuring trust calibration, explanation effects, and ethical governance. Their current psychometric limitations should not be seen as failures but as invitations to refine, to test, and

to improve. Scale development is inherently iterative; the versions presented here are offered to the scholarly community for critique, replication, and collaborative enhancement.

As a final note, this book was conducted entirely within the Turkish market. Turkey's young, digitally active population provides a valuable context for studying AI advertising, but the findings should not be exported uncritically to other cultural, regulatory, or economic environments. Replication in diverse settings is essential.

The age of AI-generated advertising has arrived. Whether it becomes an era of manipulation and distrust or an era of transparency and calibrated reliance depends not only on the technology itself but on the governance frameworks, ethical standards, and consumer protections that researchers, practitioners, and regulators build around it. This book offers empirical foundations for that work – incomplete, provisional, and imperfect, but we hope, useful. We invite others to build on it.