

Algorithmic Illusions: The Impact of AI-Driven Misleading Digital Marketing on Consumer Loyalty Dynamics

Kavita Rani, Research Scholar, Department of Commerce, NIILM University, Kaithal, Haryana
Dr. Anish Ahmad, Associate Professor, Department of Commerce, NIILM University, Kaithal, Haryana

Abstract

Background: AI-driven digital marketing enables advanced personalization but also fuels deceptive practices like dark patterns, fake social proof, emotional targeting, and opaque pricing. While short-term gains are evident, long-term impacts on trust and loyalty remain unclear.

Method: A mixed-method study was conducted with 420 urban Indian consumers (18–55) and content analysis of 60 AI-based campaigns across e-commerce, food delivery, travel, and finance. PLS-SEM was used to assess effects on trust, purchase intention, switching, and loyalty.

Results: 73.8% reported exposure to deceptive AI marketing. Key tactics included false urgency (67.2%), price discrimination (54.1%), and fake reviews (48.6%). Deception significantly reduced brand trust ($\beta = -0.41$) and loyalty ($\beta = -0.38$). Trust fully mediated loyalty in high-involvement products. Younger consumers showed greater trust loss but quicker recovery with transparency.

Conclusion: AI-driven deception undermines brand trust and loyalty. Ethical AI frameworks and transparent marketing practices are essential for sustainable consumer relationships.

Keywords: Algorithmic deception, AI marketing, dark patterns, consumer loyalty, brand trust, digital ethics.

1. INTRODUCTION

We live in an algorithmic age. Artificial intelligence determines what consumers see, in what sequence, at what price, and with what emotional context whether they swipe through a shopping app, watch a video, or read a product review. The hallmark of current internet marketing is algorithmic reality curating. At its best, it finds products you want, reminds you of things you almost forgot, and matches you with offers that meet your needs. It can produce algorithmic illusions, which this article describes meticulously built false experiences optimized to exploit user cognitive vulnerabilities for commercial metrics. The scope and intricacy of these illusions are unmatched. AI systems can now generate fake but convincing product reviews at industrial scale, dynamic countdown timers that show urgency that doesn't exist, fabricated 'X people are viewing this right now' notifications, hyper-personalised pricing that charges different users different amounts for the same product without disclosure, and emotionally manipulative advertising micro-targeted to psychologically vulnerable individuals. These are recorded behaviors with regulatory ramifications. The U.S. Federal Trade Commission penalized Fortnite creator Epic Games USD 245 million in March 2023 for utilizing dark patterns to deceive customers into buying. Amazon was sued by the FTC in 2024 for purposefully hindering subscription cancellation. The Indian Central Consumer Protection Authority (CCPA) issued 13 dark pattern prohibitions in 2022. Despite rising regulatory scrutiny, research literature on how AI-driven deceptive marketing harms consumer loyalty is few. Most investigations focus on technical dark pattern identification (Mathur et al., 2019; Gray et al., 2023) or consumer vulnerability in digital contexts (OECD, 2023). Few studies have studied how AI's ability to personalize, scale, and optimize deceptive behaviors affects consumers' trust calculus, which determines their brand loyalty.

That gap is filled by this paper. It presents survey-based findings from 420 digital consumers in urban India, one of the world's fastest-growing digital markets, on their exposure to AI-driven misleading marketing practices, their emotional and cognitive responses, and their brand trust and loyalty. By identifying and measuring unique misleading AI marketing strategies and

their differential damage to consumer loyalty, the article contributes to algorithmic marketing ethics and responsible AI deployment.

2. LITERATURE REVIEW

AI in Digital Marketing: Capabilities and Ethical Risks

AI has changed customer segmentation, content production, real-time bidding, sentiment analysis, and predicted churn modeling in digital marketing. AI-powered customisation can create five to eight times the marketing spend ROI compared to traditional methods, according to McKinsey (2023). NLP allows chatbots to mimic human empathy. Recommender systems cut product discovery cognitive load. Programmatic advertising targets consumers at certain points throughout their trip. The same qualities that enable helpful personalization also enable its misuse. In their analysis of responsible AI in marketing, Ozturkcan and Bozdog (2024) identify two key phenomena that undermine consumer trust: 'AI booring' (when AI systems visibly fail to deliver on their promised intelligence, eroding confidence) and 'AI washing' (when companies falsely claim AI capabilities to attract investment and consumer interest, AI marketing systems may target or exploit susceptible demographic groups due to algorithmic bias (Akter et al., 2022). AI systems trained on prior purchase data may perpetuate credit, insurance, and pricing discrimination without human reviewers knowing.

Dark Patterns: Definition, Taxonomy, and AI Amplification

Dark patterns, coined by UX designer Harry Brignull in 2010, are user interface and experience designs that intentionally push users toward decisions they would not make with full information and free choice. Since Brignull's classification, the taxonomy has grown. A complete taxonomy by Gray et al. (2023) identified nagging, obstruction, stealth, interface interference, forced action, and social engineering as dark pattern families. The most common dark patterns in digital marketing are false urgency (fake countdown timers and stock alerts), confirmshaming (emotionally manipulative opt-out language), hidden costs (prices revealed only at checkout), roach motels (easy to sign up, impossible to cancel), and misdirection. AI has transformed dark patterns in three ways. The same deceptive message can be dynamically adjusted to each consumer's psychology, surfing history, and known emotional state at the time of exposure using AI. Second, AI allows scale—deceptive patterns that were previously manual can now be implemented across millions of user sessions. Third, AI allows optimisation—algorithmic A/B testing optimises misleading aspects to enhance their effectiveness (Mondaq, 2024). AI-driven dark patterns are more harmful than humanly produced ones due to these three traits.

Consumer Trust, Brand Loyalty, and the Role of Perceived Deception

Brand loyalty is based on consumer trust, which is the assumption that a brand will act honestly, dependably, and in their best interests. Trust and loyalty are especially high in digital environments because consumers cannot physically verify things and must rely on vendors and algorithmic intermediaries. AI-driven deceitful marketing infringes on all three dimensions of Mayer, Davis, and Schoorman's (1995) trust model: ability, benevolence, and integrity. The AI can deceive at scale, the system isn't functioning in the consumer's interest, and the communication is dishonest. Perceived deception—the consumer's subjective assessment that a company deliberately misled them—consistently triggers negative emotional responses (anger, betrayal, helplessness) that reduce purchase intention, word-of-mouth recommendation, and long-term brand loyalty. Deceptive interfaces are more likely to cause customers to abandon the service and avoid future brand encounters, even if they were algorithmically created, according to the Nielsen Norman Group (2023). This is crucial: people believe AI-generated deceit equally to human-designed deception.

Regulatory Landscape — Global and Indian Context

The regulatory response to AI-driven fraudulent marketing has escalated since 2022. Dark patterns by large internet platforms are prohibited by the EU's Digital Services Act (DSA,

2023) The EU AI Act (2024) mandates a risk-based framework for AI systems and addresses consumer deceptive AI uses. Under Section 5 of the FTC Act, AI-enabled dark patterns are misleading or unfair commercial practices, and Amazon, Epic Games, and Intuit have been fined over USD 600 million. The CCPA (2023) published national guidelines specifying 13 prohibited dark pattern categories under the Consumer Protection Act 2019, making India one of the first nations to establish comprehensive dark pattern regulation. These legislative changes show that AI-driven misleading marketing is now a legal liability with serious financial ramifications.

3. CONCEPTUAL FRAMEWORK AND HYPOTHESES

This study combines three theoretical viewpoints. First, the Stimulus-Organism-Response (S-O-R) framework (Mehrabian and Russell, 1974) says that external stimuli (S) cause internal organismic states (O), including cognitive and emotional assessments, which lead to behavioral reactions (R). In this study, AI-driven deceptive marketing techniques are the stimulus, perceived deception, emotional reactance, and trust evaluation are the organism states, and loyalty intention, switching behavior, and word-of-mouth are the responses. The second perspective is the Technology Acceptance Model (TAM) with a trust dimension (Davis, 1989; Gefen and Straub, 2003), which recognises that customer acceptance of AI-mediated marketing depends on trustworthiness. The third is Social Contract Theory (Donaldson and Dunfee, 1994), which views the brand–consumer relationship as an implicit social contract. When a brand uses AI to deceive consumers, it breaks this contract, causing loyalty erosion.

The following hypotheses are tested using this integrated approach and the literature:

H1: Exposure to AI-driven misleading marketing practices significantly and negatively affects perceived brand trustworthiness.

H2: Perceived brand trustworthiness fully mediates the relationship between AI-driven deception exposure and consumer loyalty intention.

H3: The type of AI deceptive practice (false urgency, price discrimination, fake reviews, emotional manipulation, hidden costs, fake social proof) significantly moderates the magnitude of brand trust erosion.

H4: Age moderates the deception–trust erosion relationship, with younger consumers (18–30) experiencing stronger immediate trust damage but demonstrating faster loyalty recovery following transparent brand communication.

H5: Product involvement moderates the deception–loyalty relationship, with higher-involvement products showing stronger deception-induced loyalty erosion.

4. RESEARCH METHODOLOGY

Research Design: In this work, we use a mixed-methods approach to explain AI-driven marketing campaigns in the actual world. We interview people quantitatively and then analyze their content qualitatively. The quantitative part of the research presents data to back up the hypothesized connections between AI dishonesty and the results in customer loyalty. The results are put into context by the qualitative strand, which details particular cases of deceptive marketing in the Indian digital market that was made possible by artificial intelligence.

Quantitative Survey: In four major Indian cities—Delhi NCR (n=130), Mumbai (n=120), Bengaluru (n=100), and Hyderabad (n=70)—420 validated respondents, ranging in age from 18 to 55, who are active users of at least two digital platforms (e-commerce, food delivery, travel, personal finance, or social media) were given the survey instrument. Stratified purposive sampling was used to recruit respondents from a professional survey panel. Quotas were set up to ensure fair representation of gender (52% male, 48% female) and age groups (18–30: 38%; 31–45: 44%; 46–55: 18%). The survey was divided into four parts: (A) information about digital consumers and how they use platforms; (B) a 24-item measure of exposure to six types of AI-driven misleading practices on a 5-point frequency scale (ADMES); (C) a 16-item

measure of perceived brand trustworthiness ($\alpha = 0.87$), emotional response ($\alpha = 0.82$), and loyalty intention ($\alpha = 0.89$), derived from validated instruments; and (D) data about demographics and socioeconomics. All constructions have $CR > 0.80$, indicating internal consistency, as validated by Cronbach's Alpha and composite reliability. We used 5-point Likert scales for all of the items. We gathered this data from November 2023 to February 2024.

Qualitative Content Analysis: Twenty digital marketing initiatives involving artificial intelligence were gathered and coded from the e-commerce, food delivery, trip booking, and personal finance/insurance industries. Publicly available regulatory case data, consumer complaint reports submitted with the CCPA, and screenshot archives were used to compile the campaign materials. Two separate coders used a seven-category deceitful practice codebook that was based on the taxonomy of the CCPA (2023) and the dark patterns ontology developed by Gray et al. (2023). With Cohen's $\kappa = 0.82$, the inter-rater reliability was deemed acceptable. Using Braun and Clarke's (2006) six-phase framework, we conducted thematic analysis on consumer qualitative responses (open-ended survey items).

Structural Equation Modelling: We used SmartPLS 4.0 and Partial Least Squares-Structural Equation Modelling (PLS-SEM) to test our hypothesised model. Because it works well with complicated models and non-normal distributions, PLS-SEM was chosen. The HTMT criterion was used to confirm discriminant validity, and model fit was evaluated with the standardised root mean square residual (SRMR < 0.08 threshold). All values were found to be less than 0.85. To test for mediation effects and path coefficients, we utilized bootstrap resampling with a sample size of 5,000.

5. RESULTS AND FINDINGS

Sample Profile and Digital Platform Engagement

With an average age of 33.2 (SD=8.6) years, 420 respondents made up the final validated sample. Among the population, 61.4% had completed some college and 28.3% had earned a master's degree or higher. Digital spending per month varied between 2,000 and 25,000+, with 5,001 to 10,000 falling in the middle. Virtually every respondent utilized some kind of digital platform on a regular basis, with online shopping (92.1%) and meal delivery (84.3%) coming out on top. All of the demographic information and platform usage summaries are in Table 1.

Table 1: Respondent Demographic and Platform Usage Profile (N = 420)

Characteristic	Category	Frequency	Percentage (%)
Gender	Male	218	51.9%
Gender	Female	202	48.1%
Age Band	18–30 years	160	38.1%
Age Band	31–45 years	184	43.8%
Age Band	46–55 years	76	18.1%
Education	Graduate	258	61.4%
Education	Postgraduate	119	28.3%
City	Delhi NCR	130	31.0%
City	Mumbai	120	28.6%
City	Bengaluru / Hyderabad	170	40.4%
Most Used Platform	E-Commerce (Amazon/Flipkart)	387	92.1%

Characteristic	Category	Frequency	Percentage (%)
Most Used Platform	Food Delivery (Swiggy/Zomato)	354	84.3%
Most Used Platform	Travel Booking (MakeMyTrip/Ixigo)	276	65.7%

Source: Primary survey data, November 2023 – February 2024

Exposure to AI-Driven Misleading Marketing Practices

The sheer frequency of exposure to deceptive marketing campaigns driven by AI is the first major finding of this research. Incredibly, 73.8% of those who took the survey said they had encountered some sort of deceptive marketing campaign powered by artificial intelligence in the past three months. You can see how often each of the six types of misleading practices appears in Table 2 and Figure 1.

Table 2: Exposure to AI-Driven Misleading Marketing Practices (N = 420)

AI Deceptive Practice Category	Description	Respondents Exposed	% Exposed	Mean Frequency (1–5)
False Urgency (AI-Generated)	Fake countdown timers, fabricated 'Only 2 left' alerts	282	67.2%	3.81
Personalised Price Discrimination	Different prices for same product without disclosure	227	54.1%	3.24
AI-Generated Fake Reviews	Obviously inauthentic or AI-written reviews used to influence	204	48.6%	2.97
Emotional Manipulation Targeting	Ads served at vulnerable moments using emotional AI	189	45.0%	2.83
Hidden Cost Revelation	Charges revealed only at final checkout stage	176	41.9%	2.64
AI-Fabricated Social Proof	'1,247 people bought this today' fake real-time statistics	164	39.1%	2.41
Exposed to ≥ 1 Practice	Any of the above	310	73.8%	—

Source: Primary survey data; ADMES scale responses; N=420

Figure 1: Exposure to AI-Driven Misleading Marketing Practices (% of 420 Respondents)

False Urgency (fake timers, stock alerts)		67.2%
AI Personalised Price Discrimination		54.1%
AI-Generated Fake Reviews		48.6%




Emotional Manipulation Targeting		45%
Hidden Costs at Checkout		41.9%
AI-Fabricated Social Proof		39.1%

Figure 1: Percentage of respondents reporting exposure to each AI deceptive practice category
Figure 1: Exposure to AI-Driven Misleading Marketing Practices (% of 420 Respondents)

The vast deployment of AI-optimized scarcity messaging across Indian e-commerce and travel platforms is shown in the largest number of respondents who faced false urgency tactics (67.2%). In the qualitative research, it was found that 38 out of 60 campaigns had aspects of artificial urgency. These included countdown timers that reset with each page load and dynamic stock counters that reset on page refresh. 'I noticed the timer reset when I refreshed the page – I felt legitimately tricked and immediately lost interest in buying from that app,' recounted a 29-year-old female respondent from Mumbai, one of several who claimed recognizing these patterns.

One of the most complex and hard-to-detect types of algorithmic deceit is AI-personalised price discrimination without transparency, which 54.1% of people have experienced. Consumers aren't aware of price discrimination, in contrast to false urgency, because they only see their own price and have no idea if it varies from others. The qualitative examination of CCPA complaint records uncovered various instances of food delivery and trip booking platforms imposing higher costs on users determined by AI to be more price sensitive according to device kind, browsing history, and location data.

Emotional and Cognitive Response to AI Deception

A trend of negative affect following detected AI deceit was seen in the emotional reaction data. Out of all the emotions experienced, the majority of those exposed (61.4%) reported anger as 'strong' or 'very strong.' Next on the list were disappointment (57.8%), betrayal (49.3%), and helplessness (38.1%). These results are in line with the research on betrayal aversion (Koehler and Gintis, 2007), which demonstrates that breaches of trust in partnerships where there is a perceived authority figure (such as in the case of a brand-consumer relationship) elicit stronger emotional reactions than similar harms from impersonal sources. Psychologically, rather of simply being more skeptical about the brand in question, 68.7 percent of those who were exposed to the ads said they were more skeptical of digital marketing communications in general. Particularly noteworthy is the fact that this generalization of mistrust implies that the entire digital marketing ecosystem suffers as a result of individual brands' use of AI to engage in misleading marketing. The widespread mistrust of brands that don't act ethically has a negative impact on brands that do.

SEM Results: Impact on Brand Trust and Loyalty

The important path coefficients tested by PLS-SEM for each of the five hypotheses are shown in Table 3. Strong explanatory power was indicated by the model, which explained 58% of the variance in brand loyalty intention ($R^2 = 0.58$).

Table 3: PLS-SEM Path Coefficients — Hypothesised Relationships

Path / Hypothesis	β Coefficient	t-Value	p-Value	Supported?
H1: AI Deception → Brand Trust	-0.41	8.84	< 0.001	YES ✓
H2: Brand Trust → Loyalty Intention	-0.52	11.42	< 0.001	YES ✓

Path / Hypothesis	β Coefficient	t-Value	p-Value	Supported?
H2 (mediation): AI Deception \rightarrow Loyalty (via Trust)	-0.38	7.91	< 0.001	YES (Full Mediation)
H3: Practice Type \times Trust Erosion (Interaction)	Sig. variation	4.12–9.26	< 0.001	YES \checkmark
H4: Age \times Deception–Trust (Young: stronger erosion)	0.22 (mod.)	3.47	< 0.001	YES \checkmark
H5: Product Involvement \times Deception–Loyalty	0.31 (mod.)	5.08	< 0.001	YES \checkmark
Model R ² (Loyalty Intention)	0.58	—	—	Strong Fit

Source: PLS-SEM analysis; SmartPLS 4.0; Bootstrap n=5,000; N=420

The important path coefficients tested by PLS-SEM for each of the five hypotheses are shown in Table 3. Strong explanatory power was indicated by the model, which explained 58% of the variance in brand loyalty intention ($R^2 = 0.58$).

Loyalty Consequences: Switching and Word-of-Mouth

The behavioral effects of being exposed to AI deceit are shown in Table 4. There is a decline in customer loyalty across all three characteristics of behaviour that have been measured.

Table 4: Behavioural Loyalty Consequences of AI Deception Exposure

Loyalty Outcome Variable	Non-Exposed Group (n=110)	Exposed Group (n=310)	Mean Difference	Statistical Significance
Repurchase Intention (1–5 scale)	4.18	2.89	-1.29	t=12.4, p<0.001
Platform Switching Intention (1–5)	1.84	3.47	+1.63	t=14.8, p<0.001
Positive WOM Intention (1–5)	3.92	2.41	-1.51	t=13.1, p<0.001
Negative WOM / Online Complaint (1–5)	1.41	3.12	+1.71	t=15.3, p<0.001
Overall Loyalty Index (composite)	3.87	2.51	-1.36	t=13.9, p<0.001

Source: Primary survey data; independent samples t-test; N=420

On a 5-point scale, there is a statistically significant and commercially substantial difference of 1.29 points in repurchase intention between exposed consumers (2.89) and non-exposed consumers (4.18). Customer lifetime value is significantly threatened when repurchase intention drops by 1.29 percentage points across a consumer base. Consumers who have been exposed to AI deception have a significantly higher switching intention score (3.47 vs. 1.84 among unexposed consumers), which means that the average exposed consumer is now more likely to be considering switching than "unlikely to switch." Consumers who feel deceived are motivated to leave and warn others, which amplifies the reputational damage beyond the directly affected customer base. This is supported by the well-documented pattern in the

betrayal literature, which explains the dramatic increase in negative word-of-mouth intention (3.12 vs. 1.41) among exposed consumers.

6. DISCUSSION

The Loyalty Paradox of AI Marketing

This research proves that AI-powered digital marketing is based on a basic paradox. Marketers claim that their goal is to increase customer loyalty through fostering preference-commitment relationships that result in increased sales, word-of-mouth advertising, and reluctance to transfer brands. On the other hand, marketers are using AI's skills in ways that undermine loyalty's foundational trust. No reasonable long-term plan for managing a brand should be prepared to tolerate the economically substantial damage rate shown by the $\beta = -0.41$ path from exposure to AI deceit to erosion of brand confidence. Loyalty can't be restored without trust, and trust can't endure systematic deceit, as shown by the discovery of full mediation through brand trust (H2). This seeming contradiction is the result of a fundamental chasm between the optimization goals of AI marketing platforms and the real concerns of companies. Click-through rates, conversion rates, average order values, and subscriber counts are usually the short-term metrics that AI systems are optimized for. These short-term measures are consistently improved by deceitful techniques, such as phony urgency, hidden costs, and fraudulent reviews. However, they achieve this by spending trust capital that was built over years and bringing forward demand from future periods. The artificial intelligence is focusing on improving the measuring system rather than the core business goal. The algorithm subtly undermines the efficacy of marketing while creating the impression that it is working. This is the crux of the algorithmic illusion.

The Amplification Effect: Why AI Deception Is Different

The qualitative research found that customers react more strongly to deceit they believe is enabled by AI than they do to similar deceit perpetrated by human marketers. In describing their encounter with AI deceit, some respondents used the term "calculated"—"it felt like a machine was studying me and exploiting what it found." Beyond the effects of a typical misleading commercial, the sensation of being followed and targeted by an artificial system heightens the feeling of violation. According to psychologists, the 'personal betrayal effect' is heightened when one feels targeted rather than caught in a generic net; this is because AI deceit is personalized, meaning it is designed to exploit specific vulnerabilities. The regulatory and management response is greatly affected by this amplification impact. It implies that disclosure rules related to AI, which notify consumers when an AI system is tailoring their experience, are more than just a transparency nice-to-have; they are a real way to reduce harm. Legally sound and likely to mitigate the psychological damage caused by AI-enabled deceit are the provisions of the EU AI Act concerning transparency in AI-human contact and the FTC's recommendations about disclosure of AI-generated content.

Towards Ethical AI Marketing — A Framework

Choosing between unethical AI marketing and honest AI marketing boils down to prioritizing short-term KPIs above long-term company success, according to this study's findings, regardless of the time frame in question. It takes a lot of time and money to restore trust in a brand after it has been severely damaged. The comprehensive mediation result verifies that restoring trust is an essential first step on the road to loyalty rehabilitation. Furthermore, the age moderation finding implies that the consumer's perception of the brand grows more stable and old, making trust recovery a more challenging task. Getting the deceit right, or more accurately, getting the ethics right from the beginning, is significantly cheaper than trying to fix it.

7. RECOMMENDATIONS AND ETHICAL FRAMEWORK

Table 5: Ethical AI Marketing Framework — Principles and Practices

Principle	What It Means in Practice	Business Case
Algorithmic Transparency	Disclose when AI is personalising prices, content, or urgency signals. Use plain language, not buried legal disclaimers.	Reduces perceived deception; builds trust capital; reduces regulatory risk
Genuine Urgency Only	Deploy countdown timers and stock alerts only when backed by real inventory or deadline data. No AI-generated simulated scarcity.	Short-term conversion cost outweighed by long-term trust preservation
Honest Social Proof	Use only verified, authentic customer reviews. Disclose when reviews are AI-summarised. Remove clearly inauthentic patterns.	Fake reviews are increasingly detectable; regulatory penalties rising (CCPA, FTC)
Transparent Price Personalisation	If dynamic pricing is used, disclose the practice clearly. Provide easy mechanism for consumers to understand their price basis.	Undisclosed discrimination is a trust landmine; disclosed dynamic pricing is accepted by most consumers
Ethical Emotional Targeting	Do not serve emotionally manipulative content at identified vulnerable moments (e.g., financial stress periods). Set AI targeting guardrails.	Emotional exploitation produces strongest negative WOM; violates EU AI Act Article 5
Consumer-Controlled Personalisation	Allow consumers to review and adjust their profile data used for AI recommendations and pricing.	Control perception is a strong trust predictor (ISACA, 2024); reduces perceived manipulation
Human Override in High-Stakes Decisions	Ensure human review of AI decisions with significant consumer financial consequences (credit, insurance, high-value pricing).	Required by EU AI Act for 'high-risk' AI systems; reduces systemic harm risk
Regular Dark Pattern Audits	Conduct periodic independent audits of digital interfaces for dark patterns using AI detection tools (e.g., Fair Patterns).	CCPA mandates self-audits; early detection prevents regulatory action and loyalty damage

Source: Researcher's framework based on study findings, EU AI Act (2024), CCPA Guidelines (2023), FTC Guidance, and published literature

8. LIMITATIONS

A number of caveats should be had in mind regarding this investigation. There is a big limitation to the generalizability of the results to customers in rural areas or with lower levels of digital engagement in the Indian setting because the sample is urban and highly involved in digital activities. Even if it's starting from a different place, digital adoption in rural areas is growing rapidly, so this is particularly true. In addition, to better understand the dynamics of trust loss and recovery over time, a longitudinal study tracking the same consumers for 12-18 months would be preferable than a cross-sectional method that only captures a snapshot of deceit exposure and trust at one instant in time. Third, respondents may overstate or underestimate their level of deceit when asked to recollect instances of deceit or to apply the

social desirability effect on questions regarding self-reported exposure to deceit. The 60 campaigns that were part of the qualitative content study did not all involve AI-driven marketing in India. Recognizing the conclusions as indicative rather than exhaustive is crucial, notwithstanding the analysis's thoroughness and ratings. Finally, it's worth noting that the study doesn't specifically address the trust implications of using recommendation engines, NLP-based chatbots, and dynamic pricing algorithms, thus aggregate analysis might overlook some of them.

9. CONCLUSION

Does AI-driven deceptive marketing harm consumer loyalty? This article started with a simple but essential question. With evidence from sixty campaign analyses and four hundred and twenty-two customers, the answer is crystal clear: it does, and the harm is large from a statistical and economic perspective, and it is caused only by the decline in consumer confidence in the brand. With an overall exposure rate of 73.8% in just three months, it's clear that deceptive marketing facilitated by AI is not an isolated incident; rather, it is a common occurrence for digital customers in India. For brands that depend on AI without ethical constraints, the $\beta = -0.41$ trust degradation route and the 1.29-point repurchase intention gap are not theoretical concepts, but rather actual competitive losses. An essential aspect of AI in marketing right now is the idea of algorithmic illusions, which is the use of AI to generate false feelings of urgency, authenticity, social consensus, and pricing justice. These deceptions are not made by chance. They are the result of optimization efforts that failed to account for the need of building trust over the long run in favor of immediate conversions. The issue of fixing AI is more of a problem with values and governance than with technology. As intended, the algorithms carry out their tasks. What we want for them to do is the key question.

This study's main takeaway is that using artificial intelligence to fuel deceitful marketing is more of a trust liquidation approach than a growth strategy. For the sake of quick conversion stats that can't make up for the loyalty losses they cause, brands who use AI to deceive are cashing out on a trust account that took years to create. An ethical and commercially sound use of AI in marketing would be to use algorithms to provide customers with experiences that are truly relevant to them, honest, and empowering. Instead of being a roadblock to performance, ethical AI marketing lays the groundwork for long-term success. The expanding capabilities of AI are raising the stakes of this issue. Generative AI can now create synthetic influencer endorsements, hyper-realistic false reviews, and personalized deepfake ads. Although regulatory systems are making progress, they are not enough on their own. An industry-wide dedication to the idea that algorithms should benefit consumers rather than hurt them is necessary. Safer results—for customers, for companies, and for the entire digital ecosystem—are the result of clean code, just as they are of clean hands.

REFERENCES

1. Akter, S., McCarthy, G., Sajib, S., Michael, K., Dwivedi, Y.K., D'Ambra, J., and Shen, K.N. (2021). Algorithmic bias in data-driven innovation in the age of AI. *International Journal of Information Management*, 60, 102387.
2. Akter, S., Dwivedi, Y.K., Sajib, S., Biswas, K., Bandara, R.J., and Michael, K. (2022). Algorithmic bias in machine learning-based marketing models. *Journal of Business Research*, 144, 201–216.
3. Brignull, H. (2023). *Deceptive patterns*. Book. self-published; website: deceptive.design
4. Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
5. Central Consumer Protection Authority of India (CCPA) (2023). Guidelines for prevention and regulation of dark patterns. Ministry of Consumer Affairs, Food and Public Distribution, Government of India.

6. Davis, F.D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319–340.
7. Donaldson, T., and Dunfee, T.W. (1994). Toward a unified conception of business ethics: Integrative social contracts theory. *Academy of Management Review*, 19(2), 252–284.
8. Erasmus, V., Daha, T.J., Brug, H., Richardus, J.H., Behrendt, M.D., Vos, M.C., and van Beeck, E.F. (2010). Systematic review of compliance with hand hygiene guidelines. *Infection Control and Hospital Epidemiology*, 31(3), 283–294.
9. European Data Protection Board (EDPB) (2023). Guidelines 03/2022 on dark patterns in social media platform interfaces. Version 2.0. EDPB, Brussels.
10. European Parliament and Council of the EU (2023). Regulation (EU) 2022/2065 — Digital Services Act. *Official Journal of the European Union*.
11. European Parliament (2024). Artificial Intelligence Act (EU AI Act). Regulation (EU) 2024/1689. *Official Journal of the European Union*.
12. Federal Trade Commission (FTC) (2022). Bringing dark patterns to light: FTC staff report. U.S. Federal Trade Commission, Washington DC.
13. Federal Trade Commission (FTC) (2023). In the matter of Epic Games, Inc. (Fortnite dark patterns enforcement). FTC File No. C-4771. USD 245 million settlement.
14. Fortes, N., Rita, P., and Pagani, M. (2023). The effects of privacy concern and perceived control on the willingness to share personal data: A cross-national study. *Information and Management*, 60(3), 103762.
15. Gefen, D., and Straub, D.W. (2003). Managing user trust in B2C e-services. *e-Service Journal*, 2(2), 7–24.
16. Getty Images / VisualGPS (2024). Building trust in the age of AI. Report based on 30,000+ adults across 25 countries (2022–2024). Getty Images, New York.
17. Gray, C.M., Santos, C., and Bielova, N. (2023). Towards a preliminary ontology of dark patterns knowledge. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM.
18. Gray, C.M., Chamorro, L.S., Obi, I., and Duane, J.N. (2023). A systematic literature review mapping dark patterns scholarship. *Journal of Usability Studies*, 18(3), 111–139.
19. ISACA (2024). AI ethics: Navigating different cultural contexts. *ISACA Now Blog*, December 6.
20. Kaushal, V., and Mishra, R. (2024). Explainable AI in marketing: Enhancing transparency and consumer trust in algorithmic recommendations. *Journal of Business Research*, 178, 114641.
21. Koehler, J.J., and Gintis, H. (2007). Emotions, morality and game theory. *Games and Economic Behavior*, 61(1), 216–227.
22. Lebni, J.Y., Nazari, B., Khosravi, B., Ziapour, A., and Chaboksavar, F. (2023). The impact of social media marketing on customer loyalty: The mediating role of brand trust and brand equity. *Frontiers in Public Health*, 11, 1197697.
23. Luguri, J., and Strahilevitz, L.J. (2024). Shining a light on dark patterns. *Journal of Legal Analysis*, 13(1), 43–109.
24. Mathur, A., Acar, G., Friedman, M.J., Lucherini, E., Mayer, J., Chetty, M., and Narayanan, A. (2019). Dark patterns at scale: Findings from a crawl of 11K shopping websites. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1–32.
25. Mayer, R.C., Davis, J.H., and Schoorman, F.D. (1995). An integrative model of organizational trust. *Academy of Management Review*, 20(3), 709–734.
26. McKinsey and Company (2023). The state of AI in 2023: Generative AI's breakout year. McKinsey Global Institute, New York.
27. Mehrabian, A., and Russell, J.A. (1974). *An approach to environmental psychology*. MIT Press, Cambridge, MA.

28. Morgan, R.M., and Hunt, S.D. (1994). The commitment-trust theory of relationship marketing. *Journal of Marketing*, 58(3), 20–38.
29. Nielsen Norman Group (2023). Users' trust in digital interfaces and dark patterns. NNG Research Report.
30. OECD (2023). Consumer vulnerability in the digital age. *OECD Digital Economy Papers*, No. 347. OECD Publishing, Paris.
31. Pew Research Center (2023). AI in everyday life: Public awareness and views on artificial intelligence. Pew Research Center, Washington DC.
32. Reichheld, F.F. (2003). The one number you need to grow. *Harvard Business Review*, 81(12), 46–54.
33. Sari, F.R., Astuti, N.R., and Putri, T.H. (2024). The impact of AI-driven recommendations on consumer trust and purchase behavior: A study in Indonesian online retail. *Journal of Business and Retail Management Research*, 18(1), 14–25.

