

Review

Ethics of Digital Marketing in the AI Era: A Structured Thematic Review of Recent Research, 2023–2025

Alexios Kaponis *  and Manolis Maragoudakis 

Department of Informatics, Ionian University, Plateia Tsirigoti 7, 49100 Corfu, Greece; mmarag@ionio.gr

* Correspondence: alexios.kaponis@ionio.gr

Abstract

Artificial intelligence has become increasingly embedded in digital marketing, reshaping how firms collect consumer data, personalize content, automate persuasion, and evaluate campaign performance. While these developments offer clear strategic benefits, they also raise important ethical questions concerning privacy, transparency, fairness, manipulation, and accountability within platform-mediated marketing environments. This article presents a structured thematic review of recent research on the ethics of AI-driven digital marketing. The review focuses on English-language peer-reviewed journal articles and high-quality conference proceedings published between January 2023 and September 2025. Searches were conducted across Scopus, ScienceDirect, SpringerLink, ACM Digital Library, IEEE Xplore, MDPI, PubMed, and selected academic repositories. After deduplication, screening, and full-text assessment, 91 studies were included in the final synthesis. Methodological quality was appraised using the Mixed Methods Appraisal Tool and Joanna Briggs Institute critical appraisal criteria, while the findings were examined through thematic synthesis. The review identifies five recurring ethical domains in the literature: data privacy and GDPR compliance, algorithmic transparency and explainable AI, algorithmic fairness in targeting and automated decision-making, dark patterns and deceptive interface design, and influencer or virtual influencer disclosure. The evidence suggests that privacy, consent, deceptive design, and transparency are the most extensively discussed areas, whereas the practical effectiveness of fairness interventions, explainability tools, and disclosure mechanisms remains more mixed and context-dependent. Across these themes, ethical risks appear not only as the result of individual corporate decisions but also as outcomes shaped by platform infrastructures, ranking systems, data access arrangements, and performance-oriented advertising metrics. The article contributes by organizing recent scholarship into a coherent thematic framework and by situating AI-driven digital marketing ethics within a broader platform-governance perspective. It argues that responsible implementation requires clearer distribution of accountability among businesses, platforms, regulators, and researchers. The review is limited by its focus on the 2023–2025 period, its English-language scope, and the methodological heterogeneity of the included studies, which prevents statistical meta-analysis.



Academic Editor: Teresa Paiva

Received: 31 January 2026

Revised: 6 June 2026

Accepted: 15 June 2026

Published: 6 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: digital marketing ethics; artificial intelligence; data privacy; GDPR; algorithmic transparency; explainable AI; algorithmic fairness; dark patterns; influencer disclosure; platform governance

1. Introduction

Artificial intelligence (AI) has become deeply embedded in the infrastructures that sustain contemporary digital markets. Over the past decade—and particularly in the period between 2023 and 2025—AI-driven systems have transformed digital marketing from a largely campaign-based practice into a continuously adaptive, data-intensive ecosystem. Personalized recommendation engines, automated bidding systems, predictive customer analytics, conversational agents, and generative AI content tools now operate at scale, enabling firms to target consumers with unprecedented precision [1–3]. These technological developments promise efficiency, relevance, and measurable performance gains. Yet they also raise profound ethical and governance questions that extend far beyond traditional marketing concerns.

Much of the early discussion around AI in marketing focused on optimization: how machine learning could increase engagement, improve segmentation, and enhance return on investment [1,2]. However, as these technologies matured, scholarly attention shifted toward the ethical dimensions of algorithmic personalization. Researchers began to question how consumer data are collected and processed, how opaque systems shape decision-making, and whether automated targeting mechanisms may reproduce structural biases or manipulate vulnerable groups [4]. Increasingly, ethical concerns have moved from peripheral debates to central considerations in digital marketing scholarship.

Simultaneously, it has become clear that AI-driven digital marketing does not operate in a vacuum. It unfolds within platform-based ecosystems—social media networks, search engines, digital marketplaces, app stores, and content-sharing platforms—that serve as both technological infrastructures and governance actors. Platforms determine how content is ranked, how advertisements are distributed, how data flows between actors, and how compliance rules are enforced. They design the interfaces through which users provide consent, encounter disclosures, and engage with promotional content. In doing so, platforms mediate not only economic exchange but also ethical outcomes.

This infrastructural and governance dimension is critical. Ethical challenges in digital marketing—such as privacy protection, transparency, fairness, deceptive interface design, and influencer disclosure—are not solely the product of individual firms' choices. They are co-produced within socio-technical systems shaped by platform architecture and algorithmic curation. For example, recommendation systems embedded in social media feeds determine which advertisements become visible and which remain unseen. Interface defaults influence whether users meaningfully consent to data processing or simply click “accept” under cognitive pressure. Platform policies regulate influencer disclosures and shape the visibility of compliant or non-compliant content. In this sense, platforms function simultaneously as marketplaces, regulators, and rule-setters within digital marketing ecosystems.

The adoption of the General Data Protection Regulation (GDPR) marked a turning point in formalizing accountability for data practices in the European Union [5]. By establishing principles such as data minimization, lawful processing, transparency, and the right to erasure, the GDPR sought to rebalance power between consumers and organizations. Yet empirical research suggests that compliance often exists in tension with commercial incentives [6]. Consent mechanisms, although legally mandated, frequently manifest as complex or visually manipulative cookie interfaces that encourage acceptance rather than informed choice [7,8]. In practice, therefore, the effectiveness of data protection depends not only on regulation but also on how platforms implement interface design and enforcement mechanisms.

Parallel concerns arise in the domain of algorithmic transparency. Many AI systems used in digital marketing rely on complex models that function as “black boxes,” rendering

their internal logic inaccessible to users [9,10]. Explainable Artificial Intelligence (XAI) has emerged as a proposed solution, aiming to provide meaningful explanations for algorithmic outputs [11,12]. Within marketing contexts, this may involve clarifying why a particular advertisement is shown or identifying the data features that informed a recommendation. Evidence suggests that such explanations can enhance perceived transparency and trust [13]. However, transparency introduces trade-offs. Excessive disclosure may overwhelm users, reveal proprietary information, or reduce the persuasive efficiency of campaigns [14–17]. The tension between explainability and performance illustrates the broader governance dilemma confronting platform-mediated marketing systems.

Algorithmic fairness presents another dimension of concern. Machine learning models are trained on historical data, which may reflect entrenched social inequalities. As a result, targeting algorithms may unintentionally reproduce or amplify discrimination [4,18]. Empirical studies have documented disparities in advertising exposure related to employment opportunities, financial services, and housing [19,20]. Efforts to mitigate bias include the adoption of fairness metrics, demographic parity adjustments, and bias correction techniques [21–23]. Yet implementing fairness mechanisms may conflict with profit-maximization strategies, as hyper-targeted advertising often depends on differential segmentation [24]. Moreover, conceptions of fairness vary across cultural and regulatory contexts [25], complicating efforts to establish universal standards.

Beyond data and algorithms, interface design practices introduce additional ethical challenges. The proliferation of so-called “dark patterns”—design strategies that nudge users toward choices they might not otherwise make—illustrates how marketing persuasion can cross into manipulation [26]. Examples include pre-ticked consent boxes, hidden cancellation options, artificial scarcity cues, and drip pricing tactics [27–35]. While such practices may boost short-term conversion rates, research indicates they undermine long-term trust and provoke consumer reactance [36,37]. Regulatory responses, including provisions within the Digital Services Act and enforcement efforts by the Federal Trade Commission, attempt to curb deceptive practices, yet enforcement remains uneven across jurisdictions [38–40].

The rise of influencer and virtual influencer marketing further complicates the ethical landscape. Influencers often cultivate parasocial relationships with audiences, making their endorsements appear authentic rather than commercial [41]. Disclosure practices, such as hashtags or explicit sponsorship labels, aim to clarify promotional intent. However, their effectiveness varies depending on placement, clarity, and audience interpretation [42–44]. Research suggests that while transparent disclosures may enhance trust, they can also reduce perceived authenticity and persuasive impact in certain contexts [45,46]. The emergence of AI-generated virtual influencers intensifies these concerns, as audiences may struggle to distinguish between human and algorithmically generated personas [47–50]. These developments blur the boundaries between advertising, entertainment, and automated persuasion.

Taken together, these domains—data protection, algorithmic transparency, fairness, deceptive design, and influencer disclosure—constitute interconnected ethical pillars of AI-driven digital marketing. Yet their interdependence is often underexamined. Efforts to increase transparency may affect privacy; fairness interventions may alter targeting efficiency; and stricter disclosure rules may reshape influencer–platform relations. Ethical issues are therefore systemic rather than isolated. They arise from the interaction between commercial incentives, technical design choices, regulatory architectures, and user vulnerabilities.

This systemic character underscores the importance of adopting a platform governance perspective. Platforms structure digital marketing through algorithmic ranking,

data access controls, API restrictions, and moderation policies. They enforce disclosure guidelines, manage advertising inventories, and determine which content is amplified or suppressed. In doing so, platforms shape both opportunities and constraints for marketers and consumers.

The present review seeks to synthesize recent research (January 2023–September 2025) on the ethics of AI-driven digital marketing, while explicitly situating these developments within platform-mediated ecosystems. Drawing on 91 selected studies, the review identifies five thematic pillars that dominate current scholarship: (1) data privacy and GDPR compliance; (2) algorithmic transparency and explainable AI; (3) algorithmic fairness; (4) dark patterns and deceptive design; and (5) influencer and virtual influencer disclosures. Rather than treating these themes as discrete categories, the analysis conceptualizes them as overlapping governance challenges shaped by platform infrastructures.

First, a methodological limitation should be explicitly stated. The evidence base collated for this review is confined to peer-reviewed journal articles and peer-reviewed conference proceedings. Grey literature, such as industry white papers, consultancy and market research reports, practitioner blog posts, pre-prints, dissertations and non-peer-reviewed policy briefings, was intentionally excluded from the formal review corpus. This limitation was due to two causes. First, it grounds the synthesis in evidence that has been independently subject to scholarly quality control. This is especially important in a field where claims about algorithmic transparency, fairness, and manipulation are often made by parties with a direct commercial interest in the practices being examined. Secondly, it suggests that the same formal appraisal tool, presented in Section 4, can be systematically applied throughout the corpus as grey-literature outputs seldom reveal the design and reporting components of such an appraisal. A few legal, regulatory and official institutional documents, most notably the GDPR and the Digital Services Act, are referred to throughout, but only as normative and contextual background, not as primary evidence. These documents do not belong to the 91 studies that have been subjected to the thematic synthesis. In Section 9 we explore implications of this barrier on the scope of findings.

The study also introduces a stakeholder–challenge matrix that maps ethical concerns across four principal actors: consumers, businesses, regulators, and researchers. This matrix serves not merely as a descriptive tool but as a governance mapping framework, illustrating how ethical tensions are distributed and negotiated within platform ecosystems. By highlighting overlaps and asymmetries in responsibility, the framework contributes to ongoing debates about accountability in algorithmically mediated markets.

Ultimately, the central argument advanced in this review is that AI-driven digital marketing ethics must be understood as a question of platform governance. Ethical risks do not stem solely from individual marketing strategies; they are embedded in infrastructural architectures and monetization models that privilege engagement, scale, and data extraction. Regulatory frameworks such as the GDPR provide important guardrails, yet enforcement challenges, cross-border fragmentation, and rapid technological change complicate their implementation. Meanwhile, generative AI tools amplify both creative possibilities and manipulation risks, intensifying the urgency of coordinated governance responses.

By synthesizing recent empirical findings and situating them within a platform governance perspective, this paper aims to advance scholarly understanding of how ethical principles can be operationalized in AI-driven marketing systems. In doing so, it seeks to bridge digital marketing research with platform studies and governance scholarship, contributing to a more integrated and critical conversation about responsibility in the evolving digital marketplace.

This framework presents the five interconnected themes most frequently identified across the 91 studies included in the review (2023–2025), reflecting their majority.

- (1) Data privacy and GDPR compliance (purple) is concerned with consent mechanisms and first-party data strategies.
- (2) Algorithmic transparency and XAI (orange) examines how explainability operates within personalization systems.
- (3) Algorithmic fairness (red) addresses ways of mitigating bias in targeting or pricing.
- (4) Dark patterns/deceptive design (violet) analyzes deceptive interface practices.
- (5) Influencer and virtual influencer disclosures (teal) investigates transparency in promotional content.

All pillars converge on AI-driven digital marketing (center), unified by three cross-cutting challenges: manipulation and consumer autonomy, transparency–trust trade-offs, and accountability gaps in enforcement. The framework engages four key stakeholder groups: consumers, businesses, regulators, and researchers. As illustrated in Figure 1, our conceptual framework positions AI-driven digital marketing at the nexus of five interconnected ethical domains, each presenting distinct yet overlapping challenges for stakeholders across the digital marketing ecosystem.

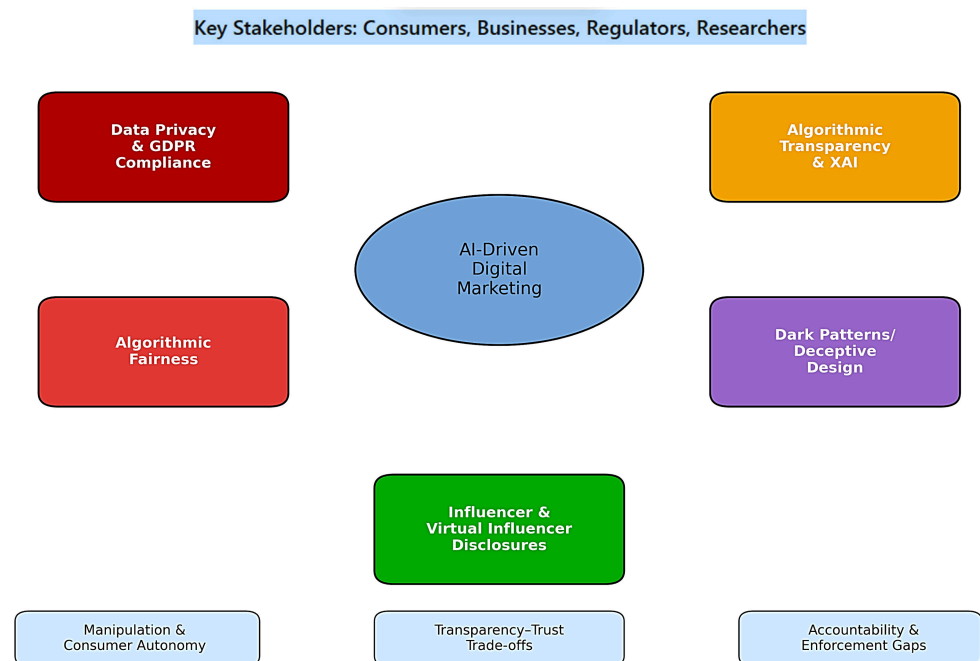


Figure 1. Conceptual framework: five thematic pillars of digital marketing ethics in the AI era.

The convergence of these themes on a central core reflects the systemic nature of ethical challenges in AI-driven marketing, where interventions in one domain inevitably affect others. For instance, efforts to increase algorithmic transparency may simultaneously impact privacy protections and raise questions about fairness in targeting practices.

Platform-Mediated Digital Marketing as Infrastructure Governance

Contemporary digital marketing can no longer be conceptualized as an autonomous managerial function operating independently of its technological environment. Increasingly, marketing activity unfolds within platform-mediated ecosystems in which social media networks, search engines, and digital marketplaces operate not merely as communication channels but as infrastructural governance systems. This shift has profound implications for how ethical questions must be framed. Unlike earlier forms of digital

marketing—such as brand-controlled websites, email campaigns, or internally managed customer relationship management systems—platform-mediated marketing is structurally shaped by algorithmic curation, recommender systems, ranking logics, and monetization architectures embedded within platform infrastructures.

This transformation reflects broader structural changes in the digital economy. Scholars of platform governance argue that platforms function simultaneously as intermediaries, market organizers, and regulatory actors [51]. They define the conditions under which visibility, engagement, and monetization occur. In AI-driven marketing environments, algorithms do not simply optimize advertiser goals; they determine which commercial messages become visible, which remain marginal, and which audiences are prioritized.

Platforms exercise governance power through multiple interlocking mechanisms. Ranking algorithms establish hierarchies of visibility by prioritizing engagement probability and behavioral relevance. Recommender systems structure exposure pathways, shaping the informational environment in which consumers encounter products and services [9,10]. Application programming interfaces (APIs) regulate access to data, redistributing informational asymmetries between firms and users. Interface architectures influence consent processes, disclosure formats, and user comprehension—particularly in areas related to GDPR compliance [5,7]. Enforcement policies moderate advertising standards, influencer disclosures, and content compliance [41,52].

These mechanisms are not neutral technical features. They embody economic priorities and governance logics. As Kitchin [53] and Pasquale [54] have argued, algorithmic systems embed normative assumptions about relevance, efficiency, and value. Within marketing ecosystems, these assumptions frequently privilege engagement optimization over transparency or fairness. For instance, targeted advertisements on social media feeds are not distributed exclusively according to advertiser configuration. They are filtered through platform ranking systems that optimize for predicted interaction rates, monetization efficiency, and user retention. Ethical concerns such as fairness in targeting [4], opacity in recommendation [9], or manipulative interface design [26] are therefore structurally mediated by platform optimization objectives.

Similarly, influencer disclosures are not simply matters of brand compliance. Their visibility and effectiveness depend on platform moderation systems, automated detection tools, and interface placement rules [43,55]. Even where regulatory frameworks require transparency—such as under EU consumer protection directives [52] or FTC guidelines [40]—enforcement remains partly delegated to platform governance infrastructures. Inconsistent application or algorithmic filtering may weaken disclosure salience despite formal compliance requirements.

Data governance further illustrates infrastructural power. While GDPR establishes clear principles of lawful processing and data minimization [5], practical implementation is often mediated by platform-designed consent interfaces. Research on cookie banners demonstrates that interface prominence and visual framing significantly influence acceptance rates [7,34]. Thus, compliance outcomes are shaped not only by legal standards but by interface architecture decisions embedded within platform ecosystems. In parallel, behavioral data extraction operates through systemic architectures that normalize tracking as a condition of digital participation. Marketing personalization systems operate within this broader data economy [56].

Regulatory asymmetry in the digital domain varies across platform types and regulatory contexts. Under the EU's GDPR, domestic advertisers face far stricter restrictions on user behavioral targeting data than their peers in low-regulation markets, a gap that undermines their ability to deploy personalized advertising configurations and generates incentives for regulatory arbitrage among multinational corporations [57]. In search adver-

tising ecosystems, auction-based visibility logic concentrates algorithmic feedback among large-budget advertisers, limiting competitive access for smaller market participants [10]. In influencer marketing, platform-specific algorithmic amplification on short-form video platforms demonstrably reduces the noticeability of sponsorship disclosures compared to static post formats, even when both environments nominally enforce identical labeling requirements [43,58]. These sector- and platform-specific divergences underline that structural asymmetries are not uniform across the digital ecosystem but are shaped by the specific governance regimes, monetization architectures, and interface logics of each platform environment. Effective governance therefore requires differentiated regulatory and design responses rather than uniform cross-sector rules.

By explicitly differentiating platform-mediated marketing from earlier forms of digital marketing, this study situates AI-driven marketing ethics within the broader field of platform governance. Ethical dilemmas—whether concerning transparency, fairness, manipulation, or disclosure—are not peripheral compliance issues but infrastructural governance outcomes. Ranking systems influence which narratives gain prominence. Data gatekeeping determines who can access behavioral insights. API asymmetries shape competitive dynamics. Enforcement delegation redistributes accountability between private platforms and public regulators.

Consequently, responsibility in AI-driven marketing cannot be confined to advertisers alone. It must be understood as distributed across platform operators, regulatory bodies, and technological design frameworks. Ethical design decisions made at the level of algorithmic architecture, interface configuration, and moderation policy shape market behavior at scale. This perspective reframes digital marketing ethics as a systemic governance challenge rather than a series of isolated managerial choices.

In doing so, the analysis aligns with emerging scholarship that conceptualizes platforms as sites of infrastructural power and normative ordering [51]. AI-driven marketing systems operate within these infrastructures, inheriting both their efficiencies and their asymmetries. Ethical outcomes are therefore co-constructed through the interaction of algorithmic design, economic incentives, and regulatory oversight.

Understanding digital marketing through this platform governance lens clarifies why issues such as algorithmic fairness [4], transparency [9], deceptive design [26], and influencer disclosure [41] persist despite formal ethical guidelines. These challenges are not merely failures of individual compliance; they are structural features of performance-driven ecosystems. Addressing them requires attention not only to corporate responsibility but to the governance architecture of platforms themselves.

2. Theoretical Framework

2.1. *Digital Marketing in the Algorithmic Era*

Digital marketing has undergone a profound transformation in recent years as artificial intelligence and machine learning technologies have become embedded in core commercial infrastructures. What was once primarily a data-supported promotional activity has evolved into a deeply algorithmic system in which automated models continuously shape visibility, pricing, targeting, and personalization decisions. Recent systematic analyses confirm that AI-driven marketing strategies now operate through predictive modeling, real-time behavioral tracking, and adaptive optimization pipelines that dynamically adjust content delivery [1,10].

AI-powered personalization represents the defining feature of this algorithmic era. Machine learning systems analyze granular behavioral signals—search histories, dwell time, transaction records, device identifiers, and geolocation metadata—to construct probabilistic profiles of individual users. These profiles determine which advertisements are displayed,

which products are recommended, and even how prices are dynamically adjusted. Such systems significantly enhance targeting precision and campaign efficiency [2], yet they simultaneously introduce new layers of opacity and asymmetry.

This transformation is closely linked to the broader logic of data-driven capitalism, where user behavior becomes a core economic resource. Digital platforms systematically extract, aggregate, and monetize behavioral data, converting consumer interactions into predictive assets. Within this environment, marketing no longer merely responds to expressed preferences; rather, algorithmic systems anticipate and subtly shape those preferences through feedback loops embedded in recommender systems and automated advertising infrastructures. As AI ecosystems become increasingly multilayered and interconnected, value creation occurs through continuous data circulation across supply chains and platform infrastructures [16].

A critical analytical distinction must therefore be made between traditional digital marketing infrastructures and platform-mediated marketing ecosystems. In earlier digital environments, firms retained substantial control over communication channels, including proprietary websites, email databases, and internally governed customer relationship management systems. By contrast, contemporary marketing increasingly unfolds within platform-centered infrastructures in which visibility is algorithmically curated rather than directly controlled by advertisers. As scholars of platform governance have argued, digital platforms function not merely as intermediaries but as infrastructural organizers that shape the conditions of participation, exchange, and visibility [51]. In this environment, brands no longer determine exposure independently; instead, their messages are filtered, ranked, and optimized within proprietary recommender systems whose logic remains only partially accessible to external actors [9,10].

This shift fundamentally alters the ethical landscape. Whereas traditional digital marketing primarily raised concerns related to persuasive messaging and data protection, platform-mediated marketing introduces infrastructural governance challenges. Platforms act as gatekeepers of data access through API design and data-sharing restrictions, thereby redistributing informational asymmetries between businesses and consumers [59]. They impose eligibility rules for advertising, structure monetization through ranking and auction systems, and enforce disclosure standards unevenly across jurisdictions [52,57]. Algorithmic systems embed normative and economic priorities within technical architectures. Consequently, ethical marketing outcomes are not shaped solely by corporate intention but by the structural configuration of platform-based algorithmic ecosystems [53,54].

Within such ecosystems, ranking algorithms determine commercial visibility hierarchies; recommender systems influence informational exposure pathways [9]; and interface architectures shape user consent experiences, often affecting GDPR compliance in practice [5,7]. These infrastructural features operate as governance mechanisms that distribute power asymmetrically across stakeholders. Consumers encounter algorithmically curated environments without full insight into prioritization logics, businesses depend on platform analytics and advertising infrastructures for reach, and regulators confront enforcement challenges within privately governed, transnational systems [39]. Ethical risks such as bias amplification [18], opacity [11], and manipulative design [26] must therefore be understood as structurally mediated phenomena rather than isolated deviations.

By differentiating platform-mediated marketing from earlier firm-controlled digital communication models, this study situates AI-driven marketing ethics within the broader scholarship on platform governance. Ethical dilemmas—whether concerning transparency, fairness, manipulation, or disclosure—are embedded within infrastructural design choices, economic optimization logics, and governance arrangements that extend beyond individual

advertisers. Responsibility, accordingly, must be conceptualized as distributed across platform operators, regulatory institutions, and technological design frameworks.

The result is a restructuring of market power. Platforms are not passive intermediaries but active governance actors that define ranking criteria, regulate data access, and shape advertising standards. Marketing campaigns operate within these infrastructural constraints, often relying on performance dashboards whose inner mechanics remain opaque even to professionals. In this context, responsibility becomes diffused across advertisers, data brokers, and platform operators. Ethical analysis must therefore extend beyond firm-level practices to encompass structural and infrastructural dimensions of algorithmic ecosystems.

Although AI-enhanced personalization improves customer relevance and engagement metrics [51], research also demonstrates that intensified personalization may trigger psychological reactance and privacy concerns when users perceive excessive intrusion. This tension illustrates a structural paradox: the same systems that generate commercial efficiency may undermine consumer trust if perceived as invasive or manipulative. Consent fatigue further complicates this dynamic. Empirical studies suggest that users frequently accept complex cookie or tracking policies without meaningful comprehension, thereby weakening the normative foundation upon which personalization systems rely [7,8].

Moreover, human–AI collaboration in marketing decision-making introduces additional governance challenges. As automated systems increasingly guide strategic choices, questions arise regarding oversight, interpretability, and responsibility allocation. Evaluations of human–AI interaction frameworks emphasize that effective collaboration requires transparency regarding model assumptions and limitations [60]. Without such clarity, marketing professionals may rely on outputs that they cannot fully interrogate, increasing the risk of unintended bias or unfair outcomes.

Thus, digital marketing in the algorithmic era must be understood as a socio-technical assemblage: a network of data extraction mechanisms, predictive algorithms, commercial incentives, and governance architectures. Ethical evaluation cannot be restricted to isolated campaigns; it must account for systemic structures embedded within platform ecosystems.

2.2. AI Ethics Principles

The growing influence of algorithmic systems in marketing has intensified calls for principled frameworks capable of guiding responsible development and deployment. Across the academic literature and policy initiatives, several recurring ethical pillars have emerged: fairness, accountability, transparency, explainability, and human oversight. While articulated in different terminologies, these principles form the normative backbone of responsible AI discourse.

Ethical AI discourse emphasizes that technological innovation must be aligned with fundamental human values rather than treated as a neutral instrument of efficiency. The European Commission’s “ethics-by-design” approach underscores the need to embed ethical reflection directly into system architecture rather than retrofitting compliance mechanisms after deployment [61]. Similarly, privacy-by-design principles advocate proactive integration of data protection safeguards within technological infrastructures [62]. These approaches shift ethics from reactive compliance toward anticipatory governance.

Fairness constitutes one of the most debated principles in AI ethics. Machine learning models are trained on historical datasets that may encode structural inequalities. When deployed in marketing contexts—such as targeted advertising or dynamic pricing—these systems can replicate or amplify disparities [4,18]. Empirical evidence demonstrates that algorithmic targeting has, in certain contexts, limited exposure to high-value opportunities for specific demographic groups [19,20,63].

However, fairness is inherently multidimensional. Different fairness metrics—such as demographic parity, equal opportunity, or individual fairness—reflect distinct normative assumptions and may conflict with one another [21,22]. Bias mitigation techniques, including reweighting training data or imposing fairness constraints, have been developed within the technical literature [23]. Yet their implementation in commercial marketing environments remains uneven, often constrained by performance optimization pressures and competitive dynamics [24]. This reveals a persistent tension between equity and efficiency.

Accountability complements fairness by clarifying responsibility across algorithmic value chains. In complex digital ecosystems involving advertisers, platforms, and data brokers, tracing decision-making processes becomes challenging. Algorithmic accountability frameworks call for documentation of data sources, model parameters, and evaluation procedures to enable auditability [14]. Without traceability, ethical principles risk remaining abstract commitments rather than enforceable practices.

Transparency and explainability are frequently treated as overlapping concepts but address distinct concerns. Transparency refers to the disclosure of relevant information about system functioning, whereas explainability concerns the interpretability of specific outputs for affected users. In marketing contexts, explainable AI (XAI) aims to clarify why a particular recommendation or advertisement was presented [11,12]. Evidence suggests that meaningful explanations can enhance perceived transparency and user confidence [13].

However, transparency presents a complex dilemma. While increased disclosure may strengthen trust, it can also reduce perceived authenticity or create cognitive overload. Recent research indicates that explicit AI disclosure can, under certain conditions, erode trust by highlighting algorithmic mediation rather than human agency [64]. Moreover, increasing interpretability sometimes requires simplifying model complexity, potentially affecting predictive performance [17]. This trade-off between transparency and efficiency remains unresolved and context-dependent.

Human oversight functions as a safeguard mechanism. The integration of human-in-the-loop systems ensures that automated outputs can be reviewed, contested, or corrected when necessary. Evaluations of human–AI collaboration emphasize that oversight mechanisms must be clearly defined to prevent responsibility diffusion [60]. In marketing applications, oversight may involve fairness audits, adversarial testing of targeting pipelines, and continuous monitoring of personalization outcomes.

Collectively, these principles form an interconnected normative architecture. Fairness without transparency limits detectability; transparency without accountability lacks enforcement; and oversight without explainability reduces effectiveness. Responsible AI in digital marketing therefore requires coordinated integration of all five dimensions.

Yet despite the proliferation of ethical guidelines, implementation gaps persist. Commercial pressures frequently prioritize optimization metrics—click-through rates, conversion rates, and engagement duration—over normative considerations. Regulatory frameworks such as the GDPR establish formal obligations for data protection [5], yet enforcement remains uneven across jurisdictions. The emerging AI regulatory landscape seeks to address these gaps, but practical translation into everyday marketing workflows remains an ongoing challenge.

In this context, AI ethics cannot be treated as an external regulatory constraint alone. It must be understood as a strategic and structural dimension of algorithmic market design. The following section extends this normative foundation by integrating AI ethics principles with transformative marketing perspectives centered on consumer well-being and societal impact.

2.3. Transformative Marketing and Consumer Well-Being

While AI ethics provides a normative foundation grounded in principles such as fairness and transparency, marketing scholarship offers a complementary perspective through transformative approaches that explicitly center consumer and societal well-being. The rise of algorithmic systems in marketing intensifies the relevance of these perspectives, as commercial communication increasingly shapes not only purchasing behavior but also broader patterns of attention, information access, and psychological experience.

Transformative Consumer Research (TCR) and Transformative Service Research (TSR) emerged as scholarly responses to the recognition that marketing activities exert profound influence on quality of life. Rather than conceptualizing consumers merely as rational utility maximizers, transformative approaches view them as socially embedded individuals whose well-being may be enhanced or undermined by market systems. Brennan and Binney argued that marketing institutions bear ethical responsibility for the societal consequences of their strategies, particularly when vulnerable groups are involved. This perspective challenges narrow profit-maximization logics by framing marketing as a social actor with moral obligations.

In parallel, Varadarajan advanced the notion that marketing strategy must evolve toward a more sustainable and stakeholder-oriented paradigm. In this view, value creation cannot be restricted to short-term financial metrics; it must incorporate long-term societal impacts, including trust, equity, and sustainability. The integration of AI into marketing infrastructures intensifies this challenge. Algorithmic personalization, predictive targeting, and behavioral nudging can enhance consumer experience when deployed responsibly, yet they can also undermine autonomy when optimized solely for engagement metrics.

Recent scholarship in responsible technology highlights this tension. Hari et al. emphasize that AI-driven marketing systems operate within a spectrum ranging from empowerment to manipulation [65]. When personalization aligns with genuine user preferences and is transparently communicated, it may improve relevance and reduce informational overload. However, when predictive models are used to exploit cognitive biases or emotional vulnerabilities, the same technological capabilities can produce ethically problematic outcomes [66].

The transformative marketing lens therefore introduces a critical evaluative question: does AI-driven personalization contribute to consumer well-being, or does it primarily optimize commercial extraction? This question becomes especially salient in the context of behavioral targeting systems that rely on psychological profiling. Choudhary characterizes AI-powered persuasion as potentially evolving into “algorithmic persuasion,” where systems learn not only what consumers prefer but also how they can be influenced most effectively [66]. Such capabilities blur the boundary between personalization and manipulation.

Well-being in digital environments encompasses multiple dimensions. It includes psychological autonomy (the ability to make informed and voluntary choices), informational integrity (access to balanced and transparent communication), distributive fairness (equitable exposure to opportunities), and relational trust (confidence in institutional actors). Transformative marketing scholarship suggests that sustainable brand equity depends on maintaining these dimensions over time rather than eroding them through short-term tactics.

Importantly, consumer well-being is not merely an ethical ideal; it also has strategic implications. Empirical research demonstrates that perceived authenticity and transparency in digital communication positively influence long-term loyalty and engagement [48]. Conversely, when consumers perceive deception or manipulation—whether through opaque algorithmic targeting or misleading interface design—trust declines and negative word-of-

mouth increases [36]. Thus, well-being-oriented strategies may align ethical and commercial objectives over the long term.

The integration of transformative marketing with AI ethics also requires attention to vulnerability. Not all consumers interact with digital systems from equal positions of power or literacy. Cultural differences in privacy perceptions and algorithmic awareness further complicate uniform ethical design [25,67]. A well-being framework therefore demands contextual sensitivity rather than universal assumptions.

In the context of influencer and virtual influencer marketing, transformative considerations become particularly pronounced. While disclosures can enhance transparency and support informed choice [42], poorly implemented or strategically obscured disclosures risk undermining autonomy. Virtual influencers introduce additional complexity, as consumers may attribute authenticity or emotional connection to algorithmically generated personas [47,49]. From a transformative perspective, such practices raise questions about emotional exploitation and identity construction within commercial environments.

Thus, transformative marketing extends AI ethics beyond compliance and technical mitigation. It frames algorithmic design decisions as moral choices with societal implications. Rather than asking merely whether systems comply with legal requirements, this perspective asks whether they contribute to human flourishing in digital ecosystems.

2.4. Integrating Machine Learning Ethics with Transformative Marketing

The preceding sections established two complementary theoretical foundations: AI ethics principles and transformative marketing theory. The integration of these perspectives provides a comprehensive framework for evaluating machine learning applications in digital marketing. While AI ethics articulates normative principles such as fairness and accountability, transformative marketing situates these principles within a broader commitment to consumer well-being and societal sustainability.

At the center of this integration lies the concept of ethical machine learning design. Ethical design refers to the deliberate incorporation of fairness constraints, transparency mechanisms, explainability tools, and oversight processes within algorithmic development pipelines. Rather than treating ethics as an afterthought, ethical design embeds normative considerations at each stage of the data lifecycle—from data collection and preprocessing to model training, deployment, and auditing [61,62].

The first theoretical linkage in the proposed model connects ethical design to perceived transparency. When consumers are provided with meaningful explanations regarding personalization practices—such as why a recommendation appears or how data are used—they are more likely to perceive communication as transparent [11,13]. However, transparency must be calibrated to avoid cognitive overload or unintended trust erosion [64]. Balanced explanations that prioritize user-relevant information appear particularly promising [17].

Perceived transparency, in turn, influences consumer trust. Trust functions as a mediating variable between technical design and relational outcomes. Studies on AI disclosure indicate that when transparency aligns with expectations of fairness and respect, trust increases; when transparency highlights impersonal automation without contextual framing, trust may decline [64]. Thus, transparency alone is insufficient—its effectiveness depends on alignment with broader ethical signals.

Trust then contributes to brand legitimacy. Legitimacy extends beyond transactional satisfaction to encompass perceptions of moral appropriateness and institutional reliability. In digital ecosystems characterized by rapid technological change and regulatory scrutiny, legitimacy becomes a strategic asset. Firms perceived as responsible stewards of data and algorithmic systems are better positioned to sustain long-term stakeholder relationships [61,65].

Finally, brand legitimacy supports long-term sustainable equity. Sustainable equity refers not only to financial performance but also to reputational capital and resilience within volatile regulatory and technological environments. As evidence accumulates that manipulative practices generate short-term gains but long-term erosion of trust [36], the strategic case for ethical design becomes clearer.

The integrated conceptual pathway can therefore be summarized as follows:

Ethical ML Design

- Perceived Transparency;
- Consumer Trust;
- Brand Legitimacy;
- Long-term Sustainable Equity.

This pathway does not imply linear causality but rather a reinforcing dynamic. Ethical design enhances transparency; transparency fosters trust; trust strengthens legitimacy; and legitimacy sustains equity. Conversely, failures at early stages may cascade into reputational damage. Importantly, this model acknowledges trade-offs. Fairness interventions may reduce immediate optimization efficiency [24]. Transparency mechanisms may reveal strategic logic that competitors could exploit. Oversight procedures require investment in governance infrastructure. Yet transformative marketing theory suggests that these costs represent long-term investments rather than short-term liabilities.

Moreover, the integration framework highlights the role of stakeholders beyond firms. Regulators establish baseline standards for data protection and accountability [5]. Platforms shape technical affordances and disclosure formats. Researchers contribute measurement tools for bias detection and deceptive design evaluation [68].

The conceptual contribution of this integrated model lies in bridging technical AI ethics with marketing theory. Much of the existing literature treats algorithmic fairness, transparency, and well-being as separate domains. By linking them within a coherent relational pathway, this framework advances understanding of how technical design decisions translate into psychological and strategic outcomes.

Machine learning in digital marketing cannot be evaluated solely through performance metrics or regulatory compliance checklists. Its legitimacy depends on its capacity to enhance rather than undermine human values. The integration of AI ethics principles with transformative marketing provides a theoretically grounded foundation for this evaluation, setting the stage for empirical investigation in subsequent sections.

3. Machine Learning in Digital Marketing, the Opportunities and Risks

The integration of machine learning into digital marketing has not simply enhanced existing tools; it has redefined the architecture of persuasion, targeting, and consumer engagement. As Section 2 demonstrated, algorithmic infrastructures now shape visibility, relevance, and even economic opportunity within digital ecosystems. However, the ethical evaluation of machine learning in marketing cannot rely solely on abstract principles. It must confront the dual reality that these systems generate substantial strategic benefits while simultaneously introducing structural risks.

This section therefore examines both dimensions: the transformative opportunities created by machine learning and the ethical vulnerabilities embedded within algorithmic market systems.

3.1. Strategic Applications of Machine Learning in Digital Marketing

Machine learning applications in digital marketing have expanded rapidly across multiple domains. At the core lies predictive personalization. Advanced models analyze consumer behavior patterns to anticipate preferences and optimize advertising delivery [1,2].

Unlike earlier segmentation techniques, machine learning systems update predictions continuously, allowing campaigns to adapt dynamically to real-time behavioral signals.

Recommender systems represent one of the most visible applications. Deep learning architectures process transactional histories, contextual cues, and network interactions to generate product suggestions tailored to individual profiles [9]. These systems increase click-through rates and conversion probabilities by presenting content perceived as relevant and timely. Similarly, predictive analytics models estimate customer lifetime value, churn probability, and cross-selling potential, enabling firms to allocate resources more efficiently [10].

Sentiment analysis and natural language processing further extend machine learning's reach. By analyzing social media posts, reviews, and user-generated content, marketers can detect emotional trends, brand perceptions, and emerging preferences at scale. This capability enhances responsiveness to consumer sentiment and allows firms to refine messaging strategies in near real time.

Dynamic pricing models constitute another significant development. Algorithms adjust prices based on demand fluctuations, competitor behavior, and inferred willingness to pay. While economically efficient, such systems also raise questions regarding distributive fairness and transparency.

More recently, generative AI has introduced automated content creation into the marketing pipeline. AI-generated advertisements, personalized email copy, chatbot interactions, and even virtual influencers are increasingly common [50,69]. These systems reduce production costs and accelerate campaign deployment, but they also intensify concerns regarding authenticity and manipulation.

Collectively, these applications illustrate that machine learning is not peripheral to digital marketing—it is foundational. It enhances scale, precision, and adaptability. Yet this same infrastructural embedding complicates ethical evaluation.

3.2. Opportunities

From a strategic perspective, machine learning delivers measurable advantages. First, personalization improves perceived relevance. Empirical studies demonstrate that AI-driven tailored journeys enhance customer satisfaction and engagement metrics when aligned with user expectations [51]. By filtering informational overload and presenting targeted recommendations, algorithmic systems can improve user experience.

Second, efficiency gains are substantial. Automated bidding systems optimize advertising placements across platforms in milliseconds, reducing manual intervention and improving return on investment [2]. Predictive models allow firms to anticipate market trends, allocate budgets effectively, and reduce waste in campaign spending.

Third, machine learning enables innovation in customer relationship management. Data-driven insights allow firms to design omnichannel strategies that integrate online and offline interactions. AI-based chatbots provide 24/7 support, increasing accessibility and responsiveness.

Fourth, inclusion potential should not be overlooked. Under certain conditions, digital technologies can enable broader market participation by reducing informational asymmetries and expanding access to services [6]. When designed responsibly, AI systems may facilitate inclusive growth rather than exacerbate inequality.

However, these opportunities are contingent upon ethical implementation. Without adequate safeguards, efficiency gains may come at the expense of autonomy, fairness, or transparency.

3.3. Ethical Risks in Algorithmic Marketing Systems

3.3.1. Algorithmic Bias and Discrimination

One of the most documented risks of machine learning systems concerns bias. Models trained on historical data may encode existing inequalities, leading to discriminatory outcomes in targeting or pricing [4,18]. In advertising contexts, empirical studies reveal unequal exposure patterns for employment and financial products across demographic groups [19,20].

Such disparities are not always intentional. They may arise from proxy variables, imbalanced training datasets, or optimization objectives focused exclusively on engagement metrics. Fairness evaluation frameworks have been proposed to mitigate these risks [21,22], yet implementation remains uneven.

A critical challenge lies in the trade-off between predictive accuracy and fairness constraints [23]. Adjusting models to satisfy fairness metrics may reduce short-term performance indicators. Consequently, firms operating under intense competitive pressure may hesitate to adopt bias mitigation strategies [24].

Thus, algorithmic bias in digital marketing is not merely a technical flaw but a structural tension between equity and optimization.

3.3.2. Data Privacy and Surveillance Intensification

Machine learning systems rely heavily on granular behavioral data. The expansion of predictive analytics intensifies data extraction practices, raising privacy concerns. Although regulatory frameworks such as the GDPR establish formal safeguards [5], consent mechanisms often fail to ensure meaningful understanding [7,8].

The reliance on first-party data strategies has been promoted as a more compliant alternative [6,70]. Yet even direct data collection may become intrusive when personalization systems infer sensitive attributes from behavioral patterns.

Moreover, international regulatory fragmentation complicates enforcement [57]. Multi-national corporations may adjust data practices according to local standards, leading to uneven protection across markets.

The ethical concern therefore extends beyond legality. It concerns the normalization of surveillance as an infrastructural condition of digital participation.

3.3.3. Manipulation and Autonomy Erosion

Perhaps the most philosophically challenging risk concerns manipulation. Machine learning systems optimize persuasion strategies based on behavioral feedback loops. As algorithms learn which emotional cues or scarcity signals maximize engagement, they may prioritize psychologically effective tactics over informational integrity.

Dark patterns exemplify this dynamic. Interface designs that obscure cancellation options or exaggerate urgency exploit cognitive biases [26,27]. Research indicates that such practices may produce short-term conversion gains but erode long-term trust [36].

Generative AI further amplifies persuasive capacity. Algorithmically generated content can adapt messaging style, tone, and timing to individual vulnerabilities [66]. This development raises questions regarding autonomy. When systems anticipate and exploit behavioral tendencies at scale, the boundary between personalization and manipulation becomes blurred.

The risk is not that all personalization is manipulative, but that the architecture of optimization prioritizes engagement metrics without evaluating normative consequences.

3.3.4. Opacity and the Transparency Dilemma

Algorithmic opacity remains a persistent challenge. Deep learning systems often function as black boxes, limiting interpretability [9,11]. Explainable AI tools attempt to address this issue by providing user-facing explanations [12].

However, recent findings suggest that AI disclosure may have complex effects. In certain contexts, explicit acknowledgment of algorithmic mediation reduces trust by highlighting impersonal automation [64]. Furthermore, increasing model interpretability may reduce predictive accuracy [17].

This creates a transparency dilemma: insufficient disclosure undermines trust, yet excessive or poorly framed disclosure may also erode confidence. The challenge lies not merely in increasing transparency but in designing context-sensitive explanations aligned with user expectations.

3.3.5. Disclosure and Authenticity in Influencer Marketing

Influencer marketing introduces additional ethical complexity. Disclosure practices aim to clarify commercial intent [42]. Yet evidence indicates that disclosure effectiveness depends heavily on placement, clarity, and cultural context [43,67].

Virtual influencers intensify these concerns. When audiences are unaware of artificial identities, perceived authenticity may increase; once disclosed, trust may decline [48,49]. This dynamic reveals a broader ethical tension between persuasive realism and informational honesty.

3.4. Structural Tensions and Systemic Interdependencies

The ethical risks identified above do not operate independently. They are interconnected through systemic dynamics:

- Enhancing personalization increases data extraction intensity.
- Increasing transparency may affect performance.
- Mitigating bias may reduce optimization efficiency.
- Restricting dark patterns may lower short-term conversion metrics.

These interdependencies demonstrate that ethical challenges in AI-driven marketing are not isolated issues but structural trade-offs embedded within platform ecosystems.

Furthermore, stakeholder responsibilities overlap. Platforms design ranking algorithms; businesses deploy targeting strategies; regulators establish compliance frameworks; and researchers develop mitigation tools.

Thus, evaluating machine learning in digital marketing requires moving beyond binary assessments of “good” or “bad.” It requires analyzing how governance structures, incentive systems, and technological architectures interact.

3.5. Reframing Opportunities and Risks Through Governance

The coexistence of opportunity and risk suggests that machine learning is neither inherently emancipatory nor inherently harmful. Its ethical character depends on design choices, oversight mechanisms, and regulatory alignment.

When ethical principles such as fairness, accountability, and transparency are embedded into model development pipelines [61], machine learning can enhance both consumer experience and long-term brand legitimacy. Conversely, when optimization logic dominates without ethical calibration, systemic vulnerabilities intensify.

The key challenge, therefore, lies in shifting from efficiency-driven to ethics-informed optimization. Rather than asking how algorithms can maximize engagement, firms and platforms must increasingly ask how algorithms can sustain trust.

4. Research Design and Methodology

4.1. Methodological Positioning and Research Design

This article presents a structured thematic review guided by transparent search, screening, appraisal, and synthesis procedures. PRISMA 2020 [71].reporting principles were used as a reference point for documenting the search and selection process where applicable, but the study is not presented as a full systematic review. No protocol was prospectively registered. Following the completion of the review, a retrospective registration was deposited at the Open science Framework and is publicly available at <https://osf.io/32sw8/> (accessed on 14 June 2026). The PRISMA framework informed the search strategy, eligibility criteria, study screening, full-text review, and reporting of the study selection process, and the corresponding flow diagram is presented in Figure 2. The formal review corpus included only peer-reviewed journal articles and peer-reviewed conference proceedings. Foundational legal, regulatory, methodological, policy, and other grey-literature sources were utilized solely to provide a conceptual and regulatory background, and were not regarded as primary evidence. Hence they were not included in the 91 studies that were included in the formal thematic synthesis.

Study selection process (January 2023–September 2025)

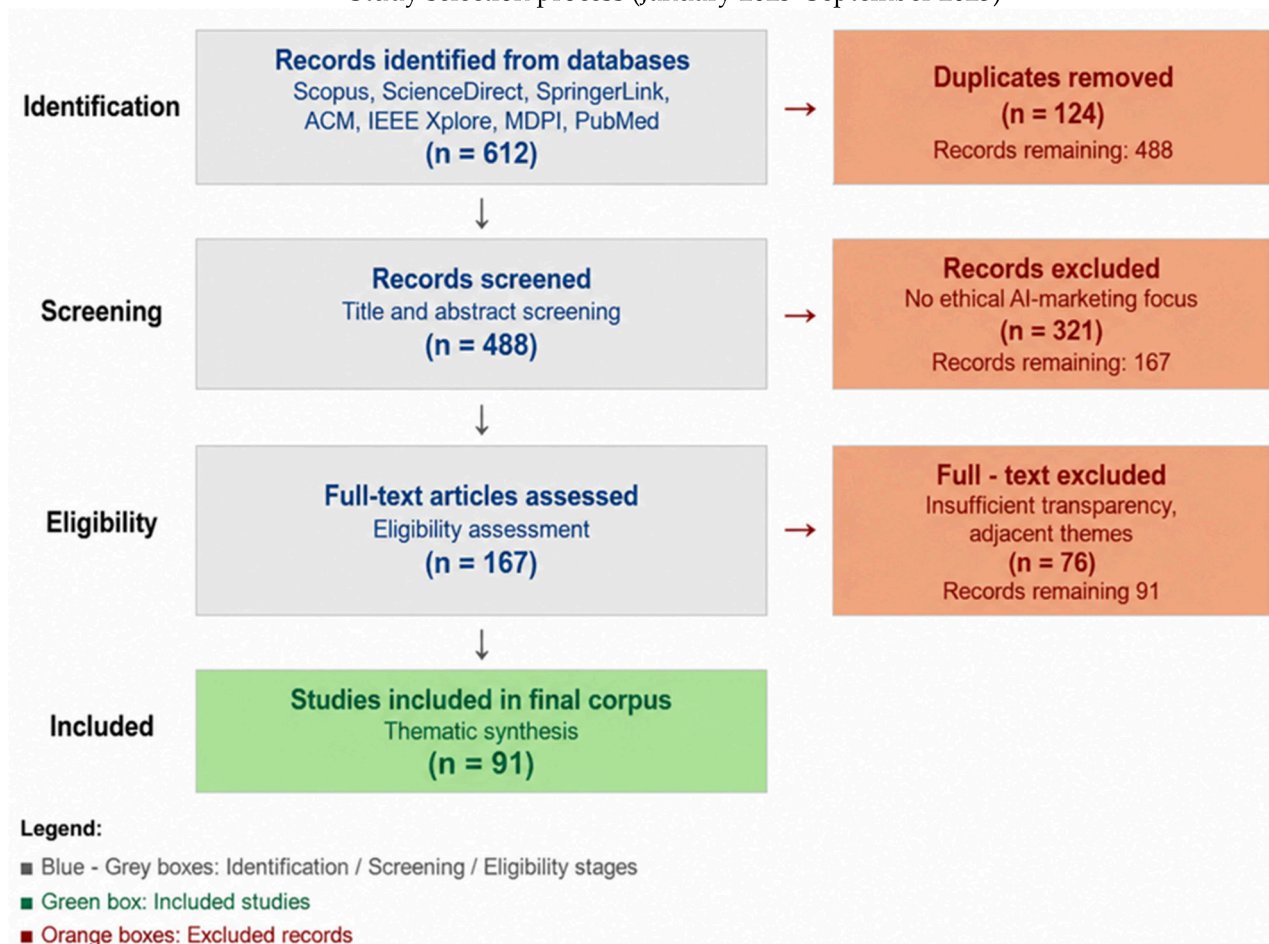


Figure 2. PRISMA-ScR flow diagram.

Conceptual, legal and regulatory background was provided only by grey-literature sources (legal and regulatory documents, policy reports, official institutional publications and foundational books). These sources were not regarded as primary empirical evidence and were not included in the thematic synthesis, which was based on peer-reviewed journal articles and conference proceedings only.

This study is designed and presented as a structured thematic review of recent research, combining systematic elements of evidence identification with interpretive thematic synthesis. The methodological approach follows established standards for systematic identification, screening, and appraisal of the literature, while recognizing the interdisciplinary and heterogeneous nature of research at the intersection of artificial intelligence, digital marketing, and platform governance. Although structured search, screening, and appraisal procedures were applied in the identification and selection phases, the study does not pursue statistical meta-analysis. The reason for this decision lies in the substantial heterogeneity of research designs, empirical contexts, and outcome measures across the selected studies. The corpus includes peer-reviewed experimental investigations, survey-based analyses, qualitative studies, conceptual frameworks, legal analyses published in academic outlets, and technical evaluations of algorithmic systems. These studies differ not only in methodology but also in their operationalization of key constructs such as transparency, fairness, manipulation, and trust. Under such conditions, statistical aggregation would risk oversimplifying conceptual nuances and conflating incommensurable measures. Instead, a thematic synthesis approach was adopted as a more appropriate interpretive strategy. Thematic synthesis enables the systematic identification of recurring ethical domains while preserving contextual specificity and methodological diversity. This approach allows the review to move beyond quantitative aggregation toward structured analytical interpretation, which is particularly suitable for emerging and rapidly evolving research areas where conceptual clarification remains as important as empirical accumulation.

The objective was not merely to catalog existing studies, but to synthesize recent empirical and conceptual contributions in order to identify recurring ethical domains, structural tensions, and governance implications within AI-driven digital marketing. The study therefore combines structured evidence identification (guided by PRISMA 2020 standards [71]) with thematic and interpretive synthesis appropriate for heterogeneous research bodies [72,73].

The chosen time frame—January 2023 to September 2025—was not arbitrary. Artificial intelligence in digital marketing is developing at exceptional speed, particularly following the expansion of generative AI systems in 2023–2024. Restricting the analysis to this period ensured that findings reflect the most current empirical evidence and regulatory developments while avoiding dilution through earlier foundational debates. Foundational texts such as the GDPR [5] and privacy-by-design principles [62] were retained as anchoring frameworks but were not included in the primary evidence corpus unless directly referenced in empirical studies.

The review therefore occupies a structured thematic position: transparent in search and selection procedures, thematic in organization, and interpretive in its synthesis of governance implications. The PRISMA 2020 checklist [71] is provided as Supplementary Table S1.

4.2. Conceptual Contribution to Platform Governance Scholarship

While existing research on AI-driven digital marketing has primarily approached ethical concerns through the lens of marketing responsibility, consumer protection, or corporate compliance, this review advances the discussion by positioning these issues within the broader field of platform governance. Rather than treating platforms merely as neutral intermediaries that host advertising activity, this paper conceptualizes them as active governance infrastructures that shape, enable, and constrain ethical outcomes in digital marketing ecosystems. In doing so, it moves beyond a firm-centric understanding of marketing ethics and situates AI-driven personalization within the structural dynamics of platform power, data asymmetry, and algorithmic mediation.

From a governance perspective, AI-enabled marketing cannot be understood solely as a strategic tool deployed by brands. It is embedded within platform architectures that determine visibility, ranking, recommendation, and monetization logics. Platforms govern digital marketing environments through ranking algorithms, recommender systems, content moderation policies, interface architectures, and data access controls. These infrastructural mechanisms do not simply distribute content efficiently; they actively shape which advertisements are seen, how personalization is calibrated, what forms of disclosure are displayed, and how consent mechanisms are designed. Ethical outcomes—whether related to fairness, transparency, privacy, or manipulation—are therefore structurally mediated by platform design choices rather than exclusively by individual corporate intent.

This review contributes to platform governance research by conceptualizing AI-driven marketing ethics as an infrastructural governance problem, where platforms shape ethical conditions through algorithmic curation, data gatekeeping, interface configuration, and enforcement policies. In this context, ethical risks emerge not only from isolated marketing practices but from systemic asymmetries embedded within platform ecosystems. Recommendation systems influence exposure; API restrictions regulate data flows; interface defaults structure user consent; and enforcement mechanisms determine whether harmful practices are penalized or tolerated. By foregrounding these structural elements, the paper reframes AI-driven digital marketing as a domain where governance is enacted technologically, economically, and institutionally.

The stakeholder–challenge matrix developed in this study further advances this platform governance framing by mapping ethical risks across consumers, businesses, regulators, and researchers. Rather than presenting ethical domains as abstract themes, the matrix reveals asymmetries in responsibility, power, and vulnerability across actors operating within platform ecosystems. Consumers disproportionately bear exposure to algorithmic opacity and manipulative design; businesses navigate performance–ethics trade-offs shaped by platform incentive structures; regulators confront enforcement limitations in opaque, cross-border digital infrastructures; and researchers occupy a mediating role between diagnosis and solution design. In this way, the matrix functions not merely as a descriptive synthesis tool, but as a conceptual instrument for analyzing governance gaps and structural accountability imbalances in AI-driven platform markets.

By integrating ethical marketing concerns with platform governance theory, the paper therefore contributes a structural perspective that situates digital marketing ethics within the infrastructural power of platforms. This reframing highlights that responsible AI in marketing cannot be achieved solely through corporate compliance or individual transparency measures; it requires attention to how platform architectures distribute visibility, shape incentives, and institutionalize norms across digital ecosystems. In doing so, the review aligns the ethics of AI-driven marketing with broader debates on platform regulation, infrastructural governance, and algorithmic accountability.

4.3. Scope Definition

This paper is designed as a structured thematic review of recent research on the ethics of AI-driven digital marketing. Its purpose is to identify, organize, and critically synthesize the main ethical concerns that emerged during the rapid expansion of AI-enabled marketing between January 2023 and September 2025. The review focuses on studies addressing data protection and GDPR compliance [5], algorithmic transparency and explainable AI [9,10], algorithmic fairness [4], dark patterns and deceptive interface design [26], and influencer or virtual influencer disclosure [41].

The study does not claim to provide a comprehensive ten-year systematic literature review. Instead, it offers a focused synthesis of recent peer-reviewed journal articles and

high-quality conference papers in order to capture current debates surrounding generative AI, algorithmic personalization, platform-mediated advertising, consumer manipulation, and disclosure practices. PRISMA 2020 principles were used selectively as a reporting reference to improve transparency in the search and screening process [71]. Thematic synthesis was adopted because the included studies were methodologically heterogeneous and not suitable for statistical meta-analysis [73].

Accordingly, the review is best understood as a structured thematic synthesis rather than a full PRISMA systematic review. Earlier foundational, legal, and methodological works were used only for contextual framing and were not treated as part of the formal review corpus unless they directly informed the analysis.

The review followed structured database searches, predefined inclusion and exclusion criteria, and transparent screening procedures. However, a quantitative meta-analysis was not conducted. This decision was based on the diversity of the included literature, which spans empirical experiments, qualitative investigations, conceptual analyses, legal commentaries, policy documents, and interdisciplinary theoretical contributions. The reviewed studies differ substantially in their methods, contexts, outcomes, and operationalization of key ethical concepts. Ethical domains such as algorithmic fairness, transparency, privacy, and manipulative design are also structurally distinct and cannot be meaningfully reduced to comparable quantitative indicators.

Given this diversity, thematic synthesis was considered the most appropriate methodological strategy. It allowed the review to identify recurring ethical patterns, structural tensions, and cross-cutting governance challenges across a heterogeneous body of literature. Rather than statistically pooling results, the review synthesizes interpretive insights in order to map conceptual convergences, methodological robustness, and stakeholder implications. This approach is suitable for interdisciplinary fields where empirical comparability is limited, but thematic integration remains analytically valuable [73].

4.4. Scope Inclusion Criteria

The scope of the review was defined through three guiding research questions:

- i. What ethical challenges are most frequently identified in AI-driven digital marketing research between 2023 and 2025?
- ii. How robust is the empirical evidence across different ethical domains?
- iii. How are responsibilities distributed among stakeholders (consumers, businesses, regulators, researchers)?

To address these questions, inclusion criteria were formulated as follows:

- Publication between January 2023 and September 2025.
- Peer-reviewed journal articles or high-quality conference proceedings.
- Empirical or theoretically grounded studies directly addressing ethical aspects of AI-driven digital marketing.
- Explicit focus on at least one of the following domains:
 - ◇ Data privacy and GDPR compliance.
 - ◇ Algorithmic transparency and explainable AI.
 - ◇ Algorithmic fairness.
 - ◇ Dark patterns and deceptive interface design.
 - ◇ Influencer or virtual influencer disclosures.

Studies focusing on AI in unrelated domains (e.g., healthcare diagnostics without marketing implications) were excluded. Purely technical machine learning papers without ethical discussion were also excluded.

The final corpus consisted of 91 studies, as documented in the PRISMA-style flow diagram (Figure 2).

More precisely, the literature search was conducted following the multi-source strategy to maximize coverage of AI-based digital marketing practices research [74]. The first search was on three general databases, namely Scopus, SpringerLink, and ScienceDirect. They were complemented with targeted searches in other subject-specific databases such as ACM Digital Library, MDPI, IEEE Xplore and PubMed and forward and backward citation searching. Regulatory and policy documents from organizations such as the Federal Trade Commission and the European Commission were consulted only as contextual sources and were not counted among the 91 studies in the formal review corpus. Search terms were based on the thematic axes of the study, including “digital marketing ethics,” “algorithmic fairness” AND “targeting”, “GDPR compliance,” “algorithmic fairness,” “explainable AI” AND “personalization”, “dark patterns” AND “e-commerce”, and “influencer disclosure” AND “AI”. Boolean operators (AND, OR) and truncation methods were used with searches restricted to full-text papers published in English. Full database-specific search strings, filters, search dates, and the number of records retrieved from each source are provided in Table 1. Forward and backward citation tracking was employed to ensure comprehensive coverage [75]. Bibliographic management software (EndNote 20) facilitated deduplication.

Table 1. Database-specific search strategy.

Database	Date Searched	Exact Boolean Search String	Filters Applied	Records Retrieved
Scopus	26 September 2025	TITLE-ABS-KEY(("artificial intelligence" OR AI OR "machine learning" OR "generative AI") AND ("digital marketing" OR advertising OR personalization OR "personalized advertising" OR "recommender systems") AND (ethic* OR privacy OR GDPR OR transparency OR explainability OR fairness OR bias OR "dark patterns" OR disclosure))	English; peer-reviewed articles and conference papers; January 2023–September 2025	145
ScienceDirect	26 September 2025	("artificial intelligence" OR AI OR "machine learning" OR "generative AI") AND ("digital marketing" OR advertising OR personalization OR "personalized advertising") AND (ethics OR privacy OR GDPR OR transparency OR fairness OR "dark patterns" OR disclosure)	English; research articles; January 2023–September 2025	118
SpringerLink	26 September 2025	("artificial intelligence" OR AI OR "machine learning" OR "generative AI") AND ("digital marketing" OR advertising OR personalization) AND (ethics OR privacy OR GDPR OR transparency OR explainability OR fairness OR "dark patterns" OR disclosure)	English; journal articles and conference papers; January 2023–September 2025	102

Table 1. *Cont.*

Database	Date Searched	Exact Boolean Search String	Filters Applied	Records Retrieved
ACM Digital Library	26 September 2025	("artificial intelligence" OR AI OR "machine learning" OR "generative AI") AND ("digital marketing" OR advertising OR personalization OR "recommender systems") AND (ethics OR privacy OR transparency OR explainability OR fairness OR bias OR "dark patterns")	English; peer-reviewed articles and proceedings; January 2023–September 2025	74
IEEE Xplore	26 September 2025	("All Metadata": "artificial intelligence" OR "All Metadata": "machine learning" OR "All Metadata": "generative AI") AND ("All Metadata": "digital marketing" OR "All Metadata": "advertising" OR "All Metadata": "personalization") AND ("All Metadata": "ethics" OR "All Metadata": "privacy" OR "All Metadata": "fairness" OR "All Metadata": "transparency" OR "All Metadata": "explainable AI")	English; journals and conference papers; January 2023–September 2025	63
MDPI	26 September 2025	("artificial intelligence" OR AI OR "machine learning" OR "generative AI") AND ("digital marketing" OR advertising OR personalization) AND (ethics OR privacy OR GDPR OR transparency OR fairness OR "dark patterns" OR disclosure)	English; research articles and reviews; January 2023–September 2025	81
PubMed	26 September 2025	((("artificial intelligence" [Title/Abstract] OR "machine learning" [Title/Abstract] OR "generative AI" [Title/Abstract]) AND ("digital marketing" [Title/Abstract] OR advertising [Title/Abstract] OR personalization [Title/Abstract]) AND (ethics [Title/Abstract] OR privacy [Title/Abstract] OR transparency [Title/Abstract] OR fairness [Title/Abstract]))	English; January 2023–September 2025	29

Note: * indicates truncation (wildcard) used to capture multiple word endings and variations of a root term (e.g., ethic* retrieves ethics and ethical; explain* retrieves explain, explained, and explainability).

This multilayered approach aimed to balance breadth and precision. The initial search yielded 612 records; after duplicate removal and screening, 91 studies met eligibility criteria. Additionally, a bibliography screening and snowballing method was used, where the reference sections of the most relevant papers were examined to identify additional

momentous studies that might not have appeared in the initial search. After deduplication through EndNote and manual screening, a final sample of 91 papers was selected. The distribution of papers across databases is outlined in Table 1.

4.5. Study Selection Process

The selection of studies was documented using PRISMA 2020 reporting principles where applicable [71], ensuring transparency and methodological consistency throughout the review process. The initial database search, complemented by targeted searches and citation tracking, produced a total of 612 records. After removing 124 duplicate entries, 488 unique studies remained for title and abstract screening. During this phase, articles that did not directly engage with ethical issues in AI-driven digital marketing were excluded. As a result, 321 records were removed because they either addressed artificial intelligence without a marketing focus or examined digital marketing without a substantive ethical component.

The remaining 167 articles were retrieved in full and assessed for eligibility. This stage involved a closer examination of methodological robustness, thematic relevance, and explicit engagement with artificial intelligence in marketing contexts. Seventy-six articles were excluded at this point. In several cases, studies addressed adjacent technological themes but did not analyze ethical implications in depth. In other instances, methodological descriptions were insufficiently transparent, limiting confidence in the reported findings. A smaller subset lacked a clear connection to AI-enabled marketing practices.

A detailed overview of the included studies is provided in Supplementary Table S2, which reports the author and year, country or region, publication type, study design, sample or data source, AI-driven marketing context, ethical domain, key findings, quality appraisal tool, quality tier, and thematic category for each study. The narrowing of the evidence base from 612 initial records to 91 selected studies is depicted in the PRISMA-style flow diagram presented in Figure 2 of the manuscript. This structured process enhances replicability and clarifies how relevance and quality considerations shaped the final corpus.

The remaining 91 studies underwent quality assessment using the Mixed Methods Appraisal Tool (MMAT) [76] and the Joanna Briggs Institute (JBI) critical appraisal criteria [77]. This resulted in a thematic synthesis which identified five major ethical domains: (1) data privacy and GDPR compliance, (2) algorithmic transparency and explainable AI (XAI), (3) algorithmic fairness, (4) dark patterns and deceptive design, and (5) influencer/virtual influencer disclosures. The review is limited to peer-reviewed literature from January 2023 up to September 2025.

The transparency and rigor of the selection process are illustrated in Figure 2, which documents the progressive narrowing of the evidence base (flow diagram) from 612 potentially eligible records to 91 studies that contributed to the final synthesis. This structured process strengthens methodological transparency and relevance for current discussions of AI-enabled digital marketing.

4.6. Quality Assessment

Following selection, each study underwent systematic quality appraisal. Two established instruments were employed: the Mixed Methods Appraisal Tool (MMAT) [76], suitable for quantitative, qualitative, and mixed-methods research, and the Joanna Briggs Institute (JBI) critical appraisal checklists [77], particularly useful for qualitative and observational designs. The use of complementary tools allowed the review to accommodate methodological diversity while maintaining evaluative consistency.

Two reviewers independently assessed each study, and any discrepancies were resolved through discussion, a practice aligned with established standards for inter-rater

reliability in systematic reviews [78]. This process minimized individual bias and strengthened the credibility of the appraisal.

The evaluation focused on clarity of research design, transparency in data collection procedures, analytical rigor, and the validity of conclusions. While most studies met acceptable methodological thresholds, recurring limitations were observed. In particular, several articles lacked detailed reporting on how algorithmic systems were evaluated, especially in fairness or explainability contexts. Others offered limited discussion of sampling strategies or relied heavily on laboratory-based experiments without field validation. Such patterns reflect broader challenges within emerging AI research, where rapid technological change often outpaces standardized empirical protocols.

Importantly, quality assessment did not operate as a rigid exclusion mechanism. Rather than discarding studies solely on the basis of minor methodological weaknesses, the appraisal informed the interpretive weighting during synthesis. Research demonstrating higher analytical rigor or clearer methodological transparency was given greater emphasis in the discussion of findings. This balanced approach ensured that the review remained inclusive while preserving analytical robustness.

4.7. Quality Appraisal Results and Interpretive Weighting

To enhance transparency beyond merely describing the appraisal tools, the review also summarizes the distribution of methodological quality across the included studies (Table 2).

Table 2. Methodological quality distribution.

Quality Tier	Number of Studies	Percentage
High	39	43%
Moderate	34	37%
Lower	18	20%
Total	91	100%

Based on MMAT and JBI evaluations, the 91 studies were categorized into three interpretive tiers according to overall methodological robustness:

- High methodological quality: 39 studies (43%);
- Moderate methodological quality: 34 studies (37%);
- Lower methodological quality: 18 studies (20%).

High-quality studies typically demonstrated clearly articulated research designs, transparent sampling procedures, detailed descriptions of algorithmic evaluation methods, and coherent alignment between research questions and analytical techniques. Many experimental and quantitative studies in this category provided statistical justification of findings and discussed limitations explicitly.

Moderate-quality studies often exhibited sound conceptual framing but lacked depth in methodological reporting. In several cases, algorithmic auditing procedures or fairness evaluations were described at a high level without sufficient reproducibility detail. Sampling strategies were occasionally underdeveloped, particularly in survey-based designs.

Lower-quality studies were not excluded, but recurring weaknesses were identified. These included limited transparency in algorithmic evaluation pipelines, absence of validation procedures, overreliance on small laboratory samples without field corroboration, and insufficient discussion of confounding variables. Conceptual papers in this category sometimes proposed normative arguments without empirical grounding.

Importantly, quality appraisal did not function as a binary inclusion filter. Instead, it informed the interpretive weighting of evidence throughout the synthesis. Higher-quality

studies shaped the strongest causal or normative claims, particularly in discussions of algorithmic fairness and transparency effectiveness. Lower-quality or conceptually thin studies were used primarily to map thematic trends and emerging concerns rather than to support definitive conclusions.

This weighted approach ensured inclusivity of emerging research while preserving analytical rigor in interpretive claims.

4.8. Thematic Coding and Synthesis

To synthesize findings across heterogeneous studies, the review employed thematic analysis, a method particularly suited to identifying recurring patterns within interdisciplinary literature [73]. The coding process unfolded in successive stages. Initially, open coding was applied to identify frequently occurring concepts, including terms such as “consent fatigue,” “algorithmic bias,” “deceptive design,” and “transparency dilemma.” These recurring expressions signaled areas of concentrated scholarly concern.

In the second stage, related codes were grouped into broader conceptual categories. For instance, discussions of cookie interfaces, privacy notices, and data minimization practices were clustered under the broader theme of data protection and regulatory compliance. Similarly, studies examining explainable AI, disclosure mechanisms, and algorithmic opacity were integrated into a transparency-related category.

Through iterative comparison and refinement, five overarching thematic domains emerged: data privacy and GDPR compliance; algorithmic transparency and explainable AI; algorithmic fairness; dark patterns and deceptive design; and influencer or virtual influencer disclosure practices. These themes were not imposed a priori but were derived inductively from recurring patterns across the 91 selected studies.

During interpretive synthesis, cross-cutting tensions became apparent. Many studies highlighted trade-offs between performance optimization and fairness safeguards, or between transparency and competitive advantage. Others revealed that interventions designed to enhance trust could inadvertently produce skepticism if perceived as strategic rather than genuine. Recognizing these tensions allowed the review to move beyond descriptive mapping toward a more integrated understanding of systemic ethical challenges.

4.9. Stakeholder–Challenge Matrix Development

To deepen analytical insight, the findings were organized into a stakeholder–challenge matrix, presented in Figure 3 of the manuscript. This matrix maps the five identified ethical domains across four primary stakeholder groups: consumers, businesses, regulators, and researchers. The purpose of this analytical device was not to quantify ethical risk but to illustrate how concerns are distributed across the digital marketing ecosystem. A concern was classified as primary when it appeared frequently across the reviewed studies and was described as having high stakeholder impact, especially in high- or moderate-quality studies. A secondary concern appeared with moderate frequency or impact. A tertiary concern appeared less frequently, was discussed indirectly, or was supported mainly by conceptual or lower-quality evidence.

Stakeholder ↓ / Ethical Challenge →	Data Privacy & GDPR Compliance	Algorithmic Transparency & XAI	Algorithmic Fairness & Bias	Dark Patterns & Deceptive Design	Influencer & Virtual Influencer Disclosure
Consumers	PRIMARY Consent fatigue; cookie manipulation; limited data control; loss of information or autonomy	PRIMARY Black-box personalization; inability to understand or contest algorithmic decisions	PRIMARY Discriminatory targeting (employment, finance, housing); unequal exposure to high-value services	PRIMARY Manipulative UX; artificial urgency; hidden cancellation; exploited cognitive biases	PRIMARY Opaque sponsorships; parasocial exploitation; inability to distinguish AI- generated personas
Businesses	SECONDARY GDPR compliance costs; first-party data transition; cross-border regulatory fragmentation	PRIMARY Transparency- performance trade- off; proprietary model protection; partial disclosure strategies	PRIMARY Structural disincentive: fairness constraints reduce targeting efficiency; no standard fairness metric	SECONDARY Short-term conversion gains vs. long-term trust erosion; ethics-by- design adoption costs	SECONDARY Authenticity- disclosure tension; influencer contract design; virtual influencer brand risk
Regulators	PRIMARY GDPR enforcement gaps; consent interface adequacy; cross-border data governance asymmetries	SECONDARY Auditing opaque algorithmic systems; limited access to proprietary model documentation	SECONDARY Lack of sector- specific fairness standards; EU AI Act applicability to marketing contexts	PRIMARY DSA / FTC enforcement; cross- jurisdictional inconsistency; regulatory arbitrage by multinationals	SECONDARY Inconsistent disclosure rules across jurisdictions; virtual influencer definitional gaps
Researchers	SECONDARY Empirical consent mechanism studies; GDPR compliance measurement; privacy-by-design evaluation	PRIMARY XAI tool development; user- centered explainability evaluation; trust- transparency interaction studies	PRIMARY Fairness metric development; bias detection pipelines; marketing audit frameworks	SECONDARY Deceptive-design detection tools; cross-platform severity indices; field experiment methodologies	TERTIARY Disclosure format experiments; virtual influencer credibility scales; platform-specific norm studies

Figure 3. Stakeholder–challenge matrix: distribution of ethical concerns across the digital marketing ecosystem. The matrix collates evidence from studies related to data privacy & GDPR compliance, algorithmic transparency & explainable AI (XAI), algorithmic fairness & bias, dark patterns & deceptive design, and influencer/virtual influencer disclosure. The categorization reflects the relative emphasis of the literature reviewed between stakeholder groups (consumers, businesses, regulators and researchers) and ethical challenge domains. The evidence was taken from studies on GDPR and privacy-by-design frameworks [5–8,57,62,70,79,80], explainable AI and algorithmic transparency [9–17,25,81], algorithmic fairness and bias [18–24], dark patterns and deceptive design [27–40,61,68,82–88], and influencer and virtual influencer disclosure [42–50,52,55,58,64,67,89–92].

The matrix reflects qualitative intensity based on frequency and emphasis within the literature. Consumers, for example, are consistently portrayed as experiencing multidimensional exposure to privacy risks, algorithmic bias, manipulative design, and opaque influencer practices. Businesses, by contrast, are frequently depicted as navigating tensions between ethical commitments and performance imperatives. Regulators face enforcement asymmetries and cross-border fragmentation, while researchers occupy a dual role, simultaneously diagnosing ethical risks and proposing mitigation frameworks.

By structuring the findings in this way, the matrix highlights the systemic and relational character of ethical challenges. It underscores that issues such as transparency or fairness cannot be attributed solely to individual actors but emerge from the interaction of platform design, market incentives, regulatory frameworks, and user behavior. The matrix therefore functions as an integrative conceptual tool that connects empirical evidence to governance analysis.

4.10. Reliability and Validity Procedures

Several strategies were adopted to enhance reliability and validity. Dual independent coding reduced the likelihood of interpretive bias, while documentation of coding decisions ensured procedural transparency. Peer debriefing sessions were conducted to review thematic categorizations and resolve ambiguities, strengthening analytic consistency [93].

Internal validity was reinforced through triangulation [94]. Findings were compared across methodological approaches, including quantitative experiments, qualitative interviews, legal analyses, and technical evaluations of algorithmic systems. Convergence across diverse methods increased confidence in recurring patterns. External validity, however, remains conditioned by the temporal and linguistic boundaries of the review. Limiting inclusion to English-language studies and a defined time frame inevitably constrains generalizability.

4.11. Reflexivity and Researcher Positioning

Given the interpretive dimension of thematic synthesis, reflexivity is essential. The categorization of ethical domains reflects analytical judgment shaped by empirical prevalence and theoretical orientation. Although the five identified pillars emerged inductively, alternative frameworks might foreground additional dimensions, such as environmental sustainability or labor ethics in AI supply chains.

The authors therefore present the synthesis as a structured mapping rather than a definitive taxonomy. The aim is not to close debate but to clarify dominant patterns and illuminate areas where scholarly consensus remains incomplete.

4.12. Methodological Limitations

Despite careful design, certain methodological constraints must be acknowledged. The temporal restriction to studies published between 2023 and 2025 enhances topical relevance but excludes earlier foundational research that continues to shape the field. The focus on English-language publications may have limited access to region-specific scholarship. Additionally, although multiple databases were consulted, niche or emerging journals may not be comprehensively indexed.

Another limitation relates to the pace of technological change. Generative AI and platform governance mechanisms evolve rapidly, and empirical findings may shift accordingly. The review captures a specific moment within an ongoing transformation.

Nevertheless, the structured search, appraisal, and synthesis procedures strengthen transparency and replicability.

4.13. Transition to Empirical Implications

The methodological approach adopted in this review enables three core contributions. First, it identifies the dominant ethical domains shaping contemporary debates on AI-driven digital marketing. Second, it evaluates the relative robustness of empirical evidence across these domains. Third, it maps the distribution of responsibility among key stakeholders within platform-mediated ecosystems.

The following section builds upon these methodological foundations to interpret the findings more critically, situating them within broader discussions of governance, accountability, and the future trajectory of ethical digital marketing.

5. Data Protection and GDPR

The protection of personal data remains a central issue in discussions on the ethics of digital marketing. With the introduction of the General Data Protection Regulation (GDPR) in the European Union in 2018 [5], a strict framework was established for the collection,

storage, and use of consumer information. The regulation set out key principles—such as data minimization, lawful processing, and the rights of access and erasure—intended to strengthen accountability. Yet, despite this comprehensive legal framework, questions persist about how consistently these principles are applied in practice. In the context of the 2023–2025 period, research shows that companies are called upon to balance the need to leverage data for personalized campaigns with compliance with the GDPR [79].

A key issue that emerges is user consent. Although consent is considered a cornerstone of the GDPR, many studies emphasize that it is often only “informed” in name, as users find it difficult to fully understand the implications of accepting cookie terms and policies [7]. This leads to a kind of “consent fatigue,” where consumers click “accept” without really understanding, limiting the effective protection of privacy [8].

To address this problem, several companies are turning to first-party data strategies. Unlike third-party data, which is at the center of legal and ethical concerns, first-party data is collected directly from the consumer through interaction with the business. Recent research indicates that relying on first-party data can strengthen GDPR compliance while also fostering greater trust and transparency between companies and consumers [6,70]. Simultaneously, the adoption of such strategies is far from straightforward. They demand substantial investment in technological infrastructure and often require marketers to re-design campaigns around new data-driven principles, a process that can be both costly and disruptive [80].

There is also the international regulatory imbalance. Even if the GDPR is considered as a global standard, the US and some other markets offer more flexibility when it comes to data usage and this makes multinational corporations adjust their strategies based on compliance with local regulations [57]. This introduces additional regulatory complexity for compliance and contradicts how consumer privacy is respected worldwide.

5.1. Algorithmic Transparency and Explainable Artificial Intelligence (XAI)

Algorithmic transparency has become one of the most pressing ethical challenges in the use of artificial intelligence within digital marketing. Personalization and recommendation systems now depend on highly complex models that often function as “black boxes,” making their decision-making processes difficult to interpret [9]. This lack of clarity can erode consumer confidence, since individuals are rarely aware of how or why they are targeted with particular campaigns [10].

The concept of explainable AI (XAI) has been introduced as a response to this problem, aiming to provide users with accessible explanations of system outputs [11]. In marketing, XAI typically takes the form of simplified feedback, such as highlighting the data points behind a product recommendation or clarifying the reasons an advertisement was displayed to a particular consumer [12]. Evidence suggests that offering such explanations can increase perceived transparency and foster more positive attitudes toward platforms [13].

Yet, the pursuit of transparency is not without complications. While greater openness can build trust and support accountability [14], excessive disclosure risks overwhelming users with information or exposing commercially sensitive details that businesses may wish to protect [15]. Some researchers point out that “partial transparency” or the use of multi-level explanations tailored to the audience’s level of understanding may be a more practical solution [16].

Another important tension lies between trust and effectiveness. While explanations can strengthen user confidence, several studies caution that they may also reduce the efficiency of algorithmic systems, particularly when models need to be modified to make them more interpretable [17]. The challenge, therefore, is not only to increase

transparency but also to balance accountability with the performance demands of digital marketing technologies.

The recent literature points toward the development of marketing-specific forms of XAI, tailored to decision-making aspects that matter most for consumer understanding [81]. For instance, clarifying why a product was recommended is often more meaningful to users than providing a technical breakdown of an algorithm's inner workings. Simultaneously, scholars emphasize the importance of cross-sectoral and cross-cultural research, as perceptions of transparency differ widely across regions and cultural contexts [25].

Beyond the technical debate over explainability tools, a deeper structural tension becomes visible. Transparency is often presented as a trust-enhancing mechanism, yet its implementation occurs within performance-driven platform ecosystems where algorithmic opacity may serve strategic interests. Platforms and advertisers operate in highly competitive environments in which predictive accuracy and behavioral optimization are central economic assets. Full transparency may therefore conflict with proprietary logics, revenue models, or competitive positioning. In this sense, the transparency dilemma is not merely cognitive or technical—it is economic. The question is not only whether algorithms can be explained, but whether existing incentive structures genuinely reward explainability over performance efficiency.

Moreover, there is emerging evidence that excessive or poorly contextualized disclosure may paradoxically reduce trust, especially when it highlights impersonal automation in domains where consumers expect human judgment. This suggests that transparency is not inherently trust-producing; its effects are mediated by expectations, design framing, and platform culture. The literature thus reveals an unresolved contradiction: transparency is normatively demanded yet structurally constrained by optimization logics embedded in platform governance architectures.

5.2. Algorithmic Fairness

The application of algorithms in digital marketing offers clear advantages in terms of efficiency and precision, yet it also raises significant concerns about fairness. Because models are frequently trained on datasets that reflect existing social inequalities, their use can reproduce or even intensify those disparities [4]. Such risks are especially evident in advertising and pricing, where consumers may face unequal treatment on the basis of demographic characteristics [18].

Recent studies published mostly between 2023 and 2025 highlight concrete cases in which targeting algorithms excluded specific groups from access to high-value products or services [19]. Lower-income consumers, for instance, were less likely to receive offers for financial products, while women and members of ethnic minorities were disproportionately shown fewer advertisements for prestigious job opportunities [20]. These outcomes not only undermine equal opportunity but also bring into focus the broader ethical responsibilities of companies deploying these technologies.

The literature consistently indicates that fair algorithms for respective steps in the advertising process should be developed and used, so that bias can be evaluated in each stage [21]. Several measures are suggested: among others, demographic parity, equal opportunity and individual fairness that look at different aspects of equality [22]. Meanwhile, tools are also undergoing development for un-biasing algorithms including scaling of example weights during the train phase or the application of fairness constraints on models [23]. But their solutions have gained little attention in the actual setting of digital marketing.

One major challenge is that corporations have little reason to develop fairness mechanisms at all: even if doing so may undermine campaigns [24]. In addition, the notion of

fairness is not culturally neutral; different markets have different ideas of what constitutes “fair” treatment [25]. Hence, setting common expectations worldwide can be complicated and necessitates engagement between investigators, regulators and industry.

While fairness metrics and bias mitigation techniques are increasingly sophisticated, their real-world adoption remains limited. A central tension lies in the economic structure of digital advertising ecosystems. Hyper-targeted marketing depends on differential segmentation, behavioral prediction, and optimization for engagement or conversion. Introducing fairness constraints may reduce predictive granularity and short-term efficiency. Consequently, businesses face a structural disincentive to prioritize equity when competitive advantage is measured primarily through performance metrics.

Additionally, fairness itself is not a universally stable concept. Normative interpretations differ across regulatory systems and cultural contexts, complicating global implementation. What constitutes equitable treatment in one jurisdiction may not align with expectations elsewhere. These discrepancies expose an under-theorized dimension of algorithmic governance: fairness is simultaneously technical, normative, and political. The literature therefore suggests that algorithmic fairness cannot be addressed solely through technical correction mechanisms. It requires institutional alignment between economic incentives, regulatory frameworks, and platform governance architectures.

6. Dark Patterns/Misleading Design

6.1. Introduction to the Phenomenon of Dark Patterns

The term dark patterns refer to design strategies in digital interfaces that intentionally steer users toward choices they might not otherwise make [26]. Common examples include pre-ticked consent boxes, misleading or intrusive notifications, deliberately hidden cancellation options, and persistent prompts encouraging purchases or subscriptions [27]. These tactics blur the line between persuasive design and manipulation, raising concerns about consumer autonomy and the ethical boundaries of digital marketing practices. In digital marketing, dark patterns are a particularly worrying practice, as they raise issues of not only an ethical but also a legal nature, especially in light of the GDPR [5].

6.2. Types of Dark Patterns

The literature recognizes various categories of dark patterns. Among these, the following stand out:

- Privacy Suckering: Nudging users into revealing more information than they want to [28].
- Roach Motel: A user interface that is easy to get into but hard to cancel [29].
- Confirm shaming: Using guilt-inducing or negative language (e.g., “No I don’t care about improving my life”) to try to prevent users saying no to an offer [30].
- Forced Continuity: An undesirable subscription and unclear expiry notification [31].
- Nagging: Annoying repeated interruptions and notifications which force users to perform a certain action [32].

In the context of e-commerce, such tactics are often justified as ways to increase conversion rates. Nevertheless, they stand in direct tension with the principles of transparency and consumer autonomy, raising broader ethical and regulatory concerns.

6.3. Dark Patterns in Cookies and Privacy Policies

The most common manifestation of dark patterns concerns cookie banners. On many websites, the “Accept All” option is more visible than the “Reject” or “Manage Settings” options, which encourages users to share more data [7,33]. This practice raises questions about whether the consent provided is truly free and informed, as required by the GDPR.

Research findings show that 70% of users in experimental studies accept all categories of cookies when “accept” is presented with visually prominent buttons, even if they had previously expressed concerns about privacy [34].

6.4. Dark Patterns in Retail Platforms

Ecommerce channels are particularly rich territory for the dark patterns. One tactic is to make the scarcity cues artificial, for example by writing “Only 1 item left!”, which unduly pressures customers into rushed buying decisions. Other techniques to create artificial scarcity are by using countdown timers indicating short usage opportunity, exaggerated high-demand messages, and fake stock availability representations. Although such strategies are likely to be effective in increasing sales, they raise ethical issues of manipulation, informed consent and trust decay on the part of consumers within online environments as in “5 other users are looking at this product right now” [35]. Additionally, time constraints by means of countdown timers can prompt impulse buying [82]. Evidence from research conducted in European markets indicates a short-term sales boost but long-term erosion of consumer confidence by such practices [83].

Also, practices of drip pricing are prevalent whereby the total purchase price is revealed only at the point of checkout [84]. This has been called out specifically as a misleading practice by regulators, because it makes price comparison hard and dilutes competition.

6.5. Legal and Regulatory Dimensions

The regulatory response to dark patterns remains in a state of development. Within the European Union, both the GDPR and the Digital Services Act contain provisions designed to curb deceptive design practices [38]. Yet enforcement is uneven across member states, leaving notable gaps in consumer protection [39]. In the United States, the Federal Trade Commission (FTC) has issued warnings to firms employing such tactics, but the absence of a comprehensive and consistent legal framework reduces the impact of these interventions [40].

Scholarly discussions highlight the importance of a coordinated international approach. Multinational companies often take advantage of uneven regulations by deploying dark patterns in jurisdictions where oversight is weaker, thereby undermining efforts made in regions with stricter enforcement [85].

6.6. Impact on Consumer Trust and Behavior

Although some dark patterns may create temporary benefit, they are more likely to cause lasting damage to the relationship between businesses and consumers. Research has found that if users perceive deception then they are additionally less likely to be loyal and more likely to share negative experiences through word-of-mouth [36]. In addition, manipulative design exposure leads to a more robust oppositional response—or reactance—which reduces the effect of later marketing attempts [37]. These findings indicate that dark pattern usage is ultimately a high-risk strategy, where the loss of customer trust and loyalty may outweigh any short-term commercial gain.

While dark patterns are frequently discussed as isolated manipulative tactics, their persistence suggests deeper structural drivers. Many platform-based revenue models depend on maximizing engagement, click-through rates, and subscription retention. Interface designs that increase friction for cancellation or emphasize urgency cues may therefore align with short-term monetization goals. In this sense, deceptive design is not just an ethical deviation, but can serve as a rational strategy within systems optimized for engagement and conversion metrics.

This effect is empirically documented in a number of platform and sector settings. Large retail marketplaces in e-commerce have been reported to use artificial scarcity cues and pre-selected upsell options that leverage cognitive load at checkout [35,84].

In the subscription economy, streaming and software services that have been the subject of FTC enforcement actions have deployed cancellation flows that purposefully add steps to the unsubscribe process [29,40]. The cookie consent interfaces of large social media platforms have been shown to display “Accept All” buttons with much higher visual salience than granular opt-out controls, with experimental evidence showing acceptance rates above 70% under those design conditions [7,34]. These sector-specific examples show that deceptive design is not evenly distributed across digital ecosystems, but rather takes on forms adapted to the affordances and revenue logics of each platform environment. The regulatory responses also differ accordingly, with the GDPR and Digital Services Act regulating EU-based interfaces [38], and the FTC framework applying in the US context [40]. Cross-border platforms can take advantage of regulatory asymmetries between countries to engage in compliance arbitrage [57,85].

Regulatory responses attempt to curb such practices, yet enforcement asymmetries and cross-border fragmentation limit deterrence. Platforms simultaneously serve as infrastructure providers and commercial actors, creating potential conflicts of interest in self-regulation. The literature thus indicates that dark patterns persist not because ethical guidelines are absent, but because governance mechanisms struggle to counteract revenue incentives embedded within digital platform architectures.

6.7. Directions for Research and Practice

Research activity suggests the development of standardized methods for measuring the degree of deceptive design so that different platforms can be evaluated comparatively [68]. Simultaneously, it is recommended that field experiments be conducted to understand the long-term effects of dark patterns on consumer behavior. From an administrative point of view, companies are called upon to adopt the principle of ethics by design, incorporating transparency and user autonomy requirements from the design stage [61].

7. Disclosures by Influencers and Virtual Influencers

7.1. Introduction to the Issue of Disclosures

The rise of influencer marketing has significantly reshaped the digital marketing landscape. By leveraging what appears to be an authentic relationship with their audience, influencers exert strong influence over consumer attitudes and purchasing decisions [41]. Yet this practice is accompanied by persistent concerns regarding transparency and ethics, as commercial collaborations are not always made explicit. The literature of 2023–2025 places particular emphasis on disclosure, underscoring its role as a critical safeguard for consumer awareness and informed choice [42].

7.2. Forms and Clarity of Disclosures

Disclosures may appear in different formats, including hashtags such as #ad or #sponsored, explicit text in the description, or graphical overlays within the content itself. The effectiveness of these signals depends largely on their clarity and placement. Empirical studies demonstrate that disclosures shown at the beginning of a post are more likely to enhance perceived transparency, whereas those placed at the end or embedded in less visible sections are often overlooked by users [43]. This suggests that the design and positioning of disclosures are not merely formalities but critical factors shaping consumer awareness and trust. Researchers have pointed out that consumers often overlook or do not fully understand disclosures, especially when they are expressed with ambiguous hashtags [44].

7.3. Impact on Consumer Reactions

The effect of disclosures on consumer reactions is complex. On the one hand, transparency enhances trust and contributes to a positive image of the influencer [42]. On the other hand, in some cases, disclosures lead to a decrease in perceived authenticity, resulting in a weakening of the message's persuasiveness [45,46]. Findings in the literature suggest that the effectiveness of disclosures is not uniform but contingent on several contextual factors. The type of product being promoted, the degree of trust already established between the influencer and the audience, and broader cultural expectations around transparency all shape how disclosures are received [67]. In markets where influencer credibility is highly valued, even clear disclosures may have limited impact, as audiences may discount their significance in favor of loyalty to the personality. Conversely, in contexts with stronger regulatory traditions, inadequate or poorly placed disclosures can generate backlash and erode both consumer trust and brand reputation.

7.4. Virtual Influencers and New Challenges

The emergence of virtual influencers—digitally created characters that promote products—introduces new dimensions to the issue of transparency. Virtual influencers are capable of producing visually striking and tightly controlled content, yet their artificial nature introduces concerns about the authenticity of communication [47]. Evidence suggests that when audiences are aware of the virtual identity of an influencer, levels of trust and purchase intention decline [48]. By contrast, when this fact is concealed, responses are generally more favorable, a dynamic that creates ethical dilemmas regarding the necessity of disclosure [49].

The growing use of generative artificial intelligence to design and manage virtual influencers deepens these concerns. Such technologies blur the line between reality and digital fabrication, making it increasingly difficult for consumers to distinguish authentic communication from algorithmically generated content [50]. In light of these challenges, scholars argue for targeted regulatory measures that would establish clear disclosure requirements and safeguard consumer autonomy [90].

7.5. Legal and Regulatory Framework

The regulatory response to influencer disclosures differs across jurisdictions. Within the European Union, the Consumer Protection Cooperation Regulation and directives issued by the European Commission mandate that promotional content must be accompanied by clear disclosures [52]. In the United States, the Federal Trade Commission (FTC) has established guidelines requiring disclosures to be both clear and conspicuous [40]. Despite these frameworks, enforcement remains inconsistent, and many collaborations are still hidden or presented in ways that obscure their commercial intent [55].

The literature further points to the problem of global asymmetry. Misleading practices are more prevalent in markets where regulation is weaker or less consistently applied [57]. This uneven landscape poses particular difficulties for multinational companies, which must adjust compliance strategies to fit varying national requirements. As a result, fragmented regulation not only undermines consumer protection but also complicates efforts to establish shared ethical standards in digital marketing.

7.6. Consumer Experience and Ethical Implications

Empirical research consistently indicates that the way disclosures are presented significantly shapes consumer response. When sponsorships, algorithmic recommendations, or commercial relationships are communicated clearly, prominently, and in language that is easy to understand, consumers tend to react more positively [42]. Transparent disclosures

do not necessarily undermine persuasion; in many cases, they appear to preserve credibility while maintaining advertising effectiveness. By contrast, when disclosures are hidden in lengthy captions, ambiguous hashtags, or visually minimized interface elements, users often perceive the communication as deceptive. This perception can trigger skepticism and distrust, sometimes extending beyond the specific campaign to the broader brand relationship [36]. Thus, the issue is not disclosure per se, but how disclosure is operationalized within digital environments.

From the consumer's standpoint, disclosure functions as a mechanism of informational balance. It provides context that allows individuals to interpret content appropriately and to distinguish between organic expression and commercial persuasion. When this contextual signal is absent or obscured, consumers are deprived of a meaningful opportunity to evaluate intent. In that sense, lack of transparency can be understood not merely as an omission, but as a form of informational asymmetry that limits autonomous judgment [42]. Ethical concerns therefore extend beyond formal compliance. The question is whether disclosure practices genuinely enable informed choice, or whether they are designed primarily to satisfy minimal regulatory thresholds while preserving persuasive advantage.

The ethical stakes become even more pronounced in influencer and virtual influencer marketing, where the boundary between personal expression and advertising is often blurred. Influencers cultivate authenticity and relational closeness with their audiences; their recommendations are frequently perceived as advice rather than promotion. In such contexts, disclosure carries symbolic as well as practical significance. It signals respect for the audience's capacity to make informed decisions. When disclosures are inconsistent, vague, or strategically ambiguous, the relational trust that underpins influencer communication may erode. This erosion is not only reputational; it touches on deeper concerns about manipulation and the exploitation of parasocial relationships.

More broadly, disclosure practices reflect a wider normative question about responsibility in digital communication ecosystems. While brands design campaigns and influencers produce content, platforms shape the technical affordances that determine how visible or salient disclosures actually are. Interface layout, algorithmic ranking, and content formatting all influence whether transparency is noticed or overlooked. Consequently, ethical accountability cannot be assigned solely to individual actors. It is distributed across brands, creators, and platform infrastructures. Clear and enforceable disclosure standards therefore serve a dual function: they support legal compliance, but they also contribute to sustaining trust as a structural feature of the digital marketplace. In environments characterized by personalization, data extraction, and algorithmic curation, disclosure becomes one of the few visible signals of fairness and respect for consumer agency.

In this light, consumer experience is inseparable from ethical design. Transparency that is intelligible, consistent, and context-sensitive strengthens not only immediate campaign credibility but also long-term relational stability. Conversely, opaque practices may yield short-term engagement gains but risk cumulative trust deficits over time. The literature suggests that disclosure should not be treated as a peripheral regulatory obligation; rather, it constitutes a foundational element of responsible digital marketing practice, shaping how consumers interpret, evaluate, and ultimately trust the increasingly algorithmic marketplace.

7.7. Recommendations for Research and Practice

Recent scholarship increasingly points to the need for more coherent and standardized disclosure frameworks that move beyond generic labeling requirements. Importantly, disclosure does not operate in a vacuum; its effectiveness depends heavily on the medium through which it is delivered. Platforms such as Instagram, TikTok, and YouTube differ

significantly in interface design, content format, user attention patterns, and algorithmic amplification. What counts as clear and salient disclosure in a long-form YouTube video may not translate effectively to short-form, fast-scrolling environments like TikTok. As a result, scholars argue that regulatory approaches must account for platform-specific dynamics rather than relying on one-size-fits-all language [58,91]. This becomes even more pressing with the emergence of virtual influencers—algorithmically generated personas that blur the boundary between human authenticity and synthetic production. Their hybrid nature raises unresolved definitional and regulatory questions: are they to be treated as corporate advertising avatars, digital characters, or autonomous creative entities? Without conceptual clarity, regulatory responses risk lagging behind technological innovation.

In addition to conceptual refinement, there is a clear call for stronger empirical grounding. Much of the existing literature relies on laboratory experiments or survey-based perceptions, which, while valuable, do not always capture the complexities of real-world digital campaigns. Field-based experimental research could provide deeper insight into how different disclosure formats perform in live environments characterized by algorithmic ranking, multitasking users, and fragmented attention [92]. For example, comparative studies examining text-based disclosures, visual tags, verbal acknowledgments, and embedded platform labels within actual campaigns could clarify not only which formats enhance understanding, but also how they influence trust, engagement, and long-term brand perception. Such research would help move the debate beyond normative assertions toward evidence-based policy and managerial recommendations.

For businesses, the literature increasingly recommends adopting transparency by design as an organizing principle rather than treating disclosure as a final compliance step [62]. This implies integrating clarity about sponsorship, data use, and commercial intent at the earliest stages of influencer collaboration planning. When transparency is embedded structurally—through contract design, briefing guidelines, and content co-creation processes—it is less likely to be perceived as reluctant or performative. From a strategic perspective, proactive transparency may also function as a reputational asset, signaling ethical commitment in an environment where consumers are increasingly sensitive to manipulation.

For policymakers, the review underscores the importance of clearer interpretive guidance and more consistent enforcement across jurisdictions [85]. Fragmented regulatory landscapes create uncertainty for multinational brands and uneven levels of protection for consumers. Harmonized standards—particularly within cross-border digital markets—could reduce ambiguity and limit opportunities for regulatory arbitrage. At the same time, enforcement mechanisms must be sufficiently resourced and technologically informed to address opaque algorithmic systems and evolving influencer practices. Strengthening clarity and consistency in disclosure regulation is therefore not only about preventing deception; it is about sustaining public confidence in digital communication infrastructures that are becoming progressively more automated and data-driven.

Taken together, these recommendations highlight that disclosure governance must evolve alongside platform architectures and AI capabilities. Conceptual clarity, empirical validation, strategic integration, and regulatory coordination are all necessary components of a disclosure ecosystem capable of supporting both innovation and consumer protection.

7.8. Authenticity vs. Regulation Tension

The debate surrounding influencer disclosures reveals a structural tension between authenticity and regulation. Influencer marketing derives persuasive power from perceived intimacy and relational trust. Regulatory requirements for clear sponsorship disclosure may protect consumer autonomy, yet they can also disrupt the narrative conti-

nuity that sustains influencer credibility. This tension is particularly pronounced in the case of virtual influencers, where authenticity is algorithmically constructed rather than organically developed.

Platform moderation practices further complicate enforcement. Disclosure standards vary across jurisdictions, and global platforms often apply uneven enforcement thresholds. As a result, ethical communication becomes contingent not only on brand intention but on platform policy design and algorithmic content ranking systems. The literature therefore suggests that disclosure challenges are not merely about clarity of labeling; they are embedded in broader questions of platform governance consistency, cross-border enforcement, and the commercialization of authenticity itself.

8. Discussion and Research Directions

This review has highlighted five key areas at the intersection of artificial intelligence and the ethics of digital marketing: data protection, algorithmic transparency, fairness, dark patterns, and influencer disclosures. A recurring theme across these domains is the growing tension between technological innovation and the safeguarding of consumer rights. The rise of generative AI amplifies these concerns, as it enhances personalization capabilities while simultaneously creating new risks of manipulation [50].

Figure 3 depicts the five key categories of ethical dilemmas identified throughout the body of literature reviewed in an AI-based digital marketing environment as presented to four key stakeholders, namely consumers, businesses, regulators and researchers. It is underpinned by 91 peer-reviewed studies and is a qualitative mapping instrument rather than a quantitative product. All cells are visually coded into three levels of priority: primary concern (dark blue) denotes issues that appear with a high frequency and emphasis in the studies reviewed; secondary concern (medium blue) denotes issues of medium salience, and partial coverage; and tertiary concern (light blue) denotes emerging or indirectly addressed issues. These levels are not formal quantitative scores but show the frequency and emphasis with which each concern was attributed to each stakeholder group across the reviewed literature.

From a consumer's point of view, the profile indicates an even spread of ethical risks across all five domains. Privacy depletion and consent fatigue emerge as leading issues [5,7,8], closely followed by algorithmic transparency and explainability challenges [9,10,13], algorithmic unfairness and discriminatory end products [18–20], manipulative interface practices and dark arts design patterns [26,27,36] and inadequate disclosure in influencer/virtual influencer marketing [42,47,48]. These results reflect the continued lack of information symmetry between consumers and producers, which constrains the capacity of consumers to make fully informed choices and reduces the efficacy of entirely market-based mechanisms for remediation.

Businesses, on the other hand, are approached as trade-offs between compliance and performance. Ethical deployment is often presented in the literature as a trade-off between ideal values of transparency, fairness and privacy on the one hand, and personalization effectiveness, targeting accuracy and commercial competitiveness on the other [14–17], aspects that extensively dominate narratives [21–24]. Although ethical-by-design strategies are receiving growing attention, actual evidence indicates that they remain an exception, regularly compromised by short-term performance drivers.

Regulators face different, but often overlapping, challenges. In different areas, regulators are found to have enforcement gaps and audit challenges with opaque algorithmic systems as well as inconsistencies due to cross-border regulatory fragmentation [38–40,57,85,87,88]. Although frameworks like the GDPR have set down formal responsibilities in data protection and accountability [5], black-box models and

regional divergence reduce the pragmatic enforceability of those rules [38–40]. This indicates that the real-world effectiveness of consumer protection may be compromised by enforcement gaps, opaque algorithmic systems, and cross-border regulatory fragmentation, even while formal regulatory safeguards exist.

Researchers occupy a dual position within the stakeholder–challenge matrix. On the one hand, they document and critically analyze the ethical risks associated with AI-driven digital marketing. On the other hand, they contribute to the development of practical tools for explainable AI, fairness assessment, deceptive design detection, and disclosure measurement. Recent work on explainable AI [11,12,81], fairness metrics and bias mitigation [21–23], deceptive design detection [61,68,86,87], and disclosure effectiveness [21,92] indicates that scholarship is both diagnostic and solution-oriented. At the same time, important gaps remain, particularly the absence of standardized tools for measuring manipulation, algorithmic unfairness, and cross-platform ethical risk [69].

Consequently, the stakeholder–challenge matrix highlights the systemic character of ethical challenges in AI-driven digital marketing. Ethical risks are not confined to single actors or realms, but created more broadly by the interplay of design choices, market incentives, regulatory architectures and user vulnerabilities. A substantial overlap of concerns across the reviewed studies reinforces the need for coordinated governance, interdisciplinary research, and stronger integration between scholarly evidence, industry practice, and regulatory enforcement.

Beyond simply categorizing ethical concerns, the stakeholder–challenge matrix makes visible the extent to which these issues are interdependent. Transparency, for instance, is often presented as a straightforward remedy for algorithmic opacity through explainable AI mechanisms [11,12,81]. Yet the literature repeatedly shows that increased transparency can conflict with commercial interests, particularly where proprietary systems and competitive advantage are at stake [14–17]. Similarly, efforts to correct algorithmic bias through fairness constraints or mitigation techniques [21–23] may affect targeting precision and, by extension, performance metrics that firms rely upon [24]. Even strengthened consent requirements under the GDPR [5,7,8], while normatively justified, may reshape data availability in ways that influence personalization models [16]. These examples illustrate that ethical reform in AI-driven marketing rarely operates in isolation. Instead, it unfolds within a landscape of competing values—efficiency, innovation, privacy, and fairness—where adjustments in one area reverberate across others. Regulatory analyses have already warned that narrowly focused interventions can produce limited impact when underlying incentive structures remain unchanged [38–40].

The matrix also reveals a marked asymmetry in how ethical risks are distributed across stakeholders. Consumers appear consistently as the group most exposed to cumulative harms: diminished data control and consent fatigue [5,7,8], limited visibility into algorithmic decision-making [9,10,13], discriminatory targeting outcomes [18–20], manipulative interface designs [26,27,36], and ambiguous influencer disclosures [42,47,48]. This pattern reflects persistent informational imbalances in platform-mediated markets. Businesses and regulators, by contrast, are portrayed as operating within structural constraints—competitive pressures, compliance costs, and jurisdictional fragmentation—that shape their responses to ethical demands [38–40,57,85]. Although not represented as a standalone category in the matrix, platforms themselves function as infrastructural gatekeepers, structuring visibility, consent flows, and disclosure mechanisms through algorithmic architectures [9,57]. As a result, ethical outcomes are not simply the product of individual firm decisions, but emerge from the interaction between technological design, economic incentives, and regulatory capacity. In contexts where opacity and behavioral nudging are systemic features, reliance on market self-correction mechanisms may be insufficient [36,66]. This asymmetry of

distribution is reflected in the visual coding of the matrix whereby consumers record the highest concentration of primary-level concerns in all five ethical domains.

Looking ahead, the matrix also points toward unresolved theoretical and empirical challenges. The expansion of generative AI systems, with their capacity for adaptive personalization and scalable persuasion [50,69], intensifies longstanding concerns about manipulation, autonomy, and fairness. While important advances have been made in explainability research [11,12,81], bias detection and fairness auditing [21–23], dark pattern identification [61,68,86], and disclosure effectiveness studies [91,92], the field still lacks standardized measures that would allow meaningful comparison across sectors and jurisdictions. In particular, the operationalization of concepts such as autonomy erosion or manipulation remains underdeveloped [25,68]. This suggests that future work must move beyond diagnosing ethical risks toward developing shared evaluative frameworks that can inform both policy and practice. In that sense, the stakeholder–challenge matrix serves not only as a synthesis of 91 studies, but as a conceptual tool for thinking more systematically about how governance structures, technological architectures, and market logics intersect in shaping the ethical trajectory of AI-driven digital marketing.

A recurring implication of the reviewed literature is that ethical implementation is constrained not only by lack of awareness but also by the metric architecture of platform-mediated marketing. Advertisers and brands are typically evaluated through ROI, click-through rates, conversion rates, engagement duration, and campaign efficiency. These metrics reward optimization, personalization intensity, and rapid experimentation, while ethical safeguards such as fairness audits, stronger consent mechanisms, or transparent explanations may reduce short-term performance. As a result, businesses face a structural disincentive: practices that improve privacy, fairness, or transparency may appear costly or inefficient within platform dashboards designed around immediate performance. Ethical AI implementation therefore requires not only firm-level responsibility but also changes in platform metrics, audit expectations, and regulatory incentives.

Operationalized Future Research Priorities

Building on the identified gaps, the following operationalized research priorities are proposed for the field. First, regarding measurement frameworks for algorithmic fairness in marketing, future research should develop and validate sector-specific fairness audit protocols adapted to the distinct data environments of social media advertising (where demographic proxies are platform-assigned), search advertising (where query-based targeting intersects with protected attributes), and e-commerce recommendation (where purchase history may encode socioeconomic disparities). Concretely, researchers could adapt existing fairness metrics such as demographic parity and equalized odds [21,22] to a standardized “Marketing Fairness Scorecard” evaluable via A/B campaign audits, enabling cross-platform comparison. The EU AI Act’s Article 9 risk management requirements and NIST AI Risk Management Framework offer regulatory anchors for such a measurement instrument.

Second, regarding audit methodologies for dark patterns, a pressing need exists for field-deployable audit toolkits capable of systematically detecting and classifying deceptive design across live commercial platforms. Future research should adapt and extend existing detection frameworks [61,68] into automated browser-based audit pipelines that can: (a) screen cookie consent interfaces against GDPR adequacy criteria using predefined visual prominence thresholds; (b) classify checkout flows against a taxonomy of dark pattern types (roach motel, drip pricing, confirm shaming) [27–32]; and (c) generate a platform-comparable “Dark Pattern Severity Index.” Such tools could be piloted in collaboration with national data protection authorities in the EU—building on the precedent of the Nor-

wegian Consumer Council’s “Deceived by Design” audits [87]—and subsequently scaled to regulatory benchmarking programs under the Digital Services Act [87].

9. Conclusions

The review concludes that AI-driven digital marketing is a complex, platform-mediated ecosystem where algorithmic design, platform infrastructure, and governance arrangements together shape ethical outcomes. Machine learning provides unprecedented accuracy in personalization and major strategic efficiencies, but these benefits often come at the cost of consumer autonomy, transparency and fairness. We identify five core ethical pillars, data privacy, algorithmic transparency, fairness, dark patterns, and influencer disclosure, and find a systemic landscape where normative safeguards are often outweighed by commercial incentives for engagement. The “transparency dilemma” and “consent fatigue,” then, expose the persistent information asymmetry that renders consumers disproportionately vulnerable to cumulative harms in these opaque digital marketplaces.

Three limitations qualify these conclusions. The review is intentionally limited to peer-reviewed journal articles and conference proceedings, and excludes grey literature. Industry practice in AI-driven marketing typically moves faster than its documentation in academic outlets. This limitation means that recent platform policy changes, proprietary system audits, enforcement actions, and practitioner-reported mitigation measures—material that circulates primarily through white papers, regulatory communications, and trade publications—fall outside the synthesized evidence base. The findings should therefore be read as a conservative account of the ethical landscape, one that may under-represent both newly emerging harms and platform-level remedies that have not yet been evaluated by independent scholars. Second, the temporal window (January 2023–September 2025) and the limitation to English-language publications further narrow the corpus and may under-represent region-specific scholarship. Third, the heterogeneity of methods in the included studies did not permit statistical aggregation and the synthesis is therefore interpretive rather than quantitative. These limitations do not invalidate the thematic structure identified here, but they do limit the empirical basis for it, and they imply a complementary research agenda for systematically appraising grey literature and practitioner sources alongside peer-reviewed evidence.

Ultimately, a responsible digital marketing future will require a move from optimization to ethics-based governance. The stakeholder–challenge matrix of the study shows that accountability should be shared across platforms, regulators and researchers, and not limited to individual brands. In the future, we need to move beyond reactive compliance, and toward “ethics-by-design” principles that incorporate fairness and explainability at the earliest stages of algorithmic development. The industry can square the circle of technological innovation and the protection of fundamental consumer rights through harmonizing regulatory frameworks and establishing common standards for the evaluation of manipulation and bias.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/platforms4030017/s1>. Table S1: PRISMA-ScR Checklist [95]. Table S2: Aggregate overview of the included studies (n = 91).

Author Contributions: Conceptualization, A.K. and M.M.; methodology, A.K.; software, A.K.; validation, A.K. and M.M.; formal analysis, A.K.; investigation, A.K.; resources, A.K. and M.M.; data curation, A.K.; writing—original draft preparation, A.K.; writing—review and editing, A.K.; visualization, A.K.; supervision, A.K.; project administration, A.K.; funding acquisition, M.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Hema Sundari, M.; Parhana, R.; Santhanalakshmi, K.; Haridharan, K.N. Artificial Intelligence in Digital Marketing: A Systematic Review and Strategic Framework for AI-Augmented Marketing. *Int. Res. J. Adv. Eng. Manag.* **2025**, *3*, 2595–2611. [CrossRef]
2. Li, T. AI Empowered Personalized Digital Advertising Marketing. *Front. Bus. Econ. Manag.* **2024**, *15*, 439–442. [CrossRef]
3. Acatrinei, C.; Apostol, I.G.; Barbu, L.N.; Chivu, R.-G.; Orzan, M.-C. Artificial Intelligence in Digital Marketing: Enhancing Consumer Engagement and Supporting Sustainable Behavior Through Social and Mobile Networks. *Sustainability* **2025**, *17*, 6638. [CrossRef]
4. Yella, S. Ethical Considerations in AI-Enhanced Marketing Automation: Balancing Personalization and Responsibility. *Int. Res. J. Mod. Eng. Technol. Sci.* **2024**, *6*, 8663–8673. [CrossRef]
5. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data (General Data Protection Regulation). Available online: <https://gdpr-info.eu/> (accessed on 14 June 2026).
6. Hui, L.; Xie, H.; Shi, B.; Lu, Y. Does digital technology enable inclusive growth? Evidence from listed firms in China. *Manag. Syst. Eng.* **2025**, *4*, 4. [CrossRef]
7. Nicola, N. Losing control: The effects of social media fatigue, privacy concerns and psychological reactance on social media advertising. *Connect. Istanbul Univ. J. Commun. Sci.* **2022**, *63*, 123–149. [CrossRef]
8. Hildt, E.; Laas, K. Informed Consent in Digital Data Management. In *Codes of Ethics and Ethical Guidelines: Emerging Technologies and Changing Fields*; Springer: Cham, Switzerland, 2022; pp. 51–88. [CrossRef]
9. Benleulmi, M.; Gasmi, I.; Azizi, N.; Dey, N. Explainable AI and deep learning models for recommender systems: State of the art and challenges. *J. Univers. Comput. Sci.* **2025**, *31*, 383–421. [CrossRef]
10. Wilson, G.; Johnson, O.; Brown, W. The Impact of Artificial Intelligence on Digital Marketing Strategies. *Preprint* **2024**. [CrossRef]
11. Nizette, F.; Hammedi, W. Explainable AI in Marketing. In *Encyclopedia of Artificial Intelligence in Marketing*; Springer: Cham, Switzerland, 2025; pp. 1–13. [CrossRef]
12. Al-Ansari, N.; Al-Thani, D.; Al-Mansoori, R.S. User-Centered Evaluation of Explainable Artificial Intelligence (XAI): A Systematic Literature Review. *Hum. Behav. Emerg. Technol.* **2024**, *2024*, 4628855. [CrossRef]
13. Szymczyk, N.; Szczuka, J.M.; Krämer, N.; Wischniewski, M. Explained or Certified? Examining the Influence of XAI and AI-Seals on Users' Trust and Understanding. In *Proceedings of the Mensch und Computer 2025*; Association for Computing Machinery: New York, NY, USA, 2025; pp. 121–129. [CrossRef]
14. Horneber, D.; Laumer, S. Algorithmic Accountability. *Bus. Inf. Syst. Eng.* **2023**, *65*, 723–730. [CrossRef]
15. Buijsman, S. Navigating fairness measures and trade-offs. *AI Ethics* **2024**, *4*, 1323–1334. [CrossRef]
16. Billones, R.K.C.; Lauresta, D.A.S.; Dellosa, J.T.; Bong, Y.; Stergioulas, L.K.; Yunus, S. AI Ecosystem and Value Chain: A Multi-Layered Framework for Analyzing Supply, Value Creation, and Delivery Mechanisms. *Technologies* **2025**, *13*, 421. [CrossRef]
17. Jafari, E. Balanced Explanations in Recommender Systems. In *Proceedings of UMAP '24: ACM Conference on User Modeling, Adaptation and Personalization*; Association for Computing Machinery: New York, NY, USA, 2024; pp. 25–29. [CrossRef]
18. Ferrara, E. Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies. *Science* **2024**, *6*, 3. [CrossRef]
19. Chen, Z. Ethics and discrimination in artificial intelligence-enabled recruitment practices. *Humanit. Soc. Sci. Commun.* **2023**, *10*, 567. [CrossRef]
20. Trautwein, Y.; Zechiel, F.; Coussement, K.; Meire, M.; Büttgen, M. Opening the 'black box' of HRM algorithmic biases—How hiring practices induce discrimination on freelancing platforms. *J. Bus. Res.* **2025**, *192*, 115298. [CrossRef]
21. Yang, M.; Guo, J.; Zhu, L.; Zhu, H.; Song, X.; Zhang, H.; Xu, T. Fairness evaluation of marketing algorithms: A framework for equity distribution. *J. Electron. Bus. Digit. Econ.* **2024**, *3*, 251–274. [CrossRef]
22. Wu, Y.; Cao, J.; Xu, G. Fairness in Recommender Systems: Evaluation Approaches and Assurance Strategies. *ACM Trans. Knowl. Discov. Data* **2023**, *18*, 10. [CrossRef]
23. Siddique, S.; Ariful Haque, M.; George, R.; Datta Gupta, K.; Gupta, D.; Hossain Faruk, M.J. Survey on Machine Learning Biases and Mitigation Techniques. *Digital* **2024**, *4*, 1–68. [CrossRef]

24. Fabris, A.; Baranowska, N.; Dennis, M.J.; Graus, D.; Hacker, P.; Saldivar, J.; Zuiderveen Borgesius, F.; Biega, A.J. Fairness and Bias in Algorithmic Hiring: A Multidisciplinary Survey. *ACM Trans. Intell. Syst. Technol.* **2025**, *16*, 1–54. [\[CrossRef\]](#)
25. Shin, D.; Chotiyaputta, V.; Zaid, B. The Effects of Cultural Dimensions on Algorithmic News: How Do Cultural Value Orientations Affect How People Perceive Algorithms? *Comput. Hum. Behav.* **2021**, *126*, 107007. [\[CrossRef\]](#)
26. Zac, A.; Huang, Y.; von Moltke, A.; Decker, C.; Ezrachi, A. Dark patterns and consumer vulnerability. *Behav. Public Policy* **2025**, *9*, 1–50. [\[CrossRef\]](#)
27. Monge Roffarello, A.; Lukoff, K.; Russis De, L. Defining and Identifying Attention Capture Deceptive Designs in Digital Interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*; Association for Computing Machinery: New York, NY, USA, 2023; pp. 1–19. [\[CrossRef\]](#)
28. Isola, C.; Esposito, F. A systematic literature review on dark patterns for the legal community: Definitional clarity and a legal classification based on the Unfair Commercial Practices Directive. *Comput. Law Secur. Rev.* **2025**, *58*, 106169. [\[CrossRef\]](#)
29. Sheil, A.; Acar, G.; Schraffenberger, H.; Gellert, R.; Malone, D. Staying at the Roach Motel: Cross-Country Analysis of Manipulative Subscription and Cancellation Flows. In *Proceedings of CHI '24: ACM Conference on Human Factors in Computing Systems*; Association for Computing Machinery: New York, NY, USA, 2024. [\[CrossRef\]](#)
30. Knyzelis, M. Creative manipulation: A case study of confirm shaming as a deceptive design pattern. *Creat. Stud.* **2024**, *17*, 335–344. [\[CrossRef\]](#)
31. Howells, G. Protecting Consumer Protection Values in the Fourth Industrial Revolution. *J. Consum. Policy* **2020**, *43*, 145–175. [\[CrossRef\]](#)
32. Kyguolienė, A.; Sokolovaitė, A. Consumers' attitude to manipulating consumers in electronic commerce. *Int. J. Bus. Manag. Stud.* **2021**, *2*, 25–32.
33. Le, H.; Nethala, S.S. Beyond the Banner: Understanding the Impact of Cookie Consent Interfaces on User Data Privacy Choices. Master Thesis, Lund University, Lund, Sweden, 2024. Available online: <https://lup.lub.lu.se/luur/download?func=downloadFile&recordId=9159294&fileId=9159297> (accessed on 14 June 2026).
34. van Bavel, R.; Rodríguez-Priego, N. *Testing the Effect of Cookie Banners on Behavior*; JRC Technical Reports; EUR 28287 EN; Publications Office of the European Union: Luxembourg, 2016. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Zamfir, M.D. Scarcity Effect and Consumer Decision Biases: How Urgency Influences the Perceived Value of Products. *J. World Econ.* **2024**, *3*, 27–34. [\[CrossRef\]](#)
36. Sukhorukov, O. The impact of dark patterns on user trust and long-term engagement: An ethical analysis. *World J. Adv. Res. Rev.* **2025**, *25*, 296–303. [\[CrossRef\]](#)
37. Spelt, H.; Dijk, E.K.-V.; Ham, J.; Westerink, J.; Ijsselstein, W. Psychophysiological Measures of Reactance to Persuasive Messages Advocating Limited Meat Consumption. *Information* **2019**, *10*, 320. [\[CrossRef\]](#)
38. Cauffman, C.; Goanta, C. A new order: The Digital Services Act and consumer protection. *Eur. J. Risk Regul.* **2021**, *12*, 758–774. [\[CrossRef\]](#)
39. European Commission. Commission Decides to Refer Czechia, Spain, Cyprus, Poland and Portugal to the Court of Justice for Lack of Effective Implementation of the Digital Services Act. Press Release. IP/25/1081. 2025. Available online: https://ec.europa.eu/commission/presscorner/api/files/document/print/en/ip_25_1081/IP_25_1081_EN.pdf (accessed on 14 June 2026).
40. Federal Trade Commission. *Bringing Dark Patterns to Light*; Staff Report; Bureau of Consumer Protection: Washington, DC, USA, 2022. Available online: https://www.ftc.gov/system/files/ftc_gov/pdf/P214800%20Dark%20Patterns%20Report%209.14.2022%20-%20FINAL.pdf (accessed on 14 June 2026).
41. Evans, N.J.; Phua, J.; Lim, H.; Jun, J. Disclosing Instagram Influencer Advertising: The Effects of Disclosure Language on Advertising Recognition, Attitudes, and Behavioral Intent. *J. Interact. Advert.* **2017**, *17*, 138–149. [\[CrossRef\]](#)
42. Vuković, D.; Pavković, T. The Ethics of Influencer Marketing: An Analysis of Transparency and Accountability in Digital Advertising. *MAP Soc. Sci.* **2025**, *6*, 40–52. [\[CrossRef\]](#)
43. Federal Trade Commission. Com Disclosures: How to Make Effective Disclosures in Digital Advertising. 2013. Available online: <https://www.ftc.gov/sites/default/files/attachments/press-releases/ftc-staff-revises-online-advertising-disclosure-guidelines/130312dotcomdisclosures.pdf> (accessed on 14 June 2026).
44. Kavoura, A.; Borges-Tiago, T.; Tiago, F. (Eds.) *Strategic Innovative Marketing and Tourism: Current Trends and Future Outlook*. In *Springer Proceedings in Business and Economics*; Springer: Cham, Switzerland, 2024. Available online: <https://pure.port.ac.uk/ws/portalfiles/portal/91471447/978-3-031-51038-0.pdf> (accessed on 14 June 2026).
45. Levitan, S. Social Media and the Mediating Role of Perceived Authenticity in Covert Celebrity Endorsement: Influencing Factors. 2020. Available online: https://www.researchgate.net/publication/373735682_Social_Media_and_the_Mediating_Role_of_Perceived_Authenticity_in_Covert_Celebrity_Endorsement_Influencing_Factors (accessed on 14 June 2026).

46. Viberg, O.; Kizilcec, R.F.; Jivet, I.; Martínez Monés, A.; Oh, A.; Mutimukwe, C.; Hrastinski, S.; Scheffel, M. Cultural Differences in Students' Privacy Concerns in Learning Analytics across Germany, South Korea, Spain, Sweden, and the United States. *arXiv* **2023**, arXiv:2312.02093.
47. Zeng, H.K.; Zhang, Y.C.; Chiang, C.H.; Tseng, W.L. Consumer Evaluation of Virtual vs. Human Influencers via Consumption Motivation: Role of Influencer Gender and Prior Brand Attitude. *J. Electron. Commer. Organ.* **2025**, *20*, 168. [CrossRef]
48. Khalfallah, D.; Keller, V. Authenticity, ethics, and transparency in virtual influencer marketing: A cross-cultural analysis of consumer trust and engagement: A systematic literature review. *Acta Psychol.* **2025**, *260*, 105573. [CrossRef] [PubMed]
49. Jayasingh, S.; Sivakumar, A.; Vanathaiyan, A.A. Artificial Intelligence Influencers' Credibility Effect on Consumer Engagement and Purchase Intention. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 17. [CrossRef]
50. Mertens, F.; Goetghebuer, J. Virtual Reality, Real Responsibility: The Regulatory Landscape for Virtual Influencers. In *Financial Law Institute Working Paper Series*; Ghent University: Ghent, Belgium, 2024. [CrossRef]
51. Poell, T.; Nieborg, D.; van Dijck, J. Platformization. *Internet Policy Rev.* **2019**, *8*, 1–13. [CrossRef]
52. Ershov, D.; He, Y.; Seiler, S. How Much Influencer Marketing is Undisclosed? Evidence from Twitter. 2025. Available online: <https://papers.ssrn.com/sol3/Delivery.cfm/4607235.pdf?abstractid=4607235&mirid=1> (accessed on 14 June 2026).
53. Kitchin, R. Thinking Critically about and Researching Algorithms. *Inf. Commun. Soc.* **2017**, *20*, 14–29. [CrossRef]
54. Pasquale, F. *The Black Box Society: The Secret Algorithms That Control Money and Information*; Harvard University Press: Cambridge, MA, USA, 2015. Available online: <https://raley.english.ucsb.edu/wp-content/Engl800/Pasquale-blackbox.pdf> (accessed on 14 June 2026).
55. Melton, H.; Abernathy, R.; Ogunyanwo, H. *What Brands Need To Know About Consumer Reviews in 2024*; Crowell & Moring LLP: Washington, DC, USA, 2024. Available online: <https://www.crowell.com/a/web/rPji2d8Vgou8DfecmLycVg/what-brands-need-to-know-for-marketing-compliance-in-2024.pdf> (accessed on 14 June 2026).
56. Zuboff, S. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*; PublicAffairs: New York, NY, USA, 2019. Available online: https://raggeduniversity.co.uk/wp-content/uploads/2024/08/1_x_Shoshana-Zuboff-The-Age-of-Surveillance-Capitalism_The-Fight-for-a-Human-Future-at-the-New-Frontier-of-Power-PublicAffairs-Books-2019.pdf (accessed on 14 June 2026).
57. Verhulst, S.; Young, A. Identifying and addressing data asymmetries so as to enable (better) science. *Front. Big Data* **2022**, *5*, 888384. [CrossRef] [PubMed]
58. Audrezet, A.; Koles, B.; Moulard, J.G.; Ameen, N.; McKenna, B. Virtual influencers: Definition and future research directions. *J. Bus. Res.* **2025**, *200*, 115647. [CrossRef]
59. Plantin, J.C.; Lagoze, C.; Edwards, P.N.; Sandvig, C. Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media Soc.* **2018**, *20*, 293–310. [CrossRef]
60. Fragiadakis, G.; Diou, C.; Kousiouris, G.; Nikolaidou, M. Evaluating Human-AI Collaboration: A Review and Methodological Framework. *arXiv* **2024**, arXiv:2407.19098.
61. European Commission. *Ethics by Design and Ethics of Use Approaches for Artificial Intelligence*, 1st ed.; DG Research & Innovation: Brussels, Belgium, 2021. Available online: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf (accessed on 14 June 2026).
62. Cavoukian, A. Privacy by Design: The 7 Foundational Principles. Information and Privacy Commissioner of Ontario. 2011. Available online: https://student.cs.uwaterloo.ca/cs492/papers/7foundationalprinciples_longer.pdf (accessed on 14 June 2026).
63. Hanna, M.G.; Pantanowitz, L.; Jackson, B.; Palmer, O.; Visweswaran, S.; Pantanowitz, J.; Deebajah, M.; Rashidi, H.H. Ethical and Bias Considerations in Artificial Intelligence/Machine Learning. *Mod. Pathol.* **2025**, *38*, 100686. [CrossRef] [PubMed]
64. Schilke, O.; Reimann, M. The transparency dilemma: How AI disclosure erodes trust. *Organ. Behav. Hum. Decis. Process.* **2025**, *188*, 104405. [CrossRef]
65. Hari, H.; Sharma, A.; Verma, S.; Chaturvedi, R. Exploring Ethical Frontiers of Artificial Intelligence in Marketing. *J. Responsible Technol.* **2025**, *21*, 100103. [CrossRef]
66. Choudhary, M. The Algorithmic Persuader: Ethical Challenges in AI-Powered Behavioral Manipulation in Digital Marketing. In *Proceedings of the RAIS Conference Proceedings*, Princeton, NJ, USA, 17–18 April 2025; pp. 13–19. [CrossRef]
67. Conti, M.; Gathani, J.; Tricomi, P.P. Virtual Influencers in Online Social Media. *IEEE Commun. Mag.* **2022**, *60*, 86–91. [CrossRef]
68. Löschner, D.M.; Pannasch, S. Measuring the Deceptive Potential of Design Patterns: A Decision-Making Game. In *Proceedings of the Workshop on Deceptive Design Patterns in Human-Computer Interaction (DDPCHI 2024)*. CEUR Workshop Proceedings, Honolulu, HI, USA, 12 May 2024; Volume 3720. Available online: <https://ceur-ws.org/Vol-3720/paper9.pdf> (accessed on 14 June 2026).
69. Hardcastle, K.; Vorster, L.; Brown, D.M. Understanding Customer Responses to AI-Driven Personalized Journeys: Impacts on the Customer Experience. *J. Advert.* **2025**, *54*, 176–195. [CrossRef]

70. Peukert, C.; Bechtold, S.; Batikas, M.; Kretschmer, J. Regulatory Spillovers and Data Governance: Evidence from the GDPR. *Mark. Sci.* **2022**, *41*, 746–768. [CrossRef]
71. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [CrossRef] [PubMed]
72. Lamichhane, B.; Lamichhane, S.; Pandey, P.; Regmi, S. Impact of Digital Marketing on Business Performance: A PRISMA Model. *J. Dev. Res.* **2025**, *10*, 33–45. [CrossRef]
73. Warr, W.; Aveyard, P.; Albury, C.; Nicholson, B.; Tudor, K.; Hobbs, R.; Roberts, N.; Ziebland, S. A systematic review and thematic synthesis of qualitative studies exploring GPs' and nurses' perspectives on discussing weight with patients with overweight and obesity in primary care. *Obes. Rev.* **2021**, *22*, e13151. [CrossRef] [PubMed]
74. Verma, S.; Sharma, R.; Deb, S.; Maitra, D. Artificial intelligence in marketing: Systematic review and future research direction. *Int. J. Inf. Manag. Data Insights* **2021**, *1*, 100002. [CrossRef]
75. Lefebvre, C.; Glanville, J.; Briscoe, S.; Littlewood, A.; Marshall, C.; Metzendorf, M.; Noel-Storr, A.; Rader, T.; Shokraneh, F.; Thomas, J.; et al. Searching for and selecting studies. In *Cochrane Handbook for Systematic Reviews of Interventions*; Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J., Welch, V.A., Eds.; John Wiley & Sons: Chichester, UK, 2019; pp. 67–107. [CrossRef]
76. Heyvaert, M.; Hannes, K.; Maes, B.; Onghena, P. Critical Appraisal of Mixed Methods Studies. *J. Mix. Methods Res.* **2013**, *7*, 302–327. [CrossRef]
77. Joanna Briggs Institute. *Critical Appraisal Tools*; Joanna Briggs Institute, University of Adelaide: Adelaide, Australia, 2020. Available online: <https://jbi.global/critical-appraisal-tools> (accessed on 26 September 2025).
78. Belur, J.; Tompson, L.; Thornton, A.; Simon, M. Interrater Reliability in Systematic Review Methodology: Exploring Variation in Coder Decision-Making. *Sociol. Methods Res.* **2021**, *50*, 837–865. [CrossRef]
79. DPO Europe. *Balancing Innovation and Data Privacy: The GDPR Expert's Role in AI-Driven Marketing*; DPO Europe Blog: Berlin, Germany, 2024. Available online: <https://data-privacy-office.eu/balancing-innovation-and-data-privacy-the-gdpr-experts-role-in-ai-driven-marketing/> (accessed on 14 June 2026).
80. Odumuwan, O.; Olanrewaju, A. Future of Security in FinTech: Balancing User Privacy, Compliance and Technological Advancements. *Int. J. Res. Publ. Rev.* **2025**, *6*, 1296–1310. Available online: https://www.researchgate.net/publication/388190214_Future_of_Security_in_FinTech_Balancing_User_Privacy_Compliance_and_Technological_Advancements.pdf (accessed on 14 June 2026). [CrossRef]
81. Saarela, M.; Podgorelec, V. Recent Applications of Explainable AI (XAI): A Systematic Literature Review. *Appl. Sci.* **2024**, *14*, 8884. [CrossRef]
82. Chou, H.Y. How countdown time units affect consumers' intentions to participate in group-buying offers. *Electron. Commer. Res. Appl.* **2019**, *35*, 100839. [CrossRef]
83. Luguri, J.; Strahilevitz, L. Shining a Light on Dark Patterns. *J. Leg. Anal.* **2021**, *13*, 43–109. [CrossRef]
84. Friedman, D.A. Regulating drip pricing. *Stanf. Law Policy Rev.* **2020**, *31*, 51–101. Available online: <https://ssrn.com/abstract=3337073> (accessed on 14 June 2026).
85. Car, P.; Cassetti, F. Regulating Dark Patterns in the EU: Towards Digital Fairness, European Parliamentary Research Service. PE 767.191. 2025. Available online: https://www.europarl.europa.eu/RegData/etudes/ATAG/2025/767191/EPRS_ATA%282025%29767191_EN.pdf (accessed on 14 June 2026).
86. Mathur, A.; Acar, G.; Friedman, M.J.; Lucherini, E.; Mayer, J.; Chetty, M.; Narayanan, A. Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. *Proc. ACM Hum.-Comput. Interact.* **2019**, *3*, 1–32. [CrossRef]
87. Forbrukerrådet. *Deceived by Design: How Tech Companies Use Dark Patterns to Discourage Us from Exercising Our Rights to Privacy*; Norwegian Consumer Council (Forbrukerrådet): Oslo, Norway, 2018. Available online: <https://storage02.forbrukerradet.no/media/2018/06/2018-06-27-deceived-by-design-final.pdf> (accessed on 14 June 2026).
88. European Data Protection Board. *Guidelines 3/2022 on Dark Patterns in Social Media Platform Interfaces: How to Recognise and Avoid Them*, 2nd ed.; European Data Protection Board: Brussels, Belgium, 2023. Available online: https://www.edpb.europa.eu/system/files/2022-03/edpb_03-2022_guidelines_on_dark_patterns_in_social_media_platform_interfaces_en.pdf (accessed on 14 June 2026).
89. Federal Trade Commission. *Trade Regulation Rule on the Use of Consumer Reviews and Testimonials*; Federal Trade Commission: Washington, DC, USA, 2024; Volume 89, pp. 68034–68114. Available online: <https://www.federalregister.gov/documents/2024/08/22/2024-18519/trade-regulation-rule-on-the-use-of-consumer-reviews-and-testimonials> (accessed on 14 June 2026).
90. European Commission. Consumer Protection Cooperation Regulation. European Commission—Law. 2025. Available online: https://commission.europa.eu/law/law-topic/consumer-protection-law/consumer-protection-cooperation-regulation_en (accessed on 26 September 2025).

91. Ciani Sciolla, J. “Dark Partners”: Transparency Obligations against Deception in Virtual Influencer, Marketing. In *The Leading Role of Smart Ethics in the Digital World*; Pelegrín-Borondo, J., Arias-Oliva, M., Murata, K., Romero, M.S., Eds.; Universidad de La Rioja: Logroño, Spain, 2025; pp. 143–154. Available online: <https://dialnet.unirioja.es/descarga/articulo/9537225.pdf> (accessed on 14 June 2026).
92. Fatima, S. *FTC Guidelines for Influencers: Everything You Need to Know in 2025*; InBeat Agency Blog: Montreal, QC, Canada, 2025. Available online: <https://inbeat.agency/blog/ftc-guidelines-for-influencers> (accessed on 14 June 2026).
93. Lincoln, Y.S.; Guba, E.G. *Naturalistic Inquiry*; Sage Publications: Newbury Park, CA, USA, 1985.
94. Carter, N.; Bryant-Lukosius, D.; DiCenso, A.; Blythe, J.; Neville, A.J. The use of triangulation in qualitative research. *Oncol. Nurs. Forum* **2014**, *41*, 545–547. [[CrossRef](#)] [[PubMed](#)]
95. Tricco, A.C.; Lillie, E.; Zarin, W.; O’Brien, K.K.; Colquhoun, H.; Levac, D.; Moher, D.; Peters, M.D.J.; Horsley, T.; Weeks, L.; et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann. Intern. Med.* **2018**, *169*, 467–473. [[CrossRef](#)] [[PubMed](#)]

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