The Dark Side of AI-Driven Persuasion: Algorithmic Manipulation and Consumer Autonomy

Dina Rizkia Dewi [1]

[1] Universitas Diponegoro, Semarang, Indonesia

Abstract

This article examines the dark side of artificial intelligence driven persuasion in digital markets by synthesising interdisciplinary research on algorithmic personalisation, dark patterns and consumer autonomy. Drawing on a systematic literature review of peer reviewed work in marketing, consumer behaviour, human computer interaction, information systems and law, the study maps how artificial intelligence based systems have shifted persuasion from discrete campaign level tactics to an always on infrastructure that continuously predicts, tests and optimises individual responses. The review identifies recurring mechanisms through which granular targeting, interface level design strategies and opaque optimisation practices blur the boundary between legitimate persuasion and manipulation and create structural risks for consumers. Across the literature, consumer autonomy emerges as the central normative concern, as artificial intelligence simultaneously supports and erodes people's capacity to act on their own reasons and values in practice. The article concludes by outlining implications for responsible design and governance that seek to retain the benefits of personalisation while preventing systematic manipulation.

1. Introduction

Artificial intelligence driven systems have rapidly become the hidden infrastructure of contemporary persuasion in digital markets. From recommendation engines and dynamic pricing to real time content optimization, marketers increasingly rely on algorithms that learn from vast traces of consumer data to personalise messages and offers at the level of the individual. Recent reviews show that artificial intelligence is now embedded across the marketing, consumer research and psychology literatures and is widely credited with improving targeting accuracy, conversion and customer experience in digital environments (Rabby et al., 2021; Mariani et al., 2022). As these systems migrate from experimental tools to always on decision engines, persuasion becomes less a matter of broad segmentation and more a continuous, data driven process that anticipates and shapes consumer responses across platforms.

Alongside this optimistic narrative, a growing stream of work highlights the ethical and societal risks created by artificial intelligence in marketing. Structured reviews of "ethical marketing AI" document concerns that data intensive systems can entrench privacy violations, amplify discrimination and generate broader social distortions when commercial objectives are optimised without explicit safeguards (Hermann, 2022; Su et al., 2023). These debates resonate with broader anxieties in digital markets regarding dark patterns, hyper personalised messaging and opaque forms of influence that are difficult for consumers, regulators and even marketers to detect and contest. The same capabilities that allow firms to tailor content to

individual preferences also enable subtle steering of attention, emotion and choice at scale.

To capture these developments, communication scholars have begun to speak of "algorithmic persuasion". Zarouali et al. (2022) conceptualise this as a dynamic chain in which behavioural and contextual data serve as inputs for algorithmic techniques and objectives, which then generate personalised persuasion attempts that shape subsequent processing and effects. In such systems, algorithms are not neutral intermediaries but active agents that decide which message appears, in what form and at what moment, based on predicted susceptibility and value. Legal scholarship goes further and introduces the notion of "manipulation by algorithms", arguing that highly granular targeting and optimisation can systematically exploit cognitive biases and information asymmetries in ways that undermine fair commercial practice and democratic deliberation (Hacker, 2021). These developments suggest that the boundary between legitimate persuasion and algorithmic manipulation is becoming increasingly blurred.

At the centre of these debates lies the question of consumer autonomy. Autonomy is often understood as the capacity of individuals to form preferences and make choices that reflect their own reasons, values and interests. Bjørlo et al. (2021) argue that artificial intelligence based decision aids both support and threaten this capacity: they can simplify complex decisions and increase relevance, yet they also create new possibilities for hidden interference that gradually depletes autonomy as a societal resource. When persuasion is orchestrated by opaque models that continuously test and refine which cues keep users scrolling, clicking or buying, it

becomes difficult for consumers to recognise where their own agency ends and algorithmic steering begins. The risk is not only that individuals are persuaded to act against their long term interests, but also that they lose the practical ability to resist or even perceive such influence.

Despite these concerns, empirical research that directly links specific AI driven persuasion strategies to perceived autonomy, manipulation and resistance remains fragmented. Existing studies tend to treat artificial intelligence in marketing as either a source of efficiency and engagement or as a broad ethical problem, with limited integration between psychological, legal and marketing perspectives on how algorithmic systems shape the experience of choosing. Reviews call for more work that scrutinises the "dark side" of AI in commercial persuasion and that clarifies when and how optimisation crosses the line into unacceptable manipulation (Mariani et al., 2022; Su et al., 2023). Responding to this gap, the present study focuses on the dark side of AI driven persuasion by examining the mechanisms and conditions under which algorithmic personalisation and targeting erode consumer autonomy. It seeks to articulate a conceptual distinction between persuasion and manipulation in algorithmic environments, to explore how consumers interpret and react to such practices, and to derive implications for the responsible design and governance of AI based marketing systems.

2. Literature Review

Research on artificial intelligence in marketing and consumer behaviour shows that AI has moved from a back end analytics tool to a central infrastructure

that shapes targeting, content delivery and real-time interaction. Syntheses of behavioural consumer research document how AI systems now personalise recommendations, automate messaging and adjust interfaces dynamically, thereby transforming the micro processes of persuasion in digital environments (Kim et al., 2023).

At the same time, conceptual work emphasises that AI in consumer markets is characterised by tensions between efficiency and ethics. Du and Xie (2021) describe "paradoxes" whereby the same predictive capabilities that increase relevance and convenience also heighten risks of surveillance, discrimination and loss of control over personal data, arguing that these trade offs are intrinsic to AI driven persuasion rather than accidental side effects.

Building on this, recent reviews in marketing ethics highlight how AI-enabled profiling and dynamic content optimisation can subtly steer consumers by leveraging psychological vulnerabilities and behavioural data. Gonçalves et al. (2023) show that concerns about opacity, asymmetric power and the exploitation of cognitive biases are central when AI is deployed to influence choice, and call for stronger governance of persuasive applications beyond traditional privacy compliance.

Parallel work in human computer interaction examines dark patterns as concrete interface strategies through which algorithmic persuasion is operationalised. Large scale crawls of e-commerce sites reveal widespread use of design elements that coerce, deceive or pressure users into purchases and data disclosure, demonstrating how user interfaces can systematically distort decision making in favour of commercial objectives (Mathur et al., 2019).

More recent information systems scholarship refines this perspective by defining dark patterns as user-interface designs that intentionally undermine user autonomy and informed consent. Kollmer and Eckhardt (2023) argue that many AI supported practices such as personalised scarcity messages, pre-selected defaults and obfuscated opt out paths should be understood as organisational strategies that monetise attention and data by exploiting behavioural biases, thereby raising fundamental questions about the protection of consumer autonomy in algorithmic markets. To synthesise these strands of evidence, Table 2.1 summarises prior studies cited in this section, highlighting their focus and key findings.

Table 2.1 Prior Research

Author(s) and Year	Article Title	Key Findings (Result)
Mathur et al. (2019)	Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites	The study develops automated detection methods to identify dark patterns at scale and applies them to a large audit of e-commerce interfaces. By analyzing approximately 53,000 product pages from around 11,000 shopping websites, the authors identify 1,818 dark pattern instances, covering 15 types across 7 broader categories, showing that manipulative interface tactics are widespread in online retail. They further flag 183 websites that exhibit deceptive practices and identify 22 third party providers that offer dark patterns as ready made solutions, indicating an ecosystem that supports their deployment. The study also proposes a taxonomy linking dark pattern characteristics to how they distort user decision making and outlines recommendations for researchers, regulators, and other stakeholders to mitigate and reduce these harmful design practices.
Du & Xie (2021)	Paradoxes of Artificial Intelligence in	The study develops a socio technical, multi layered ethical analysis of AI enabled products and argues that their ethical implications are driven by three core

	Consumer Markets: Ethical Challenges and Opportunities	product dimensions: multi functionality, interactivity, and the stage of AI intelligence. It identifies major ethical risks linked to these dimensions such as algorithmic bias, ethical design challenges, consumer privacy, cybersecurity threats, and potential harms to individual autonomy and wellbeing, alongside broader societal concerns like unemployment. Building on stakeholder and institutional theory, the paper proposes a conceptual framework for AI related corporate social responsibility, explaining how product level characteristics, firm level factors, and institutional environment conditions shape socially responsible AI actions and, in turn, influence outcomes for firms, consumers, and society, while also outlining a future research agenda on AI ethics and CSR.
Kim, Usman, Garvey, & Duhachek (2023)	Artificial Intelligence in Marketing and Consumer Behavior Research	The review finds that consumer responses to artificial intelligence in marketing are mixed and highly context dependent: AI can improve outcomes through personalization and relationship building, but it can also produce bot effects and algorithm aversion. The impact is shaped by key boundary conditions such as task agent fit (AI is often preferred for more utilitarian decisions), disclosure of AI identity (which can reduce purchase outcomes in some settings), and anthropomorphism, which may either enhance engagement or backfire depending on the consumer's state and the interaction context.
Gonçalves, Pinto, Rita, & Pires (2023)	Artificial Intelligence and its Ethical Implications for Marketing	The study proposes and empirically tests a conceptual model explaining consumers' acceptance of AI marketing services by integrating acceptance theory with perceived risk, trust, attitudes toward AI, and ethical concerns. Based on survey data from 200 consumers, the findings show that perceived risk is a key driver shaping attitudes toward AI, ethical concerns, and perceived trust. The results also indicate a significant relationship between perceived risk, ethical concerns, and social norms, suggesting that consumer ethical judgments and normative pressures

		are closely tied to how risky they believe AI-based marketing is.
Kollmer & Eckhardt (2023)	Dark Patterns	The paper defines dark patterns as interface designs that undermine user autonomy by preventing informed choices and potentially causing harm (for example, privacy loss or financial harm). It explains that dark patterns work through digital dark nudging and digital sludging, enabled by manipulation (choice composition/complexity) and deception (fabrication/omission) techniques that exploit behavioral biases. The authors also argue that existing consent-focused regulations may be insufficient and propose future research directions focusing on users, organizations, and regulators to better understand harms, drivers, and effective governance.

3. Methods

This study employs a systematic literature review approach to synthesise and integrate existing knowledge on the dark side of AI driven persuasion, algorithmic manipulation, and consumer autonomy. The review followed a structured protocol covering planning, search, screening, and synthesis stages. First, a preliminary scoping was conducted to refine the research questions and identify core constructs such as AI-driven persuasion, algorithmic targeting, dark patterns, manipulation, and consumer autonomy. Based on this scoping, a set of keyword strings and Boolean combinations was developed (for example, "artificial intelligence" AND "persuasion," "algorithmic manipulation," "dark patterns," "personalised advertising," "consumer autonomy," "algorithmic decision-making"). These terms were used to search major academic databases, including Scopus, Web of Science, ScienceDirect, and Google Scholar, focusing on recent, peer-reviewed journal

articles and high quality conference proceedings in marketing, consumer behaviour, human computer interaction, information systems, and law/ethics.

Inclusion criteria required that studies explicitly address AI or algorithmic systems used for persuasive, influence, or choice architecting purposes in commercial or consumer facing contexts, and that they discuss at least one of the following: mechanisms of personalisation and targeting, manifestations of manipulation or dark patterns, or implications for autonomy, agency, or informed consent. Conceptual, theoretical, empirical, and review articles were all eligible, while non peer reviewed work, purely technical optimisation papers without behavioural or ethical discussion, and studies unrelated to persuasion were excluded. Titles, abstracts, and full texts were screened in successive stages by applying these criteria, and the final corpus was subjected to systematic coding. For each article, information was extracted regarding the type of AI system, domain of application, persuasive or manipulative mechanisms, consumer outcomes (such as perceived fairness, trust, autonomy, and resistance), and proposed governance or design safeguards. A thematic synthesis was then conducted to identify recurring patterns and tensions across disciplines, map how the boundary between persuasion and manipulation is conceptualised, and develop an integrative framework that links AI driven personalisation, dark patterns, and consumer autonomy in algorithmic markets.

4. Results and Discussion

The systematic review confirms that artificial intelligence has evolved into a pervasive infrastructure of persuasion rather than a discrete marketing tool. Across

the corpus, studies consistently show that AI systems personalise recommendations, offers and interfaces in real time, supporting earlier claims that AI enhances targeting accuracy, conversion and customer experience (Rabby et al., 2021; Mariani et al., 2022; Kim et al., 2023). This body of work depicts AI driven persuasion as continuous and adaptive, with algorithms constantly updating their models of individuals based on behavioural and contextual data. In line with Zarouali et al. (2022) notion of "algorithmic persuasion", the findings indicate that persuasion is increasingly orchestrated by systems that determine not only which message is shown, but also when and how it is framed, based on predictions of susceptibility and value.

At the same time, the review reveals a strong and growing concern with the ethical and societal implications of this shift. Building on the paradoxes highlighted by Du and Xie (2021), several studies emphasise that the very capabilities that make AI attractive for marketers fine grained prediction, segmentation and optimisation also create structural risks of surveillance, discrimination and loss of control over personal data. Work on ethical marketing AI and governance argues that, in the absence of explicit safeguards, commercial optimisation tends to prioritise short term engagement and revenue, thereby entrenching power asymmetries between platforms and consumers (Hermann, 2022; Su et al., 2023). These findings support the view that efficiency and ethics are not independent dimensions but intertwined outcomes of how AI driven persuasion is designed and governed.

The analysis of interface level practices further demonstrates how algorithmic persuasion is operationalised through concrete design strategies that approach

manipulation. Studies of dark patterns show that e-commerce and platform interfaces routinely deploy coercive, deceptive or emotionally pressuring elements to steer purchases and data disclosure, confirming that user interfaces can systematically distort decision making in favour of commercial objectives (Mathur et al., 2019). Information systems research extends this by framing many AI supported practices such as personalised scarcity messages, pre-ticked boxes and obfuscated opt out flows as organisational strategies that monetise attention and data by exploiting behavioural biases (Kollmer & Eckhardt, 2023). Taken together with legal analyses of "manipulation by algorithms", which highlight how granular targeting and optimisation can exploit cognitive biases and information asymmetries, these findings suggest that the boundary between legitimate persuasion and manipulation is increasingly porous in algorithmic markets (Hacker, 2021).

Consumer autonomy emerges as the central normative concern that links these strands of evidence. Conceptual work on AI-based decision aids argues that such systems simultaneously support and threaten autonomy: they simplify complex choices and provide relevant options, yet they also enable new forms of hidden interference that gradually erode individuals' capacity to act on their own reasons and values (Bjørlo et al., 2021). The reviewed studies indicate that when dark patterns and opaque optimisation are layered on top of personalisation, consumers face increasing difficulty in recognising where their own agency ends and algorithmic steering begins. This aligns with ethical marketing and legal scholarship calling for stronger protections around autonomy, meaningful consent and the right to resist

or opt out of hyper personalised persuasion (Hacker, 2021; Gonçalves et al., 2023; Su et al., 2023).

Overall, the findings suggest that AI driven persuasion cannot be understood solely as a tool for improving relevance and efficiency. Rather, it constitutes a socio technical infrastructure that reconfigures how influence is exercised, how choices are structured, and how autonomy is experienced in digital markets. The review highlights the need for integrative frameworks that combine marketing, psychology, human computer interaction, information systems and law to distinguish more clearly between persuasion and manipulation in algorithmic environments. It also points to a research agenda focused on consumer interpretations and resistance strategies, as well as on concrete design and governance mechanisms that can retain the benefits of AI driven personalisation while protecting autonomy and preventing systematic manipulation.

5. Conclusion

This study has examined the dark side of AI-driven persuasion by synthesising evidence on how algorithmic personalisation, dark patterns and granular targeting reshape influence and consumer autonomy in digital markets. The review shows that AI has shifted persuasion from a discrete marketing activity to an always-on infrastructure that continuously predicts, tests and optimises individual responses. While this infrastructure undeniably enhances relevance, targeting accuracy and customer experience, it also embeds structural risks: opaque data practices, amplification of behavioural biases, and interface designs that blur the line between

legitimate influence and manipulation. Taken together, the findings suggest that AI-driven persuasion reconfigures not only how messages are delivered, but also how choices are framed and experienced, making the boundary between persuasion and algorithmic manipulation increasingly porous.

At the normative core of these dynamics lies consumer autonomy. The literature indicates that AI-based systems simultaneously support and erode autonomy: they help individuals cope with complexity, yet they introduce new forms of hidden interference that make it harder for consumers to recognise and resist undue influence. This implies that debates about AI in marketing cannot be reduced to a trade-off between efficiency and privacy, but must explicitly address how socio technical design choices affect people's capacity to act on their own reasons and values. The review therefore underscores the need for integrative frameworks and governance mechanisms that connect marketing, psychology, human computer interaction, information systems and law, with a dual aim: to harness the benefits of AI-driven personalisation for relevance and value creation, while instituting substantive safeguards at the level of algorithms, interfaces and regulation to prevent systematic manipulation and to protect consumer autonomy in algorithmic markets.

References

Bjørlo, L., Moen, Ø., & Pasquine, M. (2021). The role of consumer autonomy in developing sustainable AI: A conceptual framework. *Sustainability, 13*(4), 2332.

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

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

Hacker, P. (2023). Manipulation by algorithms: Exploring the triangle of unfair commercial practice, data protection, and privacy law. *European Law Journal, 29*(1–2), 142–175.

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

Kim, T., Usman, U., Garvey, A., & Duhachek, A. (2023). Artificial intelligence in marketing and consumer behavior research. *Foundations and Trends in Marketing, 18*(1), 1–93.

Kollmer, T., & Eckhardt, A. (2023). Dark patterns. *Business & Information Systems Engineering, 65*(2), 201–208.

Mariani, M. M., Perez-Vega, R., & Wirtz, J. (2022). AI in marketing, consumer research and psychology: A systematic literature review and research agenda. *Psychology & Marketing, 39*(4), 755–776.

Mathur, A., Acar, G., Friedman, M. J., Lucherini, E., Mayer, J., Chetty, M., & Narayanan, A. (2019). Dark patterns at scale: Findings from a crawl of 11K shopping websites. *Proceedings of the ACM on Human-Computer Interaction, 3*(CSCW), 1–32.

Rabby, F., Chimhundu, R., & Hassan, R. (2021). Artificial intelligence in digital marketing influences consumer behaviour: A review and theoretical foundation for future research. *Academy of Marketing Studies Journal, 25*(5), 1–7.

Su, Y., Wang, E. J., & Berthon, P. (2023). Ethical marketing AI? A structured literature review of the ethical challenges posed by artificial intelligence in the domains of marketing and consumer behavior. In *Proceedings of the 56th Hawaii International Conference on System Sciences* (pp. 4933–4942).

Zarouali, B., Boerman, S. C., Voorveld, H. A., & van Noort, G. (2022). The algorithmic persuasion framework in online communication: Conceptualization and a future research agenda. *Internet Research, 32*(4), 1076–1096.